Unabhängiges Landeszentrum für
Datenschutz Schleswig-Holstein

Sommerakademie 2013

Big Data – Informationelle Fremd- oder Selbstbestimmung?!

Thilo Weichert

„Wer hat Angst vor großen Daten?“

Das Thema der heutigen Sommerakademie – Big Data und Datenschutz – legten wir vor etwa einem Jahr fest. Unsere Idee war es, einen sich absehbaren Trend beim Einsatz von Informationstechnik frühzeitig auf seine datenschutzrechtlichen Auswirkungen hin zu untersuchen, um von vornherein Fehlentwicklungen zu verhindern. Der Whistleblower Edward Snowden vermittelte uns nun vor knapp drei Monaten die Erkenntnis, dass westliche Nachrichtendienste längst das praktizieren, was von uns heute präventiv kritisch hinterfragt werden sollte. Die Meldungen im vorletzten SPIEGEL zur Rezeptdatenverarbeitung bestätigen, dass auch in der Privatwirtschaft, hier im Medizinbereich, personenbeziehbares Big Data ein zentrales Betätigungsfeld der Informationstechnik und zugleich eine zentrale Bedrohung für das Recht auf informationelle Selbstbestimmung geworden ist.

Big Data ist eine logische Konsequenz der bisherigen informationstechnischen Entwicklungen. Nachdem immer mehr digitale Spuren anfallen und diese – unabhängig vom Ursprungsformat – miteinander verknüpft werden können, stellt sich die Frage, welche Erkenntnisse hieraus gezogen werden. Mit modernen Analysetools lassen sich Selektionen und Auswertungen vornehmen, die Rückschlüsse auf einzelne Datensätze und damit auf Menschen ermöglichen. Von der Industrie propagierte Datenanalysen zielen häufig auf Erkenntnisse, ohne dass ein Personenbezug nötig wäre. Den Geheimdiensten, aber auch z. B. der Werbeindustrie, geht es dagegen um personenbezogene Datennutzungen. Dabei werden grundlegende Datenschutzprinzipien herausgefordert: Die Grundsätze der Gesetzmäßigkeit, der Erforderlichkeit und der Datensparsamkeit, die Zweckbindung und grundsätzlich die Verhältnismäßigkeit werden verletzt, ebenso wie die Pflicht zur Transparenz über die Datenverarbeitung.

Wir Datenschützer dürfen angesichts der schon weit verbreiteten Praxis heute hier nicht nur kluge Fragen zur Diskussion stellen. Vielmehr müssen wir uns – wieder einmal – angesichts gewaltiger aktueller Missstände und Datenschutzverletzungen damit auseinandersetzen, wie die Chancen der Informationsverarbeitung realisiert werden können, ohne dass eine unkontrollierte Invasion in unsere Privatsphäre erfolgt. Dabei sind wir schon viel weiter als die Obama-Administration oder die Bundesregierung, die behaupten, dass Sicherheit vor Terrorismus und vor Kriminalität nur durch die Überwachung der Bevölkerung und durch massenhafte individuelle Grundrechtseingriffe verwirklichbar sei.

Ohne das Ergebnis der heutigen Diskussionen vorwegnehmen zu wollen, möchte ich schon eingangs einige zentrale Antworten auf die Fragen geben, die sich uns mit Big Data stellen. Meine Hauptthese ist: Big Data ist verfassungskonform möglich. Dabei beziehe ich mich nicht auf das fragwürdige Verfassungsverständnis, das derzeit die Obama-Administration oder auch unser Bundesinnenminister Hans-Peter Friedrich demonstrieren, bei dem unser Recht auf Privatsphäre durch ein Supergrundrecht auf Sicherheit verdrängt wird. Ich beziehe mich auf das deutsche Bundesverfassungsgericht, das in einer Vielzahl von Entscheidungen nachhaltig begründet hat, dass ein umfassend verstandenes subjektives Recht auf informationelle Selbstbestimmung unter den technischen Bedingungen unserer Informationsgesellschaft nicht nur realisiert werden kann, sondern dass es grundlegend für deren Freiheitlichkeit und für deren rechtsstaatliche und demokratische Verfasstheit ist.

Die vom Bundesverfassungsgericht gegebenen Stichworte, die auch bei Big Data relevant werden, klingen zunächst wenig erregend: Grundrechtsschutz durch Verfahren, Pseudonymisierung und Anonymisierung, Protokollierung und Dokumentation, Sicherung der Verantwortlichkeit, Transparenz und unabhängige sowie demokratische Kontrolle.

Tatsächlich hat das ULD einige Anwendungen als datenschutzkonform zertifiziert, die man als Big Data kennzeichnen kann. So gelang es einigen europäischen Internetwerbeanbietern, ihre Auswertungen digitaler Nutzungsspuren so zu aggregieren, dass gezielte Werbung geschaltet und verkauft werden kann, ohne mit dem Datenschutzrecht in Konflikt zu geraten. Aggregierungen können sowohl durch das Zusammenführen von Merkmalen als auch durch das Zusammenführen von Datensätzen umgesetzt werden. Wichtig ist, dass im Ergebnis anonymisierte Datensätze entstehen und dass eine nachträgliche Reidentifizierung nicht mehr möglich ist. Auf sensible Datenkategorien kann verzichtet werden.

Aufbewahrungszeiten werden begrenzt. Durch Auslagerung von Anonymisierungsdienstleistungen auf getrennte Stellen wird das Reidentifizierungsrisiko von Datenbeständen reduziert.

In Zusammenhang mit Big Data ist auch eine vom ULD vorgenommene Zertifizierung des „Zentralen Kassenprüfungssystems“ von Lidl spannend. Lidl hat sich ja in der Vergangenheit nicht nur durch ausgeprägtes Datenschutzbewusstsein profiliert. Mit seinem Kassensystem will es aber – und dies mit Erfolg – zu diesem Image einen Kontrapunkt setzen. Hierbei erfolgt eine umfassende Analyse sämtlicher Kassendaten zur Aufdeckung von Manipulationen im Kassierprozess. Die Auswertung beruht auf der Abgleichung der Kassendaten mit vermuteten Vorgehensweisen bei Manipulationen und auf der Überprüfung der Überschreitung von bestimmten festgelegten Grenzwerten, z. B. bei Fehlbuchungen, Stornos u. Ä. Die Daten werden nach einer kurzen Frist gelöscht, wenn sich keine Auffälligkeiten ergeben. Eine Speicherung erfolgt ausschließlich in pseudonymisierter Form; nur wenn die Auffälligkeiten sich zu einem spezifischen Verdacht konkretisieren, dürfen die Pseudonyme nach einem genau festgelegten Verfahren offengelegt werden. Auch hier erfolgt, wie bei einer zertifizierten Online-Werbedirektansprache, die Verarbeitung durch einen in der Sache nicht unmittelbar interessierten, also eher neutralen Dienstleister. Die Erkennung von Manipulationsmustern kann auch in einer derart abgestuften Form ablaufen, dass zunächst nur Gruppenanalysen vorgenommen werden und erst bei einer Konkretisierung eines Verdachts durch eine bestimmte Intensität oder Dauer Einzeldaten in der Gruppe identifiziert werden. Die Personalvertretung wird eingebunden.

Was unterscheidet nun ein solches Verfahren von den Datenauswertungen der National Security Agency – der NSA? Der Ausgangsdatenbestand sind in beiden Fällen verdachtsfreie Massendaten, ja sogar ein Gesamtdatenbestand. Durch die technischen und prozeduralen Sicherungen werden aber Falschverdächtigungen vermieden bzw. korrigiert. Streng definierte Rollen und Zugriffskonzepte reduzieren das Missbrauchsrisiko. Durch Transparenz der Verfahrensgrundlagen lassen sich sämtliche Auswertungsergebnisse nachvollziehen und überprüfen. Es wird eine interne Datenschutzkontrolle durchgeführt. Eine demokratische Legitimation ist vorgesehen, bei Lidl durch die Beteiligung der Personalvertretung.

Leider läuft bei der Auseinandersetzung um die Massendatenauswertung durch die US-amerikanische NSA oder das britische GCHQ das Hinterfragen des Verdachtsgewinnungsprozesses ins Leere. Selbst den Parlamentariern im zuständigen Kontrollausschuss wurden bisher offensichtlich in Bezug auf die strategische Telekommunikationsüberwachung des Bundesnachrichtendienstes ebenso wie bei der Kooperation mit der NSA keine prüffähigen Fakten vorgelegt. Ähnlich unkontrolliert – soweit bekannt – ergehen die Spitzelbeschlüsse der US-amerikanischen FISA-Gerichte und sind die Verfahren bei den britischen Geheimdiensten.

Transparenz wird nicht nur im staatlichen, sondern auch im privaten Bereich verweigert. Beispiel Rezeptdatenauswertung durch Apothekenrechenzentren: Diese erfolgte seit Jahrzehnten – unter Verletzung der Anonymisierungspflichten der Rechenzentren – milliardenfach. Trotz entsprechender Aufforderungen wurden z. B. dem ULD vom Dienstleister IMS Health relevante technische Details offengelegt. Apothekenrechenzentren arbeiten mit einem semantischen Trick, indem die Pseudonyme, die den Patienten zuordenbar sind, als „Patientenanonyme“ bezeichnen und ohne weiteren Beleg als

„anonym“ deklarieren. Wir erleben diesen Trick allerorten: Da die Identifizierbarkeit von Daten technisch voraussetzungsvoll ist, wird Anonymität oft unhinterfragt behauptet, wo keine Anonymität besteht.

Ein weiterer bei Big Data-Anwendungen bestehender Datenschutzmangel liegt in den Einwilligungen, die sich die Datensammler geben lassen. Das Argumentationsmuster kennen wir insbesondere von Internet-Dienstleistern wie Google und Facebook. Die viel zu unbestimmte und für den User unverständliche Zustimmung verbirgt sich hinter voreingestellten Häkchen, deren Text regelmäßig nicht aufgerufen wird. Mit Privacy by Design und Privacy by Default lassen sich aber Zustimmungen auch anlassbezogen, informiert und bewusst sowie mit technischen Mitteln einholen. Dass dies in der Praxis fast nie passiert, hat nur einen einzigen Grund: Die Anbieter wollen sich so viele Daten wie möglich ergaunern, um damit weitere Geschäfte machen zu können.

Big Data und Datenschutz zusammenzubringen, ist nicht nur juristisch und prozedural voraussetzungsvoll. Am Anfang muss immer die Technik stehen. Bisher war es das Hauptanliegen, die Datenformate so anzupassen, dass sie zusammengeführt und gemeinsam und einheitlich ausgewertet werden können. Dies ist technisch relativ banal und wird schon in Data-Warehouse-Anwendungen seit vielen Jahren praktiziert. Verantwortungsvolle Datenauswertung macht es nötig, die auszuwertenden Datenkategorien zu klassifizieren und die Datensätze mit Metadaten zu Verarbeitungsrollen, zur Relevanz und zum Zweck zu ergänzen, um eine kontrollierte und kontrollierbare Nutzung dieser Daten zu sichern. Betrachtet man sich die Datenmassen, die sich ein Bradley Manning oder ein Edward Snowden undifferenziert im Pentagon bzw. in der NSA herunterziehen konnten, so scheinen in solchen Sicherungen bisher nicht die Stärken der US-Sicherheitsbehörden zu liegen. Bei der Entwicklung des Polizeisystems @rtus kooperiert das ULD mit der Landespolizei genau vor diesem Hintergrund, worüber in einer Infobörse berichtet werden wird. In jedem Fall muss in diesen Technikfragen noch viel erforscht und entwickelt werden. Wir Datenschützer und damit auch das ULD stehen hierfür zur Kooperation bereit.

Auf der heutigen Veranstaltung wollen wir darüber diskutieren, wie Big-Data-Auswertung und informationelle Selbstbestimmung zusammengebracht werden können. Dass das geht, zeigen einzelne Pilotprojekte, über die auch berichtet werden wird. Weshalb das Machbare für mehr Datenschutz aber zumeist nicht gemacht wird, das wird uns wahrscheinlich Kopfzerbrechen bereiten wie auch die Frage, wie dieser Missstand beendet werden kann.

Die aktuellen Kommentare zum Datenschutz in der Öffentlichkeit und in den Medien sind nicht mehr: „Ich hab ja nichts zu verbergen“ und „Das machen wir doch alles freiwillig“. Seit Snowden höre ich vor allem Kommentare wie „Da kann man sowieso nichts machen“ und „Das ist alles viel zu kompliziert“. Diese neue Formen der Reaktion auf die technischen Entwicklungen sind einer freiheitlichen und demokratischen Informationsgesellschaft nicht würdig. Wenn wir gefragt werden: „Wer hat Angst vor großen Daten?“, dann sollte unsere begründete Antwort darauf sein: „Wir nicht!“

Hierfür wollen wir heute Erfahrungen, Bewertungen und Lösungen austauschen.

Ich freue mich wieder auf rege Diskussionen mit Ihnen und den Referentinnen und Referenten, wünsche Ihnen viele neue Erkenntnisse und Anregungen und schließlich auch viel Spaß!