Nicole Beranek Zanon*

Big Data und Datensicherheit

«Das Versprechen von Big Data ist, durch Analyse von Daten Probleme zu erkennen und lösen zu können, bevor sie sich gesellschaftlich ausgebreitet haben.»

Weichert, Thilo

Inhaltsverzeichnis

* Rechtsanwältin, de la cruz beranek Rechtsanwälte AG, Zug.

I. Ausgangslage

Der vorliegende Artikel befasst sich mit den Datensicherheitsaspekten[1] von Big Data bei der Bearbeitung von Personendaten durch Privatpersonen. Dazu werden zuerst die Merkmale von Big Data dargelegt und der Begriff der Datensicherheit aus datenschutzrechtlicher und technischer Sicht erläutert sowie der normative Rahmen abgesteckt. In der Folge wenden wir uns den datenschutzspezifischen Risiken von Big Data zu und erörtern die Probleme zur Gewährung der Datensicherheit bei Big Data. Zum Abschluss wird versucht, aufzuzeigen, wo gesetzgeberischer Handlungsbedarf besteht, um auch der neuen Herausforderung Big Data gewachsen zu sein, mithin Big Data nicht zu verbieten, sondern zum Vorteile unserer Gesellschaft zu nutzen.

II. Terminologie

1. Big Data

a. Einleitung

Gemäss Porter Bibb[2] wird Big Data in den nächsten Jahren keine Industrie auslassen. Derzeit wird vor allem Geld verdient durch die Speicherung von Big Data unter zur Hilfenahme von Cloud Technologien. Bereits heute entstehen neue Geschäftsmodelle durch Big Data.[3] Durch die damit zusammenhängenden Analyse bzw. Business Analytics können die Bedürfnisse von Kunden und Lieferanten besser analysiert und vorherbestimmt werden.[4] Werbemassnahmen können deshalb gezielt für spezifische Kunden-Segmente eingesetzt werden (targeted Mar-

---

1 Im Sinne des Datenschutzgesetzes.

2 Porter Bipp, Im Interview mit Yahoo Finance (http://finance.yahoo.com/video/two-companies-big-money-off-130407104.html).

3 Z.B. das Verkehrslenkungssystem von TomTom oder Wisdom Places, eine Applikation, in der man Restaurants in Echtzeit mit bestimmten Suchkriterien wie Alter, Geschlecht, kulinarische Richtung etc. verknüpft mit Facebook-Profilen, Versicherungsgesellschaften; vgl auch Stephan C Brunner, Mit rostiger Flinte unterwegs in virtuellen Welten?, in Jusletter 4. April, Ziff. 1.2, der prägnant erklärt, woher Daten «gemint» werden.

4 Die Kundenbedürfnisse und damit das Erbringen von Leistungen/Produkten entsprechend den Kundenbedürfnissen, mithin folglich die Gewinnoptimierung eines Unternehmens, ist der Zweck von Big Data, nicht das Sammeln und Auswerten als solche, dies ist nur das Mittel. Anderer Meinung: Rolf H. Weber, Big Data: Sprengkörper des Datenschutzrechts?, in: Jusletter IT 11. Dezember 2013, RZ 1.

keting[5]). Damit wird die Wirkung von Werbung effektiver und die Aufwendungen von Unternehmen oder Organisationen können zielorientiert eingesetzt werden. Ganz allgemein gesprochen – auch ausserhalb des Unternehmenskontextes – ist der Zweck von Big Data die Effizienzsteigerung, d.h. zielgerichteter Ressourceneinsatz mit besserem Ergebnis.

Big Data zeichnet sich durch vier besondere Kriterien aus: i) ***Volumen*** sowie davon abgeleitet ii) ***Geschwindigkeit*** (***Velocity),*** iii) ***Varietät***[6] die iv) ***Korrelation*** und Richtigkeit bzw. Wahrhaftigkeit (***Veracity)*** der Resultate. Auf diese Kriterien soll im Folgenden näher eingegangen werden.

b. Volumen und Geschwindigkeit

Die *Datenmengen und -berge,* die sich heute erheben, wachsen weiter über unsere Vorstellungskraft hinaus. Bereits 2012 sagte IDC (International Data Corporation) voraus, dass sich die Datenmenge bis ins Jahre 2020 verfünfzigfacht bzw. jedes zweite Jahr verdoppelt.[7] Diese Prognose ist heute überholt, nicht zuletzt, weil bereits 40% der Datenvolumen von Maschinen kreiert werden, welche mit Sensoren und IP-Adresse ausgestattet sind. Diese unheimlichen Datenberge sind alle digital und können gespeichert, durchsucht, verknüpft und geteilt werden – rund um den Globus – zu kleinen Kosten, jederzeit und in Millisekunden. Mit den zunehmenden Speichermengen wachsen die *Speichervolumen* und der Bedarf an *Rechenkapazitäten* erhöht sich. Gleichzeitig ist zu beobachten, dass die Kosten für Speichervolumen und Kapazitäten sinken.

Das geschaffene *Datenvolumen* ermöglicht es, die Daten als Grundlage für Analysen zu verwenden. Bis anhin wurden für Prognosen zukünftigen Kundenverhaltens Marktanalysen getätigt, d.h. man hat eine Stichprobe bei mindestens 1000 Kunden vorgenommen. Wenn dann z.B. eine Teilmenge der Stichprobe geantwortet hat, dass sie keine Schokolade mag, konnte nicht auf diese Stichprobeneinheit fokussiert werden, da die Teilmenge für sich aus statistischen Gründen nicht mehr aussagekräftig war. Dies ändert sich grundlegend mit Big Data. Mit Big Data kann aufgrund der grossen Datengrundlage auf ein Teilresultat fokussiert werden, bis hinab zur einzelnen Person (vgl. Ziff. III.1 unten). Genau darin liegt die datenschutzrechtliche Problematik: Mit der De-Anonymisierung werden die Resultate

---

5 Z.B. mit Einsatz von Cookies beim Browsen eines Nutzers.

6 NIST Definition von Big Data lautet: Big Data refers to digital data volume, velocity and/or variety[,veracity] that: i) enable novel approaches to frontier questions previously inaccessible or impractical using current or conventional methods; and/or ii) exceed the capacity or capability of current or conventional methods and systems.

7 IDC Report: The digital Universe in 2020: Big Data, Bigger Digital Shadows and Biggest Groth in the Far East, (<http://idcdocserv.com/1414>) (Stand 8.2.2014).

von Analysen wieder zu Personendaten, auch wenn sie zuvor anonymisiert[8] oder pseudonymisiert[9] waren.

Zudem führt das Bedürfnis nach mehr Speicherplatz und mehr Rechenleistung dazu, dass Big-Data-Anwendungen nicht mehr in eigenen Rechenzentren betrieben werden, sondern von dritter Seite als Cloud-Dienste (Infrastructure as a Service, aber auch Software as a Service) hinzugekauft werden.

c. Varietät

Das zweite Kriterium von Big Data ist *Varietät* der Daten. Bei den heutigen Datenbeständen handelt es sich um *hybride Datenbestände in strukturierter und unstrukturierter* Form. Die Daten führen somit eine grosse Varietät auf. Strukturierte Daten sind nach einem gewissen Schema abgelegt, die unstrukturierten nicht. Dies erfordert unterschiedliche Hard- und Software.

So werden strukturierte Daten in sog. SQL-Datenbanken abgespeichert. Die neuen Datenbanktypen sind sogenannte noSQL-Datenbanken; eine aufwendige Strukturierung ist nicht mehr notwendig. Um die nicht strukturierten Daten zu verarbeiten, bedarf es auch auf applikatorischer Ebene neuer Ansätze, welche Hadoop[10] liefert. Mit Hadoop wurde das *MapReduce-Konzept* eingeführt, welches

---

8 Unter Anonymisieren versteht man das (irreversible) Beseitigen der Identifikationsmerkmale von Personendaten, so dass eine Re-Identifikation ausgeschlossen respektive ohne unverhältnismässig grossem Aufwand nicht möglich ist und die ursprünglichen Personendaten zu reinen Sachdaten mutieren; m.w.H. Thomas Probst, Die Verknüpfung von Personendaten und deren rechtliche Tragweite, in: Astrid/Probst/Gammenthaler (Hrsg.), Datenverknüpfung, Problematik und rechtlicher Rahmen, Zürich 2011, S. 13 ff. sowie Urs Belser, in: Maurer-Lambrou/Vogt (Hrsg.), BSK zum Datenschutzgesetz, 2. Aufl., Basel 2006, Art. 3 DSG N 6.

9 Unter Pseudonymisieren versteht man die Trennung der persönlichen Identifikationsmerkmale von der Information, so dass eine Re-Identifikation nur (aber immerhin) durch jene Person erfolgen kann, welche den Rückschluss zwischen dem Pseudonym und der Person herstellen kann. Die De-Personalisierung geht hier weniger weit als bei der Anonymisierung. Wird nämlich bspw. bei der folgenden Information «Peter Müller, geb. 1.1.1960, hat ein Herzleiden» der Name durch ein Pseudonym ersetzt und lautet wie folgt: «P0051 hat ein Herzleiden», so ist die Person durch einen Rückschluss von «P0051» auf «Peter Müller, geb. 1.1.1960» möglich; hingegen ist durch das Anonymisieren der Information mit «jemand hat ein Herzleiden» grundsätzlich keine Identifikation mehr möglich. M.w.H. Thomas Probst (Fn. 8), 17 f.

10 Definition von Hadoop bzw. Apache Hadoop: The Apache Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Rather than rely on hardware to deliver high-availability, the library itself is designed to detect and handle failures at the application layer, so delivering a highly-available service on top of a cluster of

ermöglicht, Berechnungen auf Tausende von Rechnerknoten zu verteilen und Datenvolumina im Petabyte-Bereich verarbeiten zu können. Damit erhöht sich die Rechenkapazität.

Dieses Hadoop zugrunde liegende MapReduce-Konzept ist oft die Grundlage von standardisierten Analyse-Tools, welche uniforme Analysen ermöglichen (vgl. Ziff. II.1.e zur *Big Data Landscape*). Interessant wird es für Unternehmen aber meistens da, wo die Standards aufhören (vgl. dazu Ziff. II.1.e).

Nebst den neuen Entwicklungen auf Hardware- und Software-Ebene wird Big Data aber auch durch eine weitere Entwicklung begünstigt und führt zu mehr Varietät, nämlich *Open Data*[11]. Die öffentliche Hand stellt immer mehr Daten im Internet (und allenfalls gratis) zur Verfügung. Damit erschliessen sich für Unternehmen neue Datenquellen, die über die bisherigen hinausgehen. Dass Daten bzw. Datensammlungen aus *unterschiedlichen Quellen* stammen, erleichtert den Umgang mit ihnen keineswegs, auch dann nicht, wenn sie anonymisiert sind.

Grosse Datenmengen sind per se wertlos. Nur die Analyse derselben fördert das Gold zu Tage, nach dem heute geschürft wird: Der Einsicht über ein Nutzerverhalten und das Verhalten der Konkurrenz. Aber auch hier ist zu beachten, dass die Analysen die Vergangenheit analysieren. Über die Zukunft lassen sich damit nur Aussagen machen, wenn sich die Umstände nicht ändern. Je schneller und zeitnaher daher die Analysen gemacht werden können, umso mehr bieten sie dafür Gewähr, dass sich die Umstände zwischenzeitlich nicht geändert haben und die Resultate aussagekräftig sind. Dieses Time-Gap-Risiko hat sich mit Big Data massgeblich reduziert.

Gleichwohl lassen sich auch mit Big Data keine disruptiven Ereignisse, d.h. hoch unwahrscheinliche Ereignisse mit einer grossen Auswirkung (Black Swan[12]), voraussagen, mithin Ereignisse, welche derzeit ausserhalb des Radars unserer Vorstellungskraft sind.

Nebst dem zeitlichen Aspekt ist aber auch das Kriterium der Korrelation sehr relevant (vgl. Ziff. II.1.d nachfolgend).

---

computers, each of which may be prone to failures, (<http://hadoop.apache.org>) (Stand 8.2.2014).

11 Wie z.B. (<http://opendata.admin.ch>) (Stand 8.2.2014).

12 Nassim Nicholas Taleb, The Black Swan: The Impact of the Highly Improbable (1st ed.). London 2007.

d. Korrelation und Richtigkeit

Erst die Korrelation ermöglicht die Gewichtung des Verhältnisses von einer Aussage A zur Aussage B bzw. C. Wie viele derjenigen, die keine Schokolade mögen (Aussage A), haben eine Allergie auf Schokolade (Aussage B), und wie viele mögen sie nicht, weil sie dick macht (Aussage C)? Für den Hersteller von kalorienarmer Schokolade ist letzteres herauszufinden relevant (Verhältnis der Aussage A zu C). Damit ist die Korrelation massgeblich, d.h. das Verhältnis von Aussage A zu C. Diese kann auch ermittelt werden, wenn nur A und B bekannt sind.

Damit wird nicht mehr auf das «Warum» abgestützt (Warum mögen Sie keine Schokolade), sondern auf das «Was» (Was ist der Grund, dass sie keine Schokolade mehr mögen: die Allergie bzw. die Kalorien). Auf diese Weise stösst man mehr zum Kern dessen, was man aus Effizienzgründen wissen will.

Aufgrund der Datenmenge von Big Data hat der Schokoladenhersteller von kalorienarmer Schokolade eine wahrhaftige Aussage erhalten, auf die er die Produktion von kalorienarmer Schokolade abstützen kann (kalorienarme Schokolade nur für die Kalorienbewussten herzustellen).

Während bei der traditionellen Marktforschung eine mangelhafte Datengrundlage dazu führt, dass das Ergebnis fehlerhaft ist, ist bei Big Data eine fehlerhafte Datengrundlage nicht weiter schlimm (sofern sich natürlich die Fehler in einem gewissen Rahmen halten).[13] Aufgrund der grossen Datenmengen wird der Fehler ausgemerzt. Anders dürfte es sich nur dann verhalten, wenn die Datenbasis fast vollständig falsch ist und nicht nur einige wenige Fehler aufweist. Damit sind die Daten einer Big-Data-Analyse aus der Vogelperspektive *richtig bzw. wahrhaftig.*[14] Im Einzelfall hingegen könnten sie Fehler aufweisen. Betrifft der besagte Fehler eine Person, so ist ein Auskunfts- und Berichtigungsmechanismus vorzusehen.[15]

---

13 Im Gegensatz dazu zum bisherigen Data Warehousing und Data Mining, vgl. Urs Maurer-Lambrou, in: Maurer-Lambrou/Vogt (Hrsg.), BSK zum Datenschutzgesetz, 2. Aufl., Basel 2006, Art. 5 DSG N 4.

14 Vgl. dazu das Beispiel in Viktor Mayer-Schönberger/Kenneth Cukier, BIG DATA, A Revolution That Will Transform How We Live, Work and Think, London 2013, S. 36 ff.: Ein sehr eindrückliches Beispiel dazu ist Google Translate. Google verwendet im Hintergrund für die Übersetzungen die öffentlich übersetzten Webseiten. Diese sind nicht immer korrekt übersetzt. Die Datenbasis ist aber immens, so dass kleine Fehler untergehen. Im Gegensatz dazu hat IBM eine Übersetzungssoftware geschrieben, welche auf professionellen Übersetzungen basiert. Deren Treffsicherheit bzw. Wahrhaftigkeit ist aber aufgrund der viel kleineren Datenmenge schlechter als bei Google Translate.

15 Vgl. dazu Astrid Epiney/Tobias Fasnacht, in: Belser/Epiney/Waldmann, Datenschutzrecht, Grundlagen und öffentliches Recht, Bern 2011, §11 Rechte Einzelner N 57 ff.

e. Die Suche nach der Nadel im Heuhaufen

Für Unternehmen bestehen bereits Standard-Business- oder Web- bzw. Cloud-Analyse-Tools, um strukturierte Daten aus verschiedenen heterogenen, aber vordefinierten Datenquellen zu analysieren.

Die Big Data Landscape stellt sich damit wie folgt dar:

Log Data Apps sind Applikationen, welche Traffic-Daten erfassen, wie z.B. Cookies beim Besuch von Webseiten, Verkehrsranddaten von Verkehrsleitsystemen wie TomTom oder von Mobil-Telefonen oder -Devices. Sie erfassen damit die Spuren unserer digitalen Gesellschaft.[16]

Ad/Media Apps bezeichnen Applikationen, welche gezielte Werbung im Rahmen des Targeted Marketing i.d.R. unter Zuhilfenahme von Cookies auf Webseiten Dritter basierend auf dem Surfverhaltens des Nutzers ausliefern.

Unter *Data As A Service* werden Dienste verstanden, welche den Unternehmen Datensammlungen zur Verfügung stellen, damit Analysen gemacht werden können. Diese können Open Data sein oder aber auch von Dritten, wie z.B. Facebook/Google, zugekauft worden sein.

Unter *Business Intelligence* werden Verfahren und Prozesse zur systematischen Analyse (inkl. der Sammlung, Auswertung und Darstellung) von Daten in elektronischer Form verstanden. Es handelt sich dabei um eine Vernetzung von bereits vorhandenen Unternehmensdaten.

Unter dem Begriff der *Analytics und Visualization* werden v.a. diejenigen Dienste verstanden, welche die analytischen Resultate speziell visualisieren und graphisch darstellen. Es handelt sich damit um einen Zulieferer, der aufgrund der Analyseresultate entsprechende graphische Darstellungen erstellt.

Sodann wird auf der Infrastrukturebene unter dem Titel *Analytics Infrastruktur* entsprechende Hardware mit Software (wie Hadoop, vgl. Ziff. II.1.c) angeboten. Dies kann auch als *IaaS*-Dienst erfolgen.

Die in diesem Beitrag erwähnten Big Data Tools beziehen sich damit v.a. auf die Analyse-Tools, wie sie der Business Intelligence oder auch Business Analytics zugrunde liegen. Der Einsatz dieser Big Data Tools führt zu Veränderungen in der Unternehmensführung: Während man bis anhin in nebelumzogenen Gewäs-

[16] Die Frage der Zulässigkeit des Loggens von digitalen Spuren ist nicht Gegenstand dieses Beitrages. Es sei dazu auf den Logistep-Entscheid mit Urteil des Bundesgerichts 1C_285/2009 vom 8. September 2010 sowie dem Urteil des OLG Hamburg Az. 5 W 126/10 vom 3. November 2010 verwiesen.

sern gekreuzt ist, kann heute trotz Nebel (vorbehaltlich schwarzer Schwäne) navigiert werden.

Auch für den privaten Nutzer kann Big Data eine Veränderung bzw. insbesondere einen Vorteil bringen, so z.B. bei der Nutzung von Google Now[17]. Der Nutzer erhält bspw. aufgrund seines Kalendereintrages nicht nur eine Warnung, sondern auch gleich einen Hinweis und eine Karte, wann er wo sein muss und welchen Zug er zu besteigen hat, um rechtzeitig anzukommen.

III. Datenschutz – Grundlagen im nationalen Recht

1. Die allgemeinen Grundsätze

Was bis anhin im Data Warehouse an Daten zusammengetragen und strukturiert, aufbewahrt und aufgearbeitet wurde, ist heute mit Big Data einerseits unstrukturiert und andererseits auch auf verschiedenen Datenbanken verstreut und stammt von verschiedenen Datenquellen. Ein Warehouse wie bis anhin – d.h. eine zentrale Bearbeitung der Daten – ist mit Big Data oft als Cloud-basiertes Tool bzw. Cloud-basierten Dienst nicht mehr zwingend notwendig.

Bei Big Data werden i.d.R. anonymisierte Personendaten und/oder Sachdaten bearbeitet[18], die (noch) nicht als Personendaten i.S. des schweizerischen Bundesgesetz über den Datenschutz vom 19. Juni 1992 (SR 235.1, nachfolgend DSG) angesehen werden. Jedoch steigt aufgrund des zunehmenden Volumens der anonymen Daten und der Sachdaten sowie derer *Verknüpfung* bzw. Korrelation miteinander die Wahrscheinlichkeit einer Personenidentifikation, mithin einer Re-Individualisierung.[19] Dies wiederum bedeutet, dass die Akteure im Umfeld von Big Data nicht umhin kommen, sich früher oder später mit den datenschutzrechtlichen Vorgaben und der Bedeutung dieser Vorschriften im Kontext von Big Data auseinanderzusetzen (s. dazu auch unten Ziff. V).[20]

Das DSG und seine Ausführungsverordnung finden folglich immer dann Anwendung, wenn zumindest das Resultat einer Analyse ein Persönlichkeitsprofil ergibt

---

17 Vgl. http://www.google.com/landing/now/.

18 Heinrich Niggi Zittel, Stephan Fuhrer, Datenverknüpfungen im Versicherungswesen, 73.

19 Weber, Rolf H. (Fn. 4), Rz. 12 f.

20 M.w.H. Bruno Baeriswyl, in: Baeriswyl/Rudin/Hämmerli/Schweizer/Karjoth (Hrsg.), digma – Zeitschrift für Datenrecht und Informationssicherheit, «Big Data» ohne Datenschutz-Leitplanken, 2013.1, 14 ff.

oder die Person identifizierbar wird und wenn in der Analyse auch nur ein einziger Datensatz mit nicht anonymisierten Personendaten verwendet wird.

So sind insbesondere die *datenschutzrechtlichen Grundsätze* der Rechtmässigkeit, Zweckmässigkeit, Transparenz und Verhältnismässigkeit gemäss Art. 4 DSG zu berücksichtigen.

Die *Zweckmässigkeit* umfasst insbesondere die Frage, wofür Daten bearbeitet werden dürfen. Die Zweckbestimmung bleibt im Zusammenhang mit Big Data oftmals unbeachtet, indem Datensammlungen genutzt werden, ohne deren Zweckbestimmung zu berücksichtigen. Art. 4 Abs. 3 und 4 DSG halten den Grundsatz der Zweckgebundenheit einer Datenbearbeitung fest. Weil Datenverknüpfungen zu De-Anonymisierung von ursprünglich anonymisierten Daten bis hin zur Schaffung von Persönlichkeitsprofilen führen können – sei es nun bei Verknüpfung von Sachdaten mit Sachdaten oder bei Sachdaten mit Personendaten muss die ursprüngliche Zweckbestimmung der Datensammlung auch dann berücksichtigt werden, wenn es sich zuvor um anonymisierte Daten gehandelt hat. Dies gilt besonders bei Verknüpfung mit ortsbezogenen Daten.[21] Im Hinblick auf Big Data führt dies im Ergebnis zu einer Ausweitung des Zweckbindungsartikels von als eine rein auf Personendaten bezogene Bestimmung zu einer Norm, die sich auch auf ursprünglich als anonymisierte Daten eingestufte Sachdaten bezieht.[22] Unternehmen können sich deshalb nicht mehr darauf verlassen, dass sie nur reine Sachdaten bearbeiten und sich dabei ausserhalb der datenschutzrechtlichen Prinzipien des DSG befinden, sondern müssen dem Thema Datenschutz in ihrem Unternehmensprozess (noch) mehr Beachtung schenken: Es gilt, die Datenbestände im Hinblick auf die ursprüngliche Zweckbestimmung – sofern eine solche überhaupt vorhanden ist – im Auge zu behalten, in dem die Datensammlungen inventarisiert werden und u.a. auch den Bearbeitungszweck ergänzt werden. Noch weitergehend: Es kann unter Umständen verlangt werden, eine umfassendere vertragliche Zweckbestimmung für Sachdaten oder im Anschluss anonymisierte Personendaten vorzusehen.[23]

Dies führt dazu, dass auch die *Rechtmässigkeit* der Bearbeitung von anonymisierten Daten zu gewährleisten ist, mithin die Zustimmung der betroffenen Person für die Datenverknüpfung vorliegen muss respektive einzuholen ist.[24]

---

21 Thomas Probst (Fn. 8), 32 f.

22 Vgl. Thomas Probst (Fn. 8), 39.

23 Vgl. Astrid Epiney, in: Belser/Epiney/Waldmann, Datenschutzrecht, Grundlagen und öffentliches Recht, Bern 2011, §9 Allgemeine Grundsätze N35; bereits zum Thema Data Warehousing und Data Mining, den begrifflichen Vorläufern von Big Data wurden dazu Bedenken geäussert, vgl. m.w.H. Urs Maurer-Lambrou/Andrea Steiner, in: Maurer-Lambrou/Vogt (Hrsg.), BSK zum Datenschutzgesetz, 2. Aufl., Basel 2006, Art. 4 DSG N 14.

24 Vgl. Astrid Epiney (Fn. 23), N35.

Eine aktive Informationspolitik kann daher mithelfen, die geforderte *Transparenz* herzustellen und damit die Personendaten nach Treu und Glauben zu bearbeiten. Ein schönes Beispiel für eine positive und transparente Information der betroffenen Person findet sich bei TomTom, nachdem TomTom betreffend der Nutzung ihrer Daten in die Schlagzeilen kam.[25]

Auch das *Verhältnismässigkeitsprinzip* ist zu berücksichtigen, so dass ein geringstmöglicher Eingriff in die Rechte der betroffenen Person erfolgt. Daraus ergibt sich, dass nur diejenigen Daten gesammelt und bearbeitet werden dürfen, die für einen bestimmten Zweck objektiv tatsächlich benötigt werden und in einem vernünftigen Verhältnis zwischen Bearbeitungszweck und Persönlichkeitsrechtsverletzung stehen.[26] Diesem Grundsatz stehen oft regulatorische Vorgaben entgegen, so z.B. im Fernmeldebereich oder auch im Bankenbereich. Damit dürfen Personendaten nicht über den gesetzlich umschriebenen Umfang und die normierte Dauer aufbewahrt werden.[27] Der Ansatz von Big Data steht diesem Grundsatz diametral gegenüber. Bei Big Data ist man versucht, so viele Daten wie nur möglich zu beschaffen und aufzubewahren. Es stellt sich entsprechend die Frage, ob nicht auch ein Nutzereinverständnis über eine *verlängerte Aufbewahrungsfrist* notwendig wäre.[28]

2. Datensicherheit

a. Rechtsgrundlagen

Nebst den Grundsätzen gemäss Art. 4 DSG regeln das DSG sowie insbesondere die Verordnung zum Bundesgesetz über den Datenschutz vom 14. Juni 1993 (SR 235.11, nachfolgend VDSG) auch die Datensicherheit, die bei der Bearbeitung von Personendaten beachtet werden muss.

Die datenschutzrechtliche Datensicherheit ist in Art. 7 DSG sowie in den Artikeln 8–12 und 20–23 VDSG geregelt. Zu den in Art. 7 DSG erwähnten technischen und organisatorischen Massnahmen hat der Eidgenössische Datenschutz- und Öf-

---

25 Der von TomTom produzierte Webcast unter http://www.tomtom.com/en_gb/safeguarding-your-data/ wird dem Erwerber eines TomTom Verkehrsleitungssystem-Gerätes bei der Erstinstallation auf dem Gerät gezeigt und er hat die Möglichkeit seine Daten der «Community» zur Verfügung zu stellen oder eben auch nicht.

26 Urs Maurer-Lambrou / Andrea Steiner (Fn. 23), Art. 4 DSG N. 9.

27 M.w.H. Urs Maurer-Lambrou / Andrea Steiner (Fn. 23), N 11.

28 Nebst den allgemeinen Bedenken, vgl. Weber, Rolf H. (Fn. 4), RZ 25.

fentlichkeitsbeauftragter (EDÖB) einen Leitfaden namens «Technische und organisatorische Massnahmen» mit aktuellem Stand 2011 publiziert.[29]

Nebst den datenschutzrechtlichen Vorgaben sind insbesondere auch spezialgesetzliche Regelungen zu beachten, welche allenfalls dem DSG vorgehen.[30]

b. Definition der Datensicherheit

In der Lehre und Praxis wie auch im Gesetz wird derzeit *keine befriedigende Definition* bezüglich Datensicherheit im Kontext des DSG verwendet.[31] Eine genau umschriebene Bedeutung mit Berücksichtigung des gegenwärtigen Standes der Technik lässt sich auch im Gesetz selbst nicht finden.[32]

Die Datensicherheit i.S. des DSG hat den Schutz der Personendaten im Vordergrund und dient deshalb der *Verhinderung eines Schadens der betroffenen Person.* Dies ist m.E. insbesondere das Abgrenzungskriterium zur Informationssicherheit/IT-Security bzw. zum m.E. gleichwertigen Begriff des Informationsschutzes im technisch verstandenen Sinne (vgl. dazu Ziff. IV.2.c). Zwar ist die Datensicherheit im Sinne des DSG Bestandteil der *Informationssicherheit,* doch liegt bei letzterer der *Fokus auf sämtlichen Informationen,* wie Geschäfts- und Produktionsgeheimnissen, Dokumente, Bilder, Marken etc., und nicht nur auf denjenigen Informationen bzw. Daten, welche Personendaten beinhalten.[33]

Der Gesetzgeber hat ursprünglich die Datensicherheit im Rahmen der allgemeinen Grundsätze der Bearbeitung von Personendaten regeln wollen (Entwurf Art. 4 Abs. 6 DSG unter dem Stichwort Datensicherung[34]). Um eine bessere Gliederung zu erhalten, wurde die Datensicherheit im Rahmen eines eigenen Artikels abge-

---

29 Abrufbar unter http://goo.gl/M4huVh (Stand 8.2.2014).

30 So z.B. Fernmeldegesetz (SR 784.10, FMG), Verordnung über Fernmeldedienste (SR 784.101.1), Verordnung über Adressierungselemente im Fernmeldeverkehr (SR 784.104, AEFV), Bundesgesetz betreffend die Überwachung des Post- und Fernmeldeverkehrs (SR 780.1) und deren Verordnung über die Überwachung de sPost- und Fernmeldeverkehrs (SR 780.11, VÜPF).

31 Weder im Gesetz noch in der Verordnung wir detailliert beschrieben, was Datensicherheit ist, geschweige denn, welche Sicherheitsmassnahmen konkret zu ergreifen sind.

32 Gl. M. Kurt Pauli, in: Maurer-Lambrou/Vogt (Hrsg.), BSK zum Datenschutzgesetz, 2. Aufl., Basel 2006, Art. 7 DSG N 3.

33 Anders bei Kurt Pauli (Fn. 32), Art. 7 DSG N 2., welcher Datensicherheit m.E. falsch mit Informationsschutz gleichsetzt.

34 Botschaft zum Bundesgesetz über den Datenschutz (DSG) vom 23. März 1988, BBl 1988 II 413, 452.

handelt.[35] Dies erfolgte m.E. zu Recht, würde doch sonst der Datenschutz leerer Buchstabe bleiben. Zu beachten ist aber, dass die Datensicherheit damit eigentlich auch einen Stellenwert eines Grundsatzes des Datenschutzes geniesst.[36]

Mithin sind die Grundsätze der Datensicherheit gemäss Art. 7 DSG und Datenrichtigkeit gemäss Art. 5 Abs. 1 DSG eng verknüpft (vgl. dazu Ziff. III.2.e nachfolgend).[37]

c. Abgrenzung zur Informationssicherheit

Die Informationssicherheit aus der technischen Sicht hat die *Information als gesamter Wert im Fokus* und ist damit breiter als die Datensicherheit gemäss Art. 7 DSG. Die Informationssicherheit dient der Aufrechterhaltung der Vertraulichkeit, Integrität und Verfügbarkeit von Informationen.[38] Andere Eigenschaften wie Authentizität, Zurechenbarkeit, Nicht-Abstreitbarkeit und Verlässlichkeit können ebenfalls berücksichtigt werden. Mithin geht es also bei der Informationssicherheit um die Aufrechterhaltung des Betriebes, den Geschäftsgeheimnisschutz und schliesslich auch den Personendatenschutz. Damit ist also die Datensicherheit gemäss DSG nur ein Teilbereich der Informationssicherheit.

Entgegen Kurt Pauli ist Datensicherheit damit nicht gleich Informationsschutz bzw. Informationssicherheit im technischen Sinne.[39]

Im Rahmen der Informationssicherheit folgen Unternehmen den bereits bekannten technischen Standards wie ISO 27001:2013, COBIT, BSI-Grundschutzhandbuch etc. Es gibt derzeit keine Verpflichtung, diese Standards z.B. im Rahmen eines Information Security Management System (ISMS) einzuhalten. Es obliegt vielmehr dem einzelnen Unternehmen selbst, die Risiken zu evaluieren, die Eintretenswahrscheinlichkeit und das Schadenspotential sowie Massnahmen zur Verhinderung oder Mitigation von Risiken zu treffen.[40] Die Einhaltung dieser Standards sowie eine allfällige Zertifizierung verschafft dem Marktteilnehmer jedoch gegebenenfalls einen Wettbewerbsvorteil.

---

35 Vgl. Kurt Pauli (Fn. 32), Art. 7 DSG N 2; Amtl. Bull. 1990 II Geschäfts Nr. 88.032, S. 125-Ref. No. 20 018 586, zu Art. 4quater (neu), Wortmeldung Danioth, (<http://www.amtsdruckschriften.bar.admin.ch/detailView.do?id=20018587#1>) (Stand 8.2.2014).

36 Vgl. Astrid Epiney (Fn. 23), § 9.

37 Vgl. Astrid Epiney (Fn. 23), § 9 N 44.

38 Definition gemäss Ziff. 3.4 ISO 27001:2008. Vgl. dazu FN.

39 Kurt Pauli (Fn. 32), Art. 7 N. 2.

40 David Rosenthal/Yvonne Jöhri, Handkommentar zum Datenschutzgesetz sowie weiteren, ausgewählten Bestimmungen, 1, Zürich, Basel, Genf 2008, Art. 7 N 15.

d. Datensicherheit (Art. 7 DSG)

Den Ausgangspunkt der Datensicherheit nach Datenschutzgesetz bildet Art. 7 DSG, wonach *Personendaten* durch «*angemessene technische und organisatorische Massnahmen gegen unbefugtes Bearbeiten geschützt*» werden müssen.

Es muss sich deshalb vorab um *Personendaten i.S. von Art. 3 lit. a DSG* handeln. Im Rahmen von Big Data sind durchaus Anwendungen denkbar, bei denen weder die zugrundeliegenden Primärdaten noch das Resultat der Analyse als Personendaten zu qualifizieren sind. Dies ist z.B. der Fall bei Auswertung von Wetterdaten verknüpft mit topologischen Daten oder bei Auswertungen von Finanzdaten ohne Verknüpfung zu einem Benutzerprofil bzw. zu Benutzerangaben. Werden aber bei der Analyse i) Personendaten bearbeitet oder ii) führt das Resultat zu einem Persönlichkeitsprofil oder zu identifizierbaren Personen und damit zu Personendaten, so ist Art. 7 DSG anwendbar. Im ersten Fall i) bereits bei der Analyse, im zweiten Fall ii) beim Umgang mit dem Resultat (vgl. Ziff. III.1 hiervor).

Angemessen sind die getroffenen Sicherheitsmassnahmen, wenn sie *geeignet und verhältnismässig* sind.[41] So muss *nicht für alle Arten von Personendaten* der gleiche Sicherheitsmassstab beachtet werden. Besonders schützenswerte Personendaten z.B. geniessen ein höheres Schutzbedürfnis als normale Personendaten. In Betracht zu ziehen ist aber bei einem angemessenen Schutzniveau gemäss Art. 8 Abs. 1 VDSG insbesondere auch der Zweck der Datenbearbeitung (lit. a), die Art und der Umfang der Datenbearbeitung (lit. b), das Risiko-Assessment bei einer Personendatenverletzung für die betroffene Person (lit. c) sowie auch der gegenwärtige Stand der Technik (lit. d). Da eine Güterabwägung zu erfolgen hat, sind auch die Interessen des Datenbearbeiters zu beachten, namentlich das Verhältnis des Aufwandes für die Datensicherheit zum angestrebten Schutzzweck.[42] Letzteres dürfte aber in einem Rechtsfall jeweils rückwirkend anders betrachtet werden als im Voraus.[43] Damit braucht der Schutz nicht absolut zu sein, mithin ist kein maximaler Schutz gefordert.[44] Dies führt dazu, dass es für einige Unternehmen gelegentlich schwierig zu beurteilen ist, ob sie sich rechtskonform verhalten. Nicht zuletzt, da sie zu wenig Kenntnisse über den Stand der Technik besitzen und sich insbesondere auch nicht mit den Bedrohungslagen bzw. den Risiken sowie den anderen Kriterien zur Beurteilung der angemessenen Massnahmen (unter Berücksichtigung der Unternehmensgrösse) auskennen.

---

41 Vgl. Kurt Pauli (Fn. 32), Art. 7 DSG N 5.
42 David Rosenthal / Yvonne Jöhri (Fn. 40), Art. 7 N 3.
43 Kurt Pauli (Fn. 32), Art. 7 DSG N 3.
44 David Rosenthal / Yvonne Jöhri (Fn. 40), Art. 7 N 3.

Der *Datenbearbeiter* (und nicht nur der Dateninhaber![45]) hat folglich für den Schutz von Personendaten angemessene *technische und organisatorische Massnahmen* zu ergreifen. Mit der Erwähnung von technischen *wie auch* organisatorischen Massnahmen unterstreicht der Gesetzgeber, dass beide Massnahmekategorien notwendig sind, um einen vollständigen Schutz zu gewährleisten.

Das Gesetz *verzichtet darauf, die einzelnen Sicherheitsmassnahmen* in technischer und organisatorischer Hinsicht zu regeln. Der Gesetzgeber wollte die Definition bewusst den «Bearbeitern oder ihren Berufs- und Fachverbänden» überlassen, die für ihren Bereich nötigen Sicherheitsbedürfnisse zu definieren und die daraus folgenden Sicherheitsmassnahmen anzuordnen.[46]

Der Gesetzgeber befand, dass er im Bedarfsfall bzw. bei Schwierigkeiten, gestützt auf die Ausführungskompetenz gemäss Art. 30 Abs. 1 des Entwurfes des DSG die Details festlegen könne.[47] Im Gegensatz zum Entwurf, welcher insbesondere mit der Lesung der Botschaft nur ein Eingriff im Notfall vorsah, wurde dann in Art. 7 Abs. 2 DSG festgelegt, dass der Bundesrat die Mindestanforderungen regeln *muss.*

Unter die technischen und organisatorischen Massnahmen fallen insbesondere auch bauliche Massnahmen. Der Bundesrat wollte die Datenbearbeiter verpflichten, einen integralen Ansatz zu wählen.[48] Dies bedeutet, dass sämtliche sogenannten Layer des OSI-Modelles, des Referenzmodelles für Netzwerkprotokolle, als Schichtenarchitektur zu berücksichtigen sind, insbesondere auch der physische Layer wie die Rechenzentrumsinfrastruktur, Stecker und andere Übertragungs- und Verbindungsgeräte. Wenn nämlich das Fundament nicht sicher ist, kann auch ein darauf gebautes Gebäude nicht gerade stehen. In der Lehre wird vereinzelt postuliert, dass es keine solche Verpflichtung für ein umfassendes, ganzheitliches Sicherheitskonzept gäbe.[49] Ohne dass Datensicherheit bzw. Informationssicherheit ganzheitlich angewandt werden, sind punktuelle Massnahmen nicht zielführend, weshalb der Datenbearbeiter m.E. dann auch nicht seinen gesetzlichen Verpflichtungen nachkommt.[50]

---

45 Vgl. Astrid Epiney (Fn. 23), § 9 N 55.

46 BBl 1988 II 413, 452.

47 Gemäss Botschaft zum Bundesgesetz über den Datenschutz (DSG) vom 23. März 1988, vgl. Fn. 46.

48 Vgl. Fn. 46.

49 David Rosenthal / Yvonne Jöhri (Fn. 40), Art. 7 N 4.

50 Die meisten auf Informationssicherheit spezialisierten Berater empfehlen einen ganzheitlichen Ansatz, so z.B. swissInfosec (vgl. http://www.infosec.ch/consulting/integrale-sicherheit, (Stand 8.2.2014) oder auch InfoGuard (vgl. http://infoguard.ch/de/solutions (Stand 8.2.2014).

e. Präzisierung der Datensicherheit (VDSG)

Der Verpflichtung zur Präzisierung ist der Bundesrat nachgekommen und hat Ausführungsbestimmungen in den Art. 8–12 und 20–23 VDSG erlassen. Die Art. 8–12 VDSG gelten für Private aufgrund ihrer Systematik unter dem ersten Kapitel. Die Art. 20–23 VDSG gelten dagegen als Bestandteile des zweiten Kapitels nur für Bundesorgane. Sie verweisen auf die Art. 8–10 VDSG und treffen darüber hinaus weitere verwaltungsinterne Massnahmen. So müssen Bundesorgane z.B. bei der automatisierten Datenbearbeitung mit dem Informatik Strategieorgan des Bundes (ISB) zusammenarbeiten. Der vorliegende Beitrag konzentriert sich v.a. auf die Nutzung von Big Data durch Privatpersonen.

Auch diese Ausführungsbestimmungen in der VDSG lassen den notwendigen Präzisionsgehalt missen, wie nachfolgend aufzuzeigen ist.

Allgemeine Massnahmen (Art. 8 VDSG)

Die gesetzliche Konkretisierung dieser Massnahmen befindet sich in Art. 8 Abs. 1 VDSG, laut dem jene Privatperson, welche Personendaten bearbeitet oder ein Datenkommunikationsnetz zur Verfügung stellt, einen angemessenen Datenschutz gewährleisten muss, indem sie für die *Vertraulichkeit,* die *Verfügbarkeit* und die Integrität der Daten sorgt.

Unter *Vertraulichkeit* wird die Prämisse verstanden, dass Informationen unberechtigten Personen, Einheiten oder Prozessen nicht verfügbar gemacht oder enthüllt werden dürfen.[51]

Aufgrund der gesetzlich festgeschriebenen *Verfügbarkeit* sind einer berechtigten Einheit auf Verlangen Informationen zugänglich und nutzbar zu machen.[52]

Unter *Integrität* wird die Eigenschaft der Absicherung von Richtigkeit und Vollständigkeit von Werten (alles was für eine Organisation von Wert ist wie Daten, Informationen, Geschäftsgeheimnisse, Personendaten etc.) verstanden, mithin die Richtigkeit der behaupteten Tatsache.[53] Dies bedeutet, dass wenn in einer Datensammlung Angaben zu einer Person gemacht wurden, diese richtig sein müssen.[54]

---

[51] Mangels gesetzlicher Definition ist auf die technischen Standards abzustützen, mithin auf ISO 27001. Definition gemäss Ziff. 3.3 ISO 27001:2008. ISO 20071:2008 wurde kürzlich revidiert. Die neue ISO 2007:2013 verweist auf die Begriffsdefinitionen in ISO 2700.

[52] Vgl. dazu Fn. , Ziff. 3.2 ISO 27001:2008.

[53] Vgl. dazu FN, Ziff. 3.1 i.V.m. 3.8 ISO 27001:2008.; Vgl. Kurt Pauli (Fn. 32), Art. 7 DSG N 2, mit dem Hinweis, dass die Grundsätze vermehrt auch mit Verbindlichkeit und Authentizität ergänzt werden.

[54] Was diametral zu Big Data steht, vgl. Ziff. 1.3.

Insbesondere muss der Datenbearbeiter gemäss Art. 8 Abs. 1 lit. a-e VDSG die nicht vom Gesetzgeber spezifizierten Systeme *gegen folgende Risiken schützen:* i) unbefugte oder zufällige Vernichtung (z.B. durch Fehlmanipulation von Mitarbeitern), ii) zufälliger Verlust (z.B. durch Fehlmanipulation), iii) technische Fehler, iv) Fälschung, Diebstahl oder widerrechtliche Verwendung (z.B. durch Wirtschaftsspionage) und v) unbefugtes Ändern, Kopieren, Zugreifen oder andere unbefugte Bearbeitungen. Diese Vorgaben werden m.E. im Rahmen von Art. 9 Abs. 1 VDSG vollständig konsumiert bzw. haben über Art. 9 Abs. 1 VDSG keine hinausgehende eigenständige Bedeutung. Die detaillierte Betrachtung kann deshalb im Rahmen dieses Beitrages ausser Acht gelassen werden.

Nach welchen Massstäben bzw. Leitplanken die bereits in Art. 8 Abs. Abs. 1 und Art. 7 Abs. 1 DSG erwähnte *Angemessenheit* des Datenschutzes zu beurteilen ist, führt Art. 8 Abs. 2 in lit. a–d VDSG aus.[55] Die technischen und organisatorischen Massnahmen tragen spezifischen Kriterien Rechnung, nämlich dem Zweck der Datenbearbeitung (lit. a), der Art und dem Umfang der Datenbearbeitung (lit. b), der Einschätzung der möglichen Risiken für die betroffenen Personen (lit. c) und dem gegenwärtigen Stand der Technik (lit. d). Bezüglich der ersten drei Kriterien kann auf das unter lit. c hiervor Gesagte verwiesen werden.

Von besonderer Bedeutung ist der Stand der Technik. Der Gesetzgeber wollte damit zum Ausdruck bringen, dass die Risikoanalyse sowie die Implementierung allfälliger technischer und organisatorischer Massnahmen gegen diese Risiken eine Daueraufgabe und -pflicht sind.[56] Insbesondere sind die getroffenen Massnahmen periodisch zu überprüfen und dann auch entsprechend anzupassen. Schwerwiegende Sicherheitslücken sind selbstverständlich umgehend zu schliessen.

In organisatorischer Hinsicht sind insbesondere die folgenden im Gesetz nicht aufgeführten Massnahmen zu ergreifen:

Der Datenbearbeiter hat *Reglemente und Weisungen* an Mitarbeiter zu erlassen, *Verträge mit Datenschutzklausen* zu versehen, *Schulungen, Anleitungen und Handbücher* sowie *prozess- und serviceorientierte Vorgaben* an die Mitarbeiter zu machen.[57]

Besondere Massnahmen (Art. 9 ff. VDSG)

Nebst den allgemeinen Massnahmen gemäss Art. 8 VDSG sind besondere Massnahmen für die Datensicherheit von *Datensammlungen,* die *automatisiert* bear-

[55] Es handelt sich entsprechend um eine Präzisierung, nicht um eine Ausdehnung von Art. 7 Abs. 1 DSG, vgl. David Rosenthal / Yvonne Jöhri (Fn. 40), Art. 7 N 5.

[56] David Rosenthal / Yvonne Jöhri (Fn. 40), Art. 7 N 5.

[57] Vgl. David Rosenthal / Yvonne Jöhri (Fn. 40), Art. 7 N 9 mit konkretisierten Beispielen.

beitet werden (Art. 9 VDSG) und bei sensitiven Personendaten und Persönlichkeitsprofilen (Art. 10 VDSG), gefordert.

Dabei ist zu berücksichtigen, dass bei Art. 8 und 9 VDSG *Datensammlungen* im Sinne des DSG gemeint sind, mithin Personendatensammlungen, die so aufgebaut sind, dass die Daten nach betroffenen Personen erschliessbar sind (Art. 3 lit. g DSG). Werden folglich zuerst einmal Datensammlungen aufbewahrt, die nicht nach Personen durchsuchbar sind und auch sonst keinen Bezug zu Personen aufweisen, handelt es sich nicht um eine Datensammlung i.S. des DSG. Artikel 9 VDSG wäre für diesen Fall noch nicht anwendbar.

Automatisiert erfolgt eine Bearbeitung von Personendaten bei der Zuhilfenahme von Computern und Applikationen.

Bei Big Data *handelt es sich stets um automatisiert* bearbeitete Daten, weshalb Art. 9 VDSG stets anwendbar ist, wenn nur eine der verwendeten Datensammlungen Personendaten enthält bzw. wenn das Resultat einer Analyse zu Personendaten oder Persönlichkeitsprofilen führt.[58] Die Protokollierung hingegen gemäss Art. 10 VDSG ist nur dort zu beachten, wo sensitive Personendaten oder Persönlichkeitsprofile bearbeitet werden oder die Resultate von Analysen zu solchen führen.

Art. 9 VDSG führt für die automatisiert bearbeiteten Personendaten, welche in einer Datensammlung zusammengefasst sind, die folgenden Kontrollziele auf.

Zugriffskontrolle (Art. 9 Abs. 1 lit. g VDSG): Die wohl wichtigste technische Massnahme fordert, dass eine befugte Person nur auf jene Personendaten zugreifen kann, welche sie für die Erfüllung ihrer Aufgaben auch benötigt. Der Zugriff darf also nur auf einer Need-to-know-Basis und nur selektiv erfolgen.[59] Dazu gehören z.B. die folgenden Massnahmen: Zugriff auf Personendaten erst nach Authentifizierung mittels Passwort oder dualen Authentifizierungsmechanismen (Passwort + Token), Zugriff gestützt auf eine gewisse Rolle im Unternehmen, wie z.B. Finanzdaten nur durch die Finanzabteilung sowie Zugriff nur auf die jeweils der Rolle zugeordneten Daten entsprechend einer Datenklassifizierung. Mithin kann die freie Abfragemöglichkeit auch durch die Beschränkung der *SQL-Queries eingeschränkt* werden, d.h. indem Abfragen auf Datenbankstufe nicht erlaubt werden. Erlaubt wird i.d.R. nur der Zugriff auf applikatorischer Stufe, weil auf dieser die Zugriffskontrolle gewährleistet werden kann.

---

[58] Vgl. zu den nachfolgenden Ausführungen betreffend Kontrollziele: Kurt Pauli, in: Maurer-Lambrou / Vogt (Hrsg.), BSK zum Datenschutzgesetz, 2. Aufl., Basel 2006, Art. 7 DSG N 13.

[59] Kurt Pauli (Fn. 32), Art. 7 DSG N 13.

Zugangskontrolle (Art. 9 Abs. 1 lit. a VDSG): Der physische Zugang zum Gebäude, zum Rack, zum Cage und schliesslich auch zum Server oder einem Archivraum darf nur befugten Personen gestattet werden und muss kontrolliert und protokolliert werden. Eine solche Kontrolle wird i.d.R. sowohl technisch wie auch organisatorisch organisiert, mithin durch Eingangskontrollen mit Paches/Iris- oder Adernscan, Sichtkontrollen mit Vorlage eines Ausweises, Videoüberwachung mit Sensorkontrolle sowie der Erlaubnis, dass nur diejenigen Personen Zutritt in ein Rechenzentrum erhalten, die vorher vom Berechtigten Infrastrukturbetreiber (z.B. Cloud-Anbieter) befugt wurden.

Personendatenträgerkontrolle (Art. 9 Abs. 1 lit. b VDSG): Es gilt zu verunmöglichen, das Personendatenträger wie z.B. USB Stick, Mobile Phones und Tablets oder auch Datenbänder[60] sowie die sich im Ergebnis darauf befindenden Personendaten, nicht durch Unbefugte gelesen, verändert oder gelöscht werden können. Dies erreicht man mit entsprechender Harddisk-Verschlüsselung, Passwortschutz sowie mit Container-Ansätzen (z.B. bei Mobile Devices bzw. Bring-your-own-Device-Lösungen[61]). Bei der Nutzung von SaaS-Diensten ist diesem Thema besonders Beachtung zu schenken. Wer im Unternehmen soll auf die Daten in der Cloud Zugriff haben und wie werden diese Zugriffe verwaltet? Es muss sichergestellt werden, dass die Verwaltung der Zugriffe vom Unternehmen aus erfolgen kann oder zumindest von einem Dritten und nicht dem Cloud-Anbieter, also z.B. einem Business-Analytics-Anbieter auf der Grundlage eines SaaS-Dienstes.

Transportkontrolle (Art. 9 Abs. 1 lit. c VDSG): Bei der Bekanntgabe und beim Transport von Personendaten gibt der Dateninhaber die Kontrolle über die Daten aus der Hand. Dennoch gilt es auch hier die Vertraulichkeit, die Verfügbarkeit und die Integrität der Daten zu garantieren, indem die Daten während dem Transport verschlüsselt werden oder ein verschlüsselter Kommunikationskanal eröffnet wird (z.B. ein Virtual Privat Network, VPN). Bei der Verschlüsselung ist auf eine starke Verschlüsselung zu achten.[62]

Bekanntgabekontrolle (Art. 9 Abs. 1 lit. d VDSG): Hierbei muss die Identifikation von Datenempfängern sichergestellt werden. Dies kann durch entsprechende Sperrungen von Ports und der Authentifizierung der Empfänger gewährleistet wer-

---

60 Z.B. der Diebstahl von Datenbändern bei der Swisscom anlässlich deren Vernichtung, vgl. NZZ vom 18. September 2013, Datenleck bei der Swisscom, (<http://goo.gl/e7sPMG>) (Stand 8.2.2014).

61 Bring your own device (BYOD) aus rechtlicher Sicht, in: Jusletter IT 12. September 2012.

62 Z.B. durch Installation eines PGP Clients und individuellem Key Austausch (vgl. http://www.gpg4win.org, Stand 8.2.2014, durch Verwendung von x509v3 Zertifikaten im nativen Mailclient vgl. http://goo.gl/Za5vvC, 8.2.2014), durch Cloud based SeppMail, bekannt als Incamail der Schweizerischen Post, vgl. http://goo.gl/xEhWvD, Stand 8.2.2014) oder durch SeppMail (vgl. www.seppmail.ch, Stand 8.2.2014).

den. Problematisch sind jedoch Malware (wie Spyware) und Viren, welche durch zwangsläufig offene Ports auf Endgeräte gelangen und damit bestehende Bekanntgabekontrollen umgehen können. Bei Malware ist v.a. auf die Sensibilisierung der Mitarbeiter hinzuwirken. Sogenannte Data-Loss-Prevention-Massnahmen filtern den ausgehenden Datenverkehr bzw. unterbinden diesen, z.B. indem keine Daten auf Mobile Devices gespeichert werden können oder grössere Datenvolumen gesperrt werden. Die Authentifizierung des Empfängers kann durch Zugriffskontrollen sowie durch Zertifikate (z.B. SSL) gewährleistet werden. In diesem Zusammenhang ist die Protokollierung der Zugriffe und der Datenflüsse notwendig.

Speicherkontrolle (Art. 9 Abs. 1 lit. e VDSG): Diese Vorgabe verlangt, dass Massnahmen getroffen werden, die verhindern, dass Personendaten auf dem Datenträger eingegeben, gesichtet, verändert oder v.a. gelöscht werden.

Benutzerkontrolle (Art. 9 Abs. 1 lit. f VDSG): Die Benutzerkontrolle verhindert, dass unbefugte Personen die automatisierten Datenverarbeitungssysteme anwenden können und so Personendaten bearbeiten. Damit sind v.a. Intrusion-Massnahmen zu ergreifen, d.h. Massnahmen, die verhindern, dass Dritte nicht z.B. mittels einer Cyberattacke in die Systeme des Dateninhabers eindringen und Personendaten entwenden können.[63]

Eingabekontrolle: Hierbei wird eine retrospektive Kontrolle der getätigten Änderungen an den Personendaten verlangt, womit letzten Endes ein einwandfreies Monitoring und eine lückenlose, revisionsfähige Protokollierung von Eingaben auf Datenbankebene einhergehen.[64] Dies trägt wiederum zur Integrität der Daten bei. Es muss deshalb eruiert werden können, wer an den Daten etwas verändert hat.

Nebst den besonderen Massnahmen für Datensammlungen i.A., die automatisiert bearbeitet werden, sind noch die *weitergehenden Protokollierungspflichten bei besonders schützenswerten Personendaten und Persönlichkeitsprofilen* gemäss Art. 10 VDSG zu beachten, wenn präventive Massnahmen nicht greifen. Das Protokoll dient dem Nachweis, dass die Daten tatsächlich nur für denjenigen Zweck bearbeitet werden, für den sie erhoben wurden.[65] Ein solches Protokoll (auch History genannt) muss beinhalten, wer (Benutzer), wann (Zeitangabe), wie (inhaltliche Änderung) und womit (welches Device) Daten bearbeitet hat.

Darüber hinaus hat der private Inhaber einer meldepflichtigen automatisierten Datensammlung ein *Bearbeitungsreglement* (Art. 11 VDSG) zu erstellen. Das Bearbeitungsreglement beschreibt die interne Organisation der Datenbearbeitungspro-

---

63 Z.B. der Datenklau bei Target, dem zweitgrössten Retailer der USA anfangs 2014, (<http://goo.gl/GQcG9S>) (Stand 8.2.2014).

64 Kurt Pauli (Fn. 32), Art. 7 DSG N 13.

65 Kurt Pauli (Fn. 32), Art. 7 DSG N 15.

zesse und -kontrollen.[66] Es betrifft nur diejenigen Dateninhaber, die regelmässig besonders schützenswerte Personendaten oder Persönlichkeitsprofile bearbeiten oder regelmässig Personendaten an Dritte bekannt geben und die nicht aufgrund von Art. 11a Abs. 5 Buchstaben b-d DSG von der Meldepflicht ausgenommen sind. Dies trifft damit all diejenigen Dateninhaber, die für Big Data ausländische Anbieter von Analyse-Tools mit Cloud-basierten Angeboten nutzen (Art. 6 DSG, so wie nachfolgend Ziff. IV.2.b).

Das Bearbeitungsreglement beschreibt die interne Organisation sowie das Datenbearbeitungs- und Kontrollverfahren und enthält die Unterlagen über die Planung, die Realisierung und den Betrieb der Datensammlung und der Informatikmittel. Insbesondere ist damit auch ein Dateninventar zu erstellen, welches den Zweck der Personendatenbearbeitung sowie die Zustimmungserklärung und deren Umfang umfasst.[67]

3. Datensicherheit gemäss sektorspezifischen Vorgaben

Für die Datensicherheit sind zudem die sektorspezifischen Vorgaben zu berücksichtigen. So gibt es z.B. weitergehende Bestimmungen für die öffentliche Hand[68], die Sozialversicherungs- und Privatversicherungsträger, die Banken[69], die Fernmeldedienstanbieter[70] sowie auch für Träger von Berufs- und Geschäftsgeheimnissen.[71] Im Rahmen dieses Beitrages sollen nur die grundsätzlichen Datensicherheitsvorgaben erläutert werden, damit für die nachfolgende Analyse der Datensicherheitsthemen im Rahmen von Big Data eine Grundlage geschaffen werden kann.

---

66 Kurt Pauli (Fn. 32), Art. 7 DSG N 16.

67 Für die spezifische Berücksichtigung von Big Data Risiken im Rahmen eines Bearbeitungsreglements vgl. Jon Neiditz, Baking your big data information governance programm; in celebration of 2013, in: Lexology 2013, (<http://goo.gl/HYpQ6D>) (Stand 8.2.2014).

68 Z.B. mit der Verordnung über den Schutz von Informationen des Bundes (SR 510.411, Informationsschutzverordnung, ISchV).

69 Z.B. mit dem (FINMA Rundschreiben FINMA Outsourcing für Banken- und Finanzbranche 2008/7, (http://goo.gl/GWCSds) (Stand 8.2.2014).

70 Vgl. Fn. 30.

71 Z.B. für den Arzt, EDÖB Erläuterungen zum Datenschutz in der Arztpraxis, (<http://goo.gl/IwWL8h>) (Stand 8.2.2014).

4. Datenbekanntgabe ins Ausland

Aufgrund der notwendigen hohen Rechen- und Speicherkapazität dürften bei Business-Analytics-Anwendungen auch öfter Cloud-Dienste zur Anwendung kommen. Damit sind auch die bei der Nutzung von Cloud-Computing-Diensten relevanten datenschutzrechtlichen Überlegungen im Allgemeinen und auch im Hinblick auf die Datensicherheit zu beachten, dies unabhängig davon, wo sich der Cloud-Anbieter befindet.[72]

Werden nun (Personen-)Daten zu einem Cloud-Diensteanbieter (z.B. einem Business-Analytics-Anbieter) im Rahmen des grenzüberschreitenden Datenverkehrs ins Ausland transferiert respektive erfolgt die Bearbeitung von (Personen-)Daten auf/über im Ausland gelegene Server, so sind die Bestimmungen zur grenzüberschreitenden Datenbekanntgabe gemäss Art. 6 DSG zu berücksichtigen.

Das schweizerische Datenschutzgesetz knüpft die Datenbekanntgabe bei Personendaten aus der Schweiz ins Ausland an besondere Voraussetzungen (Art. 6 DSG), so dass die übermittelten Personendaten angemessen geschützt bleiben müssen. Im Hinblick auf die ausländische, im Einzelfall anwendbare Rechtsordnung hat der EDÖB eine unverbindliche Liste erstellt, welche die jeweilige Datenschutzgesetzgebung der Staaten untersucht und beurteilt, ob im spezifischen Staat ein angemessener Datenschutz besteht.[73]

Der EDÖB kategorisiert diesen sogenannten Drittstaatendatenschutz in drei Kategorien: i) angemessener Schutz für die natürlichen Personen, ii) angemessener Schutz unter bestimmten Voraussetzungen und iii) ungenügender Schutz.[74] Die Staaten der EU fallen alle unter die Kategorie i). Unter die Kategorie ii) fällt nur die USA. Hier besteht ein Schutz, wenn ein Datenbearbeiter eine sogenannte Safe Harbour-Deklaration für die Bearbeitung von Personendaten, welche aus der Schweiz stammen, abgegeben hat und auf der Liste des U.S. Department of Commerce verzeichnet ist.[75] In diesem Falle soll im gesetzlichen Sinne ein angemessener Schutz im Sinne von Art. 6 Abs. 1 DSG gewährleistet sein. M.E. ist diese Selbstdeklaration aber mit Vorsicht zu geniessen. Oftmals betrifft sie nur

---

[72] Carmen De la Cruz, Cloud Computing: Alter Wein in neuen Schläuchen, in: Jusletter 15. Mai 2013.

[73] EDÖB, Stand des Datenschutzes weltweit, vom 18. Januar 2013, (<http://goo.gl/jpGl2v>) (Stand 8.2.2014).

[74] In der Schweiz sind nebst den natürlichen Personen immer auch die «Personendaten» von juristischen Personen geschützt. Diese Konzept ist im Ausland i.d.R. nicht bekannt, weshalb die Auflistung des EDÖB dies unberücksichtigt lässt.

[75] Federal Trade Commission FTC 600 Pennsylvania Avenue, NW Washington, DC 20580 www.export.gov/safeharbor/.

den Schutz der Personendaten des Kunden, nicht aber den Schutz der Kundendaten, welche ebenfalls Personendaten darstellen können.

Datenübermittlungen ins Ausland in Staaten ohne angemessenen Datenschutz, wie insbesondere die Kategorie iii), müssen dem EDÖB gemeldet werden. Es bedarf dann eines Abschlusses eines Vertrages, über dessen Bestimmungen der EDÖB zu informieren ist. Wer sich dabei an die EU-Standardvertragsklauseln für die Übermittlung personenbezogener Daten in Drittländer hält, kann unbesorgt von der rechtmässigen Personendatenbearbeitung ausgehen.[76]

Nebst der Datenkategorisierung bzw. selbst bei anonymisierten Daten, welche zur Analyse verwendet werden, muss deshalb berücksichtigt werden, dass eine solche Meldung an den EDÖB vorliegt.

IV. Datenschutzrisiken von Big Data

1. Im Allgemeinen

Big Data führt uns folglich zu neuen Rechtsfragen, welche bis anhin nicht oder nicht so bedeutend waren. So werden sämtliche Grundsätze des Datenschutzgesetzes tangiert.

Big Data birgt das Risiko der *Zweitverwendung* von bestehenden Datensammlungen im Unternehmen, ohne dass eine Zustimmung der betroffenen Person oder eine andere Grundlage für die *rechtmässige Bearbeitung* gegeben wäre.

Diesem Risiko kann nur entgegengewirkt werden, indem sich der öffentliche und private Sektor eine gewisse Disziplin aneignet und Zustimmungserklärungen bzw. gesetzliche Grundlagen minutiös erfasst und vom Kunden entsprechende Zustimmungen einholt und die Datensammlungen im Dateninventar damit ergänzt. Damit kann transparent gemacht werden, ob eine Datensammlung für eine Zweitverwendung zur Verfügung steht oder nicht.

Grundsätzlich problematisch ist das Thema der *Zustimmung der betroffenen Person.* Die betroffene Person muss umfassend und transparent informiert worden sein, was mit ihren Daten geschieht. Die Formulierung, «die Daten dürfen zu Analysen verwendet werden», dürfte dazu nicht genügen (vgl. dazu den einlässlichen Beitrag von Prof. Dr. Florent Thouvenin in diesem Tagungsband). Die Balance zwischen der Wahrung der unternehmerischen Freiheit und der Informationspflicht gegenüber den betroffenen Personen wird damit deutlich anspruchsvoller.

[76] Vgl. <http://goo.gl/1tyNZF> (Stand 8.2.2014).

Ein weiteres spezifisches Thema von Big Data ist die Möglichkeit der *De-Anonymisierung bzw. der Re-Identifizierung von Personen* trotz Verwendung von anonymisierten bzw. pseudonymisierten Daten aufgrund der *Datenverknüpfung.*[77] Dies ist v.a. ein Thema bei der Forschung, welche gehalten ist, bei Austausch von Daten entsprechende Vereinbarungen abzuschliessen, dass keine re-identifizierten Daten öffentlich zur Verfügung gestellt werden dürfen.

Des Weiteren kommt das Risiko hinzu, dass grosse Unternehmen, welche Big Data verwenden und Persönlichkeitsprofile ihrer Kunden anlegen, interessanter für Datendiebe werden. Sie müssen nur schon deshalb ihre Sicherheitsvorkehrungen, in technischer, organisatorischer aber auch in rechtlicher[78] Hinsicht erhöhen.

Werden für Analysen Daten in die Cloud oder zu einem Business Intelligence Agent oder Business Analytics übertragen, so sind auch bei Big Data die Datensicherheitsbestimmungen durch diesen zu erfüllen. Der Dateninhaber hat sicherzustellen, dass sein *Datenbearbeiter* die Datensicherheit gewährleistet (Art. 10a Abs. 2 DSG i.V.m. Art. 7 DSG). Dieser kann dies aber auch nur im Rahmen der technischen Gegebenheiten garantieren (siehe dazu nachfolgend Ziff. 2).

2. Spezifische Datensicherheitsrisiken bei Big Data

a. Überblick

Aus der Big-Data-Thematik ergeben sich spezifische Probleme der Gewährleistung der Datensicherheit insbesondere auch im datenschutzrechtlichen Sinne.

Die Risiken lassen sich in folgende Gruppen aufteilen:[79]

a) Risiken, welche die Sicherheit der Infrastruktur und Zugriffskontrolle (vgl. Ziff. IV.2.b) betreffen, b) Risiken beim Daten-Management (Ziff. IV.2.c) betreffend sicherer Datenspeicherungs- und Transportprotokolle sowie granularer Audits und c) Risiken der Datenintegrität und reaktiven Sicherheitsmassnahmen (Ziff. IV.2.d) im Zusammenhang mit Eingabekontrolle – und Protokollierung sowie Echtzeit-Monitoring.

---

77 Vgl. dazu einlässlich Thomas Probst (Fn. 8).

78 Z.B. durch geeignete IT-Governance und Schulung der eigenen Mitarbeiter, durch hohe Konventionalstrafen für den Datenbearbeiter bei Verletzung der gesetzlichen und vertraglichen Datenschutzbestimmungen gegenüber dem Dateninhaber.

79 In Anlehnung an den Bericht der Big Data Working Group der Cloud Security Alliance, Expanded Top Ten Big Data Security and Privacy Challenges, April 2013, (<http://goo.gl/zsDoJL>) (Stand 8.2.2014).

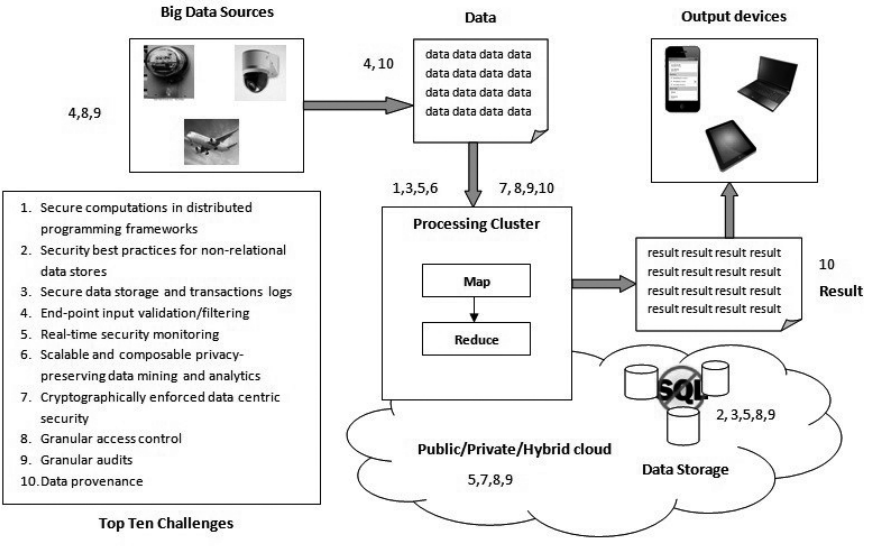

Fig. 1: Big-Data-Sicherheitsrisiken, Quelle: vgl. Fn. 79.

b. Sicherheit der Infrastruktur

Datenintegrität bei verteilten Programm-Strukturen

Wie wir gesehen haben, ist im Rahmen der Datensicherheit gemäss DSG die Datenintegrität eine Anforderung der Datensicherheit. Datenbanktechnisch wurde dies bis anhin mit SQL Datenbanken, d.h. strukturierten relationalen Datenbanken abgebildet bzw. durch entsprechende Applikationen oder mit anderen systemspezifischen Mechanismen. Bei unstrukturierten NoSQL Datenbanken, welche Hadoop nutzen sowie aufgrund der Menge der Datengrundlagen ist die Integrität der Daten für Big Data aber gerade kein notwendiges Kriterium, wie das vorher erwähnte Beispiel von Google Translate (vgl. Fn. 14) eindrücklich veranschaulicht.

Wird aber für die Business Analytics mit Personendaten gearbeitet, so dass man schlussendlich auf die einzelne Person schliessen kann oder zuverlässig können muss, so besteht das Risiko, dass man Personen etwas zuordnet, was allenfalls infolge mangelnder Datenintegrität eines Datensatzes zu falschen Resultaten führt. Entsprechend ist die Datenintegrität eigentlich bei Big-Data-Analysen sicherzustellen, damit Personen nicht falscher Tatsachen beschuldigt werden.

Technisch ist dieses Problem alles andere als trivial. So werden bei verteilten Programmierungsstrukturen, wie sie durch das MapReduce angewandt werden, par-

allele Rechen- und Speicherprozesse gestartet, um die Datenmengen zu bewältigen (Vgl. Ziff. II.1.b). Dabei werden die Rechenjobs in kleine Datenpakete bzw. Jobs zerteilt und einem Rechenknotenpunkt zur Verarbeitung zugeteilt. Nach deren Verarbeitung werden die Datenpakete wieder zusammengefügt.

Daraus können zwei Risiken entstehen: Zum einen, dass die sogenannten Mapper, welche jedem kleinen Datenpaket «angehängt» werden, nicht korrekt und vertrauenswürdig sind. Zum anderen ergibt sich das Risiko, dass sich nicht vertrauenswürdige Mapper einschleichen, welche schwierig zu identifizieren und zu beseitigen sind. Nicht vertrauenswürdige Mapper können v.a. im Rahmen von sogenannten Man-In-The-Middle-Attacken auf dem Transportweg eingeschleust werden. Man-In-The-Middle-Attacken zu entdecken ist dann offenbar beinahe unmöglich. Ausserdem werden ganz allgemein DDos-Attacken verwendet, die bewirken, dass die Mapper nicht richtig übermittelt werden können. Damit ist auch die Transportkontrolle schwierig bis unmöglich zu gewährleisten.

Beide Risiken können dazu führen, dass das Resultat korrumpiert wird, womit die Datenintegrität nicht gewährleistet werden kann. Dies kann je nach Kontext der Datenbearbeitung verheerend sein.

Zugriffskontrolle

Bei den bis anhin eingesetzten SQL-Datenbanken wurden Sicherheitsvorkehrungen im Datenbank-Managementsystem und in der Middleware (d.h. zwischen Datenbank und Applikation) realisiert. Bei im Rahmen von Big Data verwendeten noSQL-Datenbanken werden Sicherheits-Anwendungen derzeit noch nicht auf Datenbankstufe ermöglicht oder werden nur mit schwachen Sicherheitsvorkehrungen ausgeliefert. Dies ist deshalb so, weil es sich bei noSQL noch um relativ neue Technologie handelt. Damit ist die Robustheit von Sicherheitsvorkehrungen nicht gewährleistet, insbesondere nicht die Zugriffskontrolle.

Will z.B. die Marketingabteilung eine Big-Data-Analyse machen, greift sie direkt d.h. via die Hintertüre – unter Umgehung von Identity Access Management – auf die NoSQL-Datenbank zu. Der Zugrifft kann i.d.R. nicht protokolliert werden und erfolgt damit unter Umgehung der ansonsten bei SQL-Datenbanken auf Middleware bzw. applikatorischer Stufe gewährleisteten Sicherheit durch Identity Access Management.

Die geringen Authentifizierungsmechanismen von NoSQL bzw. der von NoSQL verwendeten Kommunikationsprotokolle (XML und REST) machen damit die Datenbank anfällig für Cybercrime-Attacken.

Dateninhaber sind somit angehalten, ihre bestehenden (Middleware)-Sicherheitskonzepte zu überdenken und **neue Sicherheitsmechanismen für NoSQL-Datenbanken** zu implementieren, ohne deren Funktionstüchtigkeit einzuschränken.

Dabei müssen sie sowohl das Szenario eines internen wie auch externen nicht berechtigten Zugriffes berücksichtigen. Grundsätzlich kann dabei aber die Schwierigkeit bestehen die Zugriffskontrolle für die internen und externen Zugriffswege vollumfänglich zu gewährleisten. Diese Problematik ist aber jedem komplexen System inhärent.

NoSQL wurde ohne Fokus auf Sicherheitsaspekte entwickelt. Dies rächt sich heute, denn die beiden Vorteile von NoSQL, nämlich Performance und Skalierbarkeit, erweisen sich als die grössten Sicherheitsrisiken.

Intern können die Risiken, wenn auch nicht technisch, doch wenigstens mit organisatorischen Massnahmen mitigiert werden, indem klare Vorgaben erlassen werden, wer Zugriff auf die Datenbanken oder auf die Business Analytics Tools haben darf. Eine Vorgabe, dass «Privacy by Design»[80] und «Privacy Impact Assessment»[81] verfolgt werden soll, reicht m.E. hier nicht aus, da dies technisch aufgrund der aufgezeigten Zugriffs-Problematik nicht umsetzbar ist.[82]

Vielmehr braucht es eine granulare Zugriffskontrolle, d.h. eine Zugriffskontrolle, die Nutzern Zugang zu denjenigen Daten gewährt, die aufgrund ihrer Rolle, ihrer Zugriffsart (Mobile Device oder nicht) und -berechtigung sowie aufgrund der entsprechenden Datenklassifikation der Information notwendig ist, welche eben auch Business Analytics Tools umfassen muss sowie die daraus gewonnenen Resultate. Dies erfordert eine seriös implementierte Daten-Policy sowie deren Umsetzung, welche aber in der Praxis oft fehlt.

c. Daten-Management

Sichere Datenspeicherung und Transportprotokolle

Heute werden oftmals Datenverarbeitung, Datenbanken, Applikationen etc. logisch voneinander getrennt und damit in unterschiedlichen Netzwerkzonen platziert. Diese sogenannte Multi-Tier-Architektur erfordert bei Big Data aufgrund der grossen Datenmengen eine automatisierte Lösung zur Verschiebung von Daten. Sie können nicht mehr wie bis anhin manuell verschoben werden. Damit entgleitet dem IT-Manager die Kontrolle darüber, welche Daten in welches Tier verschoben werden. Diese automatisierten sogenannten Auto-Tiering-Lösungen protokollieren zudem nicht restlos, wohin Daten abgespeichert werden.

---

80 «Privacy by Design» verfolgt das Konzept, dass Datenschutzthemen bereits bei der Architektur von Systemen und Prozessen zu berücksichtigen ist.

81 «Privacy Impact Assessment» verfolgt das Konzept, dass v.a. auf die Auswirkungen der Persönlichkeitsrechtsverletzung zu achten ist.

82 Wie dies z.B. die Richtlinie Big Data, der Datenschutzstelle des Fürstentum Liechtensteins in Ziff. 6.2 fordert, (<http://www.llv.li/pdf-llv-dss-richtlinie-big-data.pdf>), (Stand 8.2.2014).

Dies kann dazu führen, dass kritische Informationen wie Personendaten oder Forschungsergebnisse / Geschäftsgeheimnisse in ein Tier verschoben werden, welches über keine genügenden Datensicherheitsvorkehrungen verfügt. Die daraus konkret resultierenden Risiken sind die Beeinträchtigung der Verfügbarkeit und der Datenkonsistenz, die genauen Bestimmung der Quelle der Daten[83] sowie die Möglichkeiten spezifischer Cyber-Attacken.

Als Lösung dieser Probleme kann die Verschlüsselung der Daten sowie des Transportes der Daten bis zu einem gewissen Grad weiterhelfen. Die Verschlüsselung der Daten greift aber für die Absicherung der Datenintegrität zu kurz, da zumindest auch verschlüsselte Daten ungeeignet verschoben oder unbefugt gelöscht werden können. Eine verschlüsselte Bearbeitung der Daten durch Dritte ist zum jetzigen Zeitpunkt noch nicht möglich.[84].

Die bestehenden Lösungsansätze wie die erwähnte Verschlüsselung sowie weitere Lösungen sind aber alles individuelle Lösungsansätze. Derzeit besteht weder ein ganzheitlicher Lösungsansatz noch Standards. Dies gilt umso mehr, als Sicherheitskonzepte für die verschiedenen Ebenen (Tiers) nicht aufeinander abgestimmt sind. Die Ausbalancierung der verschiedenen Zielkonflikte zwischen Sicherheit, Benutzerfreundlichkeit, Komplexität und Kosten muss erst noch in ein ganzheitliches Sicherheitskonzept einfliessen.

Granulare Audits

Wird Big Data bzw. Analytics für das Monitoring von Security Vorfällen verwendet, so sind für die Analyse nicht entdeckter Attacken detaillierte Audit-Informationen notwendig.[85] Die Audit-Informationen müssen vollständig und zugänglich sein. Ihre Datenintegrität muss gewährleistet und der unautorisierte Zugriff muss ausgeschlossen werden können.

Damit zeigt sich, dass auch Big-Data-Infrastrukturen auditierbar sein müssen. Die sogenannte Chain of Custody wird damit erweitert d.h. der Verantwortungsbereich in Punkte IT-Sicherheit, welcher durch IT-Forensik überprüfbar sein muss. Dazu müssen bei den verschiedenen Komponenten der Big-Data-Infrastruktur Daten geloggt werden können.

---

83 Was problematisch ist, weil dann weder die Quelle, noch die Zweckbestimmung oder die Nutzerzustimmung im Dateninventar erfasst werden kann.

84 Vgl. dazu (<http://goo.gl/AUTgNM>) (Stand 8.2.2014).

85 Z.B. auf Grund von Sarbenes-Oxley oder für eDiscovery-Zwecke.

d. Datenintegrität und Reaktive Sicherheitsmassnahmen

Eingabekontrolle – und protokollierung

Im Rahmen von verschiedenen Sicherheitsmonitoring-Systemen, wie z.B. bei der Data Leackage Prevention, werden heute viele Transaktions- und Zugriffsdaten von Endgeräten protokolliert. Dabei stellt sich die Herausforderung, inwiefern man diesen Daten effektiv trauen kann oder ob sie unter Umständen manipuliert sind.[86]

Denkbar wäre, dass die Transaktions- und Zugriffsdaten von einem Angreifer kompromittiert werden, dass ID-Spoofing-Attacken[87] oder Man-in-the-Middle-Attacken ausgeführt werden.

Um das Risiko falscher Input-Daten zu minimieren, gibt es zwei Möglichkeiten: präventiv, indem sichergestellt wird, dass kein falscher Input erfolgen kann und reaktiv, indem falscher Input ausgefiltert wird. Ersteres kann abgesichert werden, indem mit Zugriffskontrollen und Zertifikaten gearbeitet wird. Zertifikate stellen allerdings nur die Authentifizierungsproblematik sicher. Wenn anschliessend vom Device trotz Zertifikats falsche Daten gesendet werden, wird dies durch ein Zertifikat nicht verhindert. Zweiteres, die Filterung von falschem Input hingegen, ist bei grossen Datenmengen problematisch. Der falsche Input wird durch bestehende Monitoring-System gegebenenfalls nicht sichtbar, weil sich diese auf aussergewöhnliches fokussieren.

Echtzeit-Monitoring

Echtzeit-Monitoring kann im Kontext von Big Data für zwei Anwendungsfälle eingesetzt werden: Einerseits um die Big-Data-Infrastruktur selbst zu überwachen, indem geprüft wird, ob alle Rechenknoten[88] korrekt funktionieren und andererseits um ein Echtzeit-Monitoring im Rahmen von Business Analytics Tools[89] zu verwenden.

Echtzeit-Monitoring ist eine Herausforderung aufgrund der Masse bei Big Data. Aufgrund der grossen Datenmenge kann dieses Monitoring je nach angewandter Regel zu einer grossen Anzahl an falsch-positiven Ergebnissen führen. Je grösser diese Anzahl ist, um so mehr bereitet deren manueller Abarbeitung Mühe. Dennoch ergeben sich aus Anomalien-Analysen für die Monitoring-Anwender Vorteile, solange die Sicherheit der Resultate gewährleistet werden kann.

---

86 V.a. im Zusammenhang mit Bring your own device (BYOD).

87 Bei ID-Spoofing-Attacken werden falsche Identitäten verwendet, die dann falsche Daten liefern.

88 Vgl. zur Funktionsweise der Nodes hiervor unter Ziff. 2.2.1.

89 Z.B. um Kreditkartenmissbrauch in Echtzeit aufzudecken (Fraud Detection).

Dies kann nur durch ein Zusammenspiel verschiedener Sicherheits-Aspekte erreicht werden, nämlich indem gewährleistet wird, dass z.B. die Public Cloud welche für die Berechnungen verwendet werden sicher ist, dass das Hadoop Cluster und die Monitoring-Applications sicher ausgestaltet sind sowie eine Eingabekontrolle und Protokollierung (Ziff. IV.2.d hiervor) der verwendeten Daten gewährleistet werden kann.

V. De lege ferenda

Da durch den Einsatz von Big Data auch Personendaten in Datenanalysen miteinbezogen werden können bzw. durch Verknüpfung Sachdaten auch wieder infolge der Re-Individualisierung zu personenbezogene Informationen führen, stellt sich die Frage, inwiefern unsere heutige Datenschutzgesetzgebung mit dem Paradigmenwechsel Big Data zu kooperieren vermag.

So macht im Lichte von Big Data die Unterscheidung in Personen- und Sachdaten keinen Sinn mehr, wenn mit wenigen Sachdaten bereits auf Personen geschlossen werden kann.[90]

Könnte denn die Abkehr vom Grundsatz der Bearbeitung hin zur Frage nach dem legitimen Einsatz-Zweck zielführend sein? Dabei würde berücksichtigt zu was die Personendaten bearbeitet werden und geprüft, ob der Zweck dazu legitim ist. Eine Wertung würde damit in den Vordergrund gerückt. Zwar hat der Einsatz von Big Data eine Effizienzsteigerung zum Zweck, doch nicht immer ist dies nur im Interesse eines einzelnen Unternehmens – Dritte wie Kunde oder der Staat können durchaus auch ein Interesse an der Effizienzsteigerung haben. Effizienzsteigerung heisst schlussendlich immer, dass vorhandene Ressourcen besser zur Zielerreichung eingesetzt werden, wie z.B. einer Kostenreduktion oder einer Gewinnsteigerung. Es bleibt ein Unbehagen. Dieses Unbehagen gründet auf der Intransparenz, wozu Daten über Personen bearbeitet werden und darauf, dass so unser Verhalten beeinflusst werden kann und damit steuerbarer ist.[91] Auch ist den wenigsten bewusst, dass durch die Vorhersehbarkeit unseres Handelns, uns der Zufall abhanden kommt[92] und «Big Brother is watching you»[93] längst Realität ist.

---

90 Gleicher Ansicht, aber skeptisch betreffen der Umsetzung: Rolf H. Weber (Fn. 4), Rz. 24.

91 Meldet z.B. ein Navigationsgerät an einer bestimmten Stelle Stau und gibt es eine Umfahrungsstrecke an, nutzen die meisten Benutzer diese Information und verwenden die Umfahrungsstrasse. Ein Navigationsgerätehersteller könnte somit diese Information auch missbrauchen.

92 Miriam Meckel, Next: Erinnerungen an eine Zukunft ohne uns, Rowolt 2011.

93 George Orwell, 1984, London 1949.

Eine Grenzziehung ihm Rahmen dieses Einsatz-Zweckes-Ansatzes wäre schwierig zu bewerkstelligen, weshalb der Ansatz zu verwerfen ist.

Der Kontrollverlust durch Big Data kann m.E. auch nicht durch ein Verbot der Verwendung von Big Data bzw. von Datenverknüpfungen wieder hergestellt werden.

Das einzige Instrument, welches hilfreich ist, ist der *Ausgleich des Machtverhältnisses zwischen demjenigen, der Big Data einsetzt und Betroffenen.*[94] Betroffenen Personen sollen Instrumente in die Hand gegeben werden, welche eine effiziente Durchsetzung ihrer Rechte ermöglichen. Das beinhaltet nebst den materiellen Grundsätzen eines Rechts auf informationelle Selbstbestimmung auch ein darin enthaltenes Recht auf Vergessen. Formell muss eine *effiziente Rechtsdurchsetzung* möglich gemacht werden durch ein einfaches und v.a. kostenloses Verfahren für den Betroffenen gegenüber dem Datenverletzter analog dem Schlichtungsverfahren in Miet- oder Arbeitsrechtssachen. Dazu muss der Betroffene die Möglichkeit haben, Einblick zu erhalten, wie das über ihn erstellte Persönlichkeitsprofil erstellt wurde und auf welchen Datenquellen die Resultate beruhen. Der *Informations- und Auskunftsanspruch von Betroffenen* muss also ausgedehnt werden, auf die Frage woher die Daten kommen. Sodann muss der *Datenschützer mit mehr Kompetenzen* ausgestattet werden. Datenschutzverletzungen dürfen auch nicht mehr grösstenteils straflos sein. Vielmehr sollen Unternehmen, welche gerade der Datensicherheit zu wenig Beachtung schenken geahndet werden können.

VI. Zusammenfassung

Zusammenfassend kann festgehalten werden, dass die Datensicherheit nach DSG wie auch die technische Datensicherheit bei Big Data schwierig einzuhalten ist. Dies nicht nur aus der rechtlichen Warte, sondern v.a. auch aus der technischen Perspektive. Wird Big Data im Rahmen von Tools eingesetzt so können v.a. Zugriffsprobleme besser kontrolliert werden, wie wenn dies ausserhalb derselben erfolgt.

Bei der Anwendung von Big Data müssen folglich die verschiedenen Zielkonflikte abgewogen werden: Effizienz, Compliance, Kosten etc. versus Schutz der Persönlichkeit und der informationellen Selbstbestimmung.

Der Gesetzgeber ist gefordert, Möglichkeiten aufzuzeigen, wie Big Data gesetzeskonform eingesetzt werden kann. Das Verbieten von Big Data würde m.E. ge-

[94] Ebenso Rolf H.Weber (Fn. 4), Rz. 21.

genüber Staaten, welche liberaler mit dem Datenschutz umgehen, dazu führen, dass ein gewichtiger Wettbewerbsnachteil für Länder mit restriktiverem Datenschutz (EU/Schweiz) entsteht. Big-Data-Technologien werden bereits heute verwendet, ein Verbot ist daher faktisch nicht durchsetzbar.

Folglich muss der Gesetzgeber Betroffene wie auch Wächter des Datenschutzes wie den EDÖB mit mehr rechtlichen Mitteln ausstatten.