

Reinhard Riedl

Welchen Regulierungsbedarf schaffen die Paradigmenwechsel von Big Data?

In this article, varying types of Big Data, the several phases of the Big Data process and some of the most important application areas of Big Data are outlined. Then the paradigm shifts, chimeras and illusions coming along with Big Data are discussed, in order to derive how companies and administration can practice Good Governance and where regulatory action is needed. (ah)

Category: Beiträge

Field of law: Big Data, Open Data & Open Government

Region: Switzerland

Citation: Reinhard Riedl, Welchen Regulierungsbedarf schaffen die Paradigmenwechsel von Big Data?, in: Jusletter IT 21 May 2015

Inhaltsübersicht

- 1 Einführung
 - 1.1 Formen von Big Data
 - 1.2 Geschichte und aktueller Diskussionsbedarf
- 2 Nutzung und Nutzen
 - 2.1 Vier Schritte des Big Data Prozesses
 - 2.2 Das Stack-Prinzip
 - 2.3 Nutzungsbereiche
 - 2.4 Praktische Herausforderungen
- 3 Kritische Paradigmenwechsel und Schimären
 - 3.1 Paradigmenwechsel
 - 3.2 Resultierende Chancen und Risiken
 - 3.3 Weitergehende Hoffnungen
 - 3.4 Schimären der Wissenschaft und Fiktionen der Praxis
 - 3.5 Schlussfolgerungen
- 4 Regulierungsbereiche
 - 4.1 Datenschutz
 - 4.1.1 Respektierung des Datenschutzes in Organisationen unter Schweizer Recht
 - 4.1.2 Wahrung der Rechte des Individuums
 - 4.2 Beschränkung der Datennutzung und Datenbereitstellung
 - 4.2.1 Beispiele möglicher Diskriminierungen
 - 4.2.2 Schlussfolgerungen
 - 4.3 Staatliche Förderung von Big Data
 - 4.4 Die Rolle der Wissenschaftä
- 5 Zusammenfassung

1 Einführung

[Rz 1] Es gibt keine etablierte Definition für Big Data, sondern nur Umschreibungen wie die 3-V-Beschreibung (Volume, Variety, Velocity), die 6-V-Beschreibung (plus Veracity, Validity und Volatility) oder N-V-Beschreibungen (plus Value, Visualization ...). Die wesentlichen Eigenschaften sind hohe Quantität, fragwürdige Qualität und fragwürdiger aber potentiell extrem hoher Nutzen. In diesem Beitrag werden wir mit folgender Definition arbeiten.

[Rz 2] *Arbeitsdefinition: Bei Big Data geht es darum, implizit in Daten vorhandene Informationen explizit zu machen, um Geschäftsnutzen zu erzeugen.*

[Rz 3] Wir betrachten Big Data unter der Prämisse, dass geschäftskritische Informationen aus Daten abgeleitet werden sollen, bzw. allgemeiner Informationen, die für die Big Data Anwender einen signifikanten Wert darstellen. Typischerweise geht es dabei um sehr grosse Datenmengen, die aus unterschiedlichen Quellen zusammengeführt werden und deren Qualität ebenso unklar ist wie ihr Nutzen. Die Ableitung der geschäftskritischen Information geschieht unter Verwendung von Datenmodellen, Algorithmen und Computern. Dabei können die verwendeten Modelle und Algorithmen mehr oder weniger «intelligent» sein, i.e. Kontextwissen über die Daten und Sachlogik des Big-Data-Verwendungszwecks enthalten.

1.1 Formen von Big Data

[Rz 4] Es gibt cum grano salis sechs Kernformen von Big Data: *traditionelles Information Retrieval (IR)*, *klassisches Big Data*, *Freestyle Big Data*, *Big Data auf dem Graphen*, *Big Data Profiling* und

komplexes Big Data.

[Rz 5] Traditionelles IR sucht in der Regel die sprichwörtliche Nadel im Heuhaufen, wofür es die Suchanfrage mit Wissen über den Suchenden (Vergangenheit und aktueller Kontext) und Wissen über Such- und Finderverhalten anderer kombiniert und mit Eigenschaft von Elementen aus einer Suchmenge, z.B. Websites bei der Google-Suche, vergleicht. Klassisches Big Data sucht dagegen mehr den roten Faden im Heuhaufen: Es schätzt nicht explizit bekannte Attribute und basiert darauf, dass man zuvor aus grossen Datenmengen abgeleitet hat, wie Eigenschaften respektive Attribute zusammenhängen. Diese Erkenntnis wird für das Schätzen unbekannter geschäftskritischer Eigenschaften durch die Nutzung bekannter andere Eigenschaften benutzt.¹ Freestyle Big Data strebt an, Zusammenhänge in Daten zu erkennen, von deren Existenz man zuvor nichts wusste. Man spricht dabei auch vom explorativen Data Mining oder umgangssprachlich vom wissenschaftlichen Fischen im Trüben. Big Data auf dem Graphen versucht, die Zusammenhänge zwischen sehr unterschiedlichen Datensets auszuwerten und daraus Erkenntnisse abzuleiten, wobei weitgehend andere Algorithmen als bei den anderen Kernformen verwendet werden. Big Data Profiling versucht Daten zu de-anonymisieren und personenbezogen zu integrieren, um ein digitales alter Ego von natürlichen Personen oder auch von Organisationen zu schaffen. Und komplexes Big Data verwendet alle bisher beschriebenen Formen, um Konzepte oder allgemeiner Problemlösungen zu finden. Wobei die hier verwendete begriffliche Klassifikation primär dazu dienen soll, verständlich zu machen, was unter anderem bei Big Data passiert (und nicht als wissenschaftliche Ontologie zu verstehen ist).

[Rz 6] Beim IR haben wir es mit mehreren Ingredienzien zu tun: formalen Repräsentationen von Informationsobjekten und von der Suchanfrage und formale Repräsentationen von Such-Historien, wobei letztere im Idealfall sowohl für den Suchenden als auch für andere mit ähnlichen Interessen bereitstehen. Daraus wird eine Rangliste der Objekte aus der Suchmenge definiert, die beschreibt (genauer: schätzt), wie relevant ein gesuchtes Objekt für den Suchenden in seiner jeweiligen Situation ist. Aus rechtlicher Perspektive besonders kritisch sind dabei zwei Punkte: Erstens kann Wissen über den Suchenden und seine aktuelle Situation genutzt werden, um die Qualität der erstellten Rangliste zu verbessern. Der Preis ist der Verlust von Privatsphäre. Zweitens kann Wissen über das Verhalten anderer ebenfalls genutzt werden, um die Qualität der Rangliste zu verbessern. Analog beim klassischen Big Data.

[Rz 7] Der Einfachheit halber kann man hier von drei Ingredienzien sprechen: die geschäftskritische Frage, empirische Daten verschiedenster Art aus denen ableitbar ist, wie die richtige Antwort auf die gestellten Fragen mit bestimmten Eigenschaften korreliert und bekannten Daten zu diesen Eigenschaften. Auch hier sind oft die relevanten empirischen Daten personenbezogen, wenn auch möglicherweise (oder hoffentlich) anonymisiert.

[Rz 8] Beim Freestyle Big Data gibt es nur Daten und menschliche Intuitionen, die in den Daten nach Mustern suchen, die dann mit Hilfe der Intuition interpretiert werden. Der Spezialfall, dass man nach einer Menge fix definierte Muster sucht, ist eine Form des klassischen Big Data. In der Praxis können so klassisches Big Data und Freestyle Big Data kombiniert werden. Je mehr Wissen über die Daten und den Echtweltbereich, den sie beschreiben, existiert, desto grösser sind die Erfolgchancen.

¹ Mathematisch betrachtet ist traditionelles IR ein Spezialfall von klassischem Big Data (weil der Fitness-Wert von Dokumenten bezogen auf eine Suchanfrage geschätzt wird).

[Rz 9] Bei Big Data auf dem Graphen sind die beiden entscheidenden Ingredienzien Daten und Beziehungs-Metadaten (= Daten über die Beziehung zwischen Daten). Ausgehend von Linked Data (www.w3.org/wiki/LinkedData) werden Graphen konstruiert, deren Knoten Daten und deren Kanten Beschreibungen sind, wie sich die Daten zueinander verhalten. Auf diesen Graphen funktionieren gewisse Arten des Suchens besonders gut, insbesondere das Suchen nach verteilten und nach fehlenden Informationen. Letzteres kann man auch allgemeiner als Suche nach generischen Mustern sehen. Insofern lässt sich Big Data auf dem Graphen nicht gegen andere Formen von Big Data klar abgrenzen.

[Rz 10] Beim Big Data Profiling werden in der Regel zuerst Daten zusammengeführt, die sich auf die gleiche, vorerst noch anonyme Person beziehen, um dann durch Abgleich mit bekannten explizit personenbezogenen Daten die Person zu ermitteln. Oft wird die Person erst sehr spät identifiziert oder es wird auf den eigentlichen Deanonymisierungsschritt ganz verzichtet. Solch ein verkürztes Profiling ist auch ein wesentliches Element von IR und anderen Big Data Formen. Auch ohne den eigentlichen Deanonymisierungsschritt stellt es einen Eingriff in die Privatsphäre der Betroffenen dar. Rechtlich ist das Stattfinden der De-Anomysierung sowieso nach dem Schweizer DSG irrelevant, weil ja der letzte Schritt in der Regel mit geringerem Aufwand durchführbar ist als die davor durchgeführten Schritte.

[Rz 11] Neben den skizzierten fünf Kernformen gibt es noch viele andere Formen von Big Data, beispielsweise die menschliche Interpretation von visualisierten Daten oder die Ableitung von Strategieentscheiden mittels datenbasierten Simulationen. Erstere spielt eine zunehmende Rolle im Freestyle Big Data, letztere wird verwendet, um interessierende Konzepte zu identifizieren, beispielsweise vielversprechende Strategien. Die quasi industrielle Strategieproduktion durch Big Data ist ein Beispiel für die sechste Kernform, die wir komplexes Big Data nennen und die so vielfältig sein kann, dass wir hier nicht im Detail darauf eingehen. Ebenfalls nicht diskutieren werden wir Big Data als Instrument zur Steigerung der Relevanz von an sich bekannten Fakten (Big Data als Fassade).

1.2 Geschichte und aktueller Diskussionsbedarf

[Rz 12] Sieht man von der Verwendung von Computern ab, dann gibt es Big Data seit Jahrtausenden. Einer der ersten grossen Big Data Erfolge der Neuzeit war Keplers Entdeckung seiner Planetengesetze. Was sich verändert hat, ist, dass eine ehemalige Teildisziplin der Philosophie und spätere Teildisziplin vieler Wissenschaften zu einem Handwerk und für manche auch zu einem Religionsersatz wurde. Während in früheren Jahrhunderten Big Data stets mit einem Denken in Modellen verbunden war, ja meist zur Bestätigung oder Neufindung von Modellen diente, und dabei oft nichtmathematische philosophische Überlegungen eine Rolle spielten, wird heute in Big Data von vielen eine Möglichkeit gesehen, sich von dem Denken in Modellen und vom fachlichen Praxisbezug zu emanzipieren und Datenwissenschaft als neue generische Wissenschaft zu etablieren, mit der man aufgrund der vorhandenen Datenfülle alles und jedes besser beforschen kann als mit traditionellen Forschungsmethoden. Das ist teilweise wahr, teilweise grober Unfug. Big Data wird sich massvoll eingesetzt als ebenso nützlich, wie masslos oder kriminell eingesetzt als gefährlich erweisen. Deshalb ist es wichtig, Potential und Wirkung von Big Data zu analysieren und politisch zu reflektieren, inwieweit es neuen Regulierungsbedarf gibt. Optimistische Übertreibungen sind dabei ebensowenig angebracht wie pessimistische Tiefschwarzmalerei.

[Rz 13] Was heute schon absehbar ist, ist der Bedarf die Datenschutzgesetze grundsätzlich zu

überdenken, weil Big Data Profiling aufzeigt, wie kurzlebig die Eigenschaft «anonymisierte Daten» sein kann, und weil auch die anderen Formen von Big Data demonstrieren, wie viel weiter als bisher vom Gesetzgeber angenommen der Personenbezug und die besondere Schutzwürdigkeit von personenbezogenen Daten² zu verstehen sind.

[Rz 14] Es stellen sich aber viele weitergehende rechtliche Fragen. Unsere digitalen Datenspuren sind vielfältig. Sie können untereinander und mit Sachdaten verbunden werden. Die Durchführung von Big Data setzt fast immer auch auf den Abgleich mit den Daten anderer und die letztendliche Nutzung der Big Data Ergebnisse führt dazu, dass das Verhalten anderer sich auf unsere Chancen im Leben auswirkt. Dies gilt besonders dann, wenn bei der Durchführung von Big Data Fehler gemacht werden, aber auch dann, wenn Big Data richtig gehandhabt wird, weil es ja darum geht, aus der Kenntnis des Verhaltens der vielen Schlüsse auf den situativen Einzelfall (d.h. den eigenen Einzelfall) zu machen. Inwieweit das zulässig ist und nicht die Rechte des Einzelnen unzulässig einschränkt, ist zu hinterfragen.

[Rz 15] Weiters sind in der Datenwelt die Rechte-Konzepte der Informatik dem Gegenstand angemessener als das Eigentumsverständnis der physischen Welt. Der Zusammenhang zwischen Ordnung und Ordnung im Sinn von Carl Schmitt fällt bei virtuellen Cloud-Architekturen weg.³ Und der Zuwachs an Wissen durch Big Data hat zur Folge, dass alle Bereiche, in denen Nichtwissen die Ursache von gesellschaftlich wünschenswertem Verhalten war, politisch neu zur Diskussion stehen. Beispielsweise macht Nichtwissen über die Zukunftschancen Einzelner Solidarität und soziales Teilen einfacher: Mit jemandem solidarisch zu sein, der das sicher nie zurückgeben wird, ist schwieriger als Solidarität bei einem potentiell reziproken Anspruch auf seine/ihre Solidarität. Big Data wirft also viele neue rechtliche und politische Fragen auf.

2 Nutzung und Nutzen

[Rz 16] Big Data wird eingesetzt, um geschäftlich benötigte — oder noch allgemeiner: interessierende — Informationen zu bekommen. Wobei der Begriff Big Data meist nur verwendet wird, wenn wir dafür einen Computer benötigen und dieser über hoher Leistung oder/und «intelligente» Algorithmen verfügt.

[Rz 17] Im einfachsten Fall geht es bei Big Data um ein Finden der passenden Informationen in einer grossen Menge von Daten und Dokumenten. In komplexeren Fällen geht es um ein Finden und Kombinieren der Informationen. Und im Normalfall geht es um ein Erraten (bzw. Schätzen) von Informationen basierend auf anderen, explizit vorhandenen und zugreifbaren Informationen. Dabei arbeiten Mensch und Computer zusammen, wobei oft, aber nicht immer, der Computer fast die ganze Arbeit erledigt.

[Rz 18] Der Nutzen von Big Data ist ganz einfach der Erwerb von relevantem Wissen. Dieser Nutzen ist individuell und situativ, eventuell auch flüchtig, und er muss durch Folgemaassnahmen valorisiert werden.

² Vergleiche: MATHYS, ROLAND, Was bedeutet Big Data für die Qualifikation als besonders schützenswerte Personendaten?, in: Jusletter IT 21. Mai 2015.

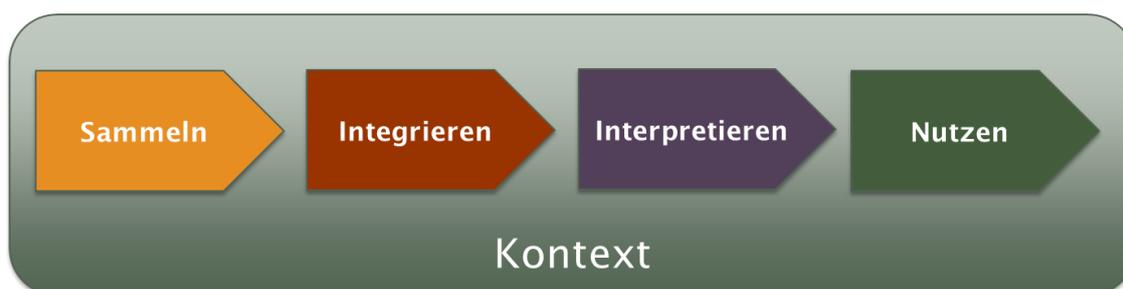
³ KOVAL, PETER / RIEDL, REINHARD, Zur 2.0 (zweiten) Ordnung von E-Government, Proceedings IRIS 2009, Salzburg (2009).

2.1 Vier Schritte des Big Data Prozesses

[Rz 19] Für ein erstes Einarbeiten in das Thema, ist es sinnvoll, Big Data als Prozess zu verstehen, der aus vier Schritten besteht. Die ersten drei Schritte stellen dabei die Durchführung von Big Data dar, der vierte Schritt die eigentliche Nutzung bzw. Valorisierung. Die Schritte sind:

1. Sammeln von Daten
2. Integrieren von Daten
3. Interpretieren von Daten
4. Nutzen der Interpretationsresultate

[Rz 20] Im Fokus der Spezialisten steht häufig nur der dritte Schritt. Hier kommt die Mathematik entscheidend zum Zug. Mittels Datenmodellen und Algorithmen wird aus Daten Wissen extrahiert, das vorher nicht explizit vorhanden war, wobei auch hier die Zusammenarbeit von Maschine und Mensch zentral ist. Streng analytisch argumentierend würde es Sinn machen, den 3. Schritt in Unterschritte zu zerlegen, die das Wechselspiel von Mensch und Maschine aufzeigen. Wir meinen aber, dass eine solches geschäftsprozessmässiges Bürokratisieren leicht in die Irre führt, weil es nahe legt, dass die Verantwortung für den algorithmischen Teil und die Verantwortung für die echtweltbezogene Interpretation der Resultate der Algorithmik trennbar sind — was sie aber nicht sind. Nachstehend die vier Phasen im Diagramm.



[Rz 21] Es ist wichtig, sich stets bewusst zu sein, dass nur in einer idealen Welt, die vier Schritte problemlos getrennt betrachtet werden können und beispielsweise ein richtig durchgeführter dritter Schritt immer zu richtigen Resultaten führt, die höchstens aus Sicht von Schritt 4 wertlos sein können. In einer Welt mit Unzulänglichkeiten beeinflusst aber ein vorgängiger Schritt nicht nur den Nutzen, sondern auch die Qualität der Arbeit in den nachfolgenden Schritten. Werden im ersten Schritt obskure Daten gesammelt oder werden Daten im zweiten Schritt obskur integriert, so wird das nicht notwendigerweise im dritten Schritt erkannt, weshalb auch in diesem Schritt potentiell obskur gearbeitet wird. Eine bestmögliche Entkopplung der vier Schritte ist nur durch Zweierlei möglich: eine korrekte und gut verständliche Dokumentation des Outputs plus Direktkommunikation mit den Akteuren im nächsten Schritt. Das kann man so auf den Punkt bringen: Die Teilung der Arbeit funktioniert im Big Data nur sauber bei enger Zusammenarbeit.

[Rz 22] Aus organisationsstrategischer wie auch aus politischer und rechtlicher Perspektive sind in den vier Schritten ganz unterschiedliche Aspekte relevant. Aus Gesamtperspektive gilt es naturgemäss zu beurteilen, ob das Handeln im Big Data Prozess rechtskonform ist und ob dadurch gesellschaftlicher Schaden oder Nutzen entsteht. Doch das ergibt sich einerseits aus einer Vielzahl von Teilaspekten der einzelnen vier Schritte und andererseits aus der Wechselwirkung mit dem Kontext, in dem die vier Schritte stattfinden.

[Rz 23] Im ersten Schritt werden die Datenquellen für die Durchführung von Big Data identifiziert und die Daten aus diesen Quellen werden gesammelt, um sie in den nachfolgenden Schritten für Big Data zu nutzen. Dabei haben die Rechte auf die Quellen, das Wissen über die Quellen, die Qualität der Quellen und das Passen der Quelle zum Zweck der Big Data Anwendung sowie das gut nachvollziehbare Dokumentieren von allem kritische Bedeutung. Je nach Fall werden Daten bei Big Data ad hoc oder strategisch geplant eingesetzt und es werden nur eigene oder auch externe Quellen genutzt. Dabei ist die Datenqualität kritisch.

[Rz 24] Im zweiten Schritt werden die gesammelten Daten aufbereitet und zusammengefügt zu einer Quelle für die Anwendung von Big Data. Das kann je nachdem sehr aufwendig sein oder sehr trivial. Die Abgrenzung zwischen zweitem und drittem Schritt ist auch nicht unbedingt scharf. Ziel des zweiten Schritts ist es, dass im dritten Schritt auf eine Datenquelle mit einem Datenmodell zugegriffen werden kann. Das heisst aber nicht, dass alle Daten die gleiche Syntax und Semantik besitzen müssen. Kritisch ist allerdings die Qualität der Integration. Werden Kraut und Rüben, oder Gemüse unterschiedlicher Güteklassen gemischt, so dass dies im nachfolgenden Schritt nicht mehr unterscheidbar ist, so entsteht eine Datenquelle mit sehr schlechter Qualität für den nächsten Schritt. Je nachdem ob identische oder unterschiedliche Datenmodelle den Daten zugrunde liegen (oder diese gar nicht bekannt sind) und je nachdem die Qualität der Daten ähnlich oder unterschiedlich ist, sollten darum andere Techniken eingesetzt werden. Ein wesentliches Instrument zur Integration ist im Fall heterogener Datenquellen die Linked Data Technologie.

[Rz 25] Im dritten Schritt findet die eigentliche Bearbeitung der nunmehr integrierten Big Data Datenquelle statt, mit dem Ziel neue Erkenntnisse zu schaffen. Hierbei tun sich unter den Wissenschaftlern ideologische Gräben auf. Cum grano salis geht es um die Frage, ob die Auswertung der Daten mit oder ohne Modell des Entstehungs- und Nutzungskontexts geschieht. Oder präziser darum, wie aussagekräftig das Modell sein muss, bzw. wie viele Interpretationsannahmen drin stecken muss, bzw. wie viel Wissen über die Wirklichkeit genutzt werden muss. Die Hoffnung besteht, dass man in Zukunft in der wissenschaftlichen Forschung weder Modelle noch eine Reflexion der Resultate benötigt, weil die Daten allein schon alles aussagen.

[Rz 26] Der vierte Schritt besteht im Anwenden des erworbenen neuen Big Data Wissens. Hier werden die Resultate valorisiert. Kritisch ist, ob die Anwendung automatisiert oder durch kompetente Experten geschieht, das heisst präzise, inwieweit Menschen darin involviert sind, welche Qualifikationen sie besitzen, welche Regeln es gibt und welche Kontrolle. Die Bandbreite in der Praxis ist sehr gross, von vordefinierten Prozessen bis zur Nutzung von Big Data als explorativem wissenschaftlichem Instrument und von kreativem ad hoc Einsatz in Freiräumen bis zum Einsatz unter klaren Regeln und mit Auditprozessen. Letzteres ist nach unserem Wissensstand sehr selten der Fall.

2.2 Das Stack-Prinzip

[Rz 27] Für ein besseres Verständnis der Auswirkungen (wie auch mancher Diskussionen rund um Rohdaten) ist es wichtig, das Stack-Prinzip der Informatik zu kennen. Um die grosse Komplexität zu reduzieren und die Wiederverwendung von Lösungen zu ermöglichen werden Aufgaben in Teilaufgaben zerlegt, für die jeweils ein separates Lösungsdesign stattfindet. Kommunikation wird so durch einen Protokoll-Stack durchgeführt. Bei Datenanalysen findet oft eine Datenverarbeitungspipeline Anwendung, in der verschiedene Verarbeitungselemente mit jeweils

ganz spezifischen Aufgaben hintereinander geschaltet werden. Man sollte sich Big Data auch als Zusammenstecken von Verarbeitungsmodulen vorstellen (das natürlich nach gewissen Regeln stattfindet). Wichtig daran für den Nichtinformatiker ist, dass nach dem Black-Box-Prinzip der Informatik eine Trennung des WAS und WIE stattfindet: Um ein Modul zu nutzen muss man nur verstehen, was das Modul leistet und welche Schnittstellen es hat, nicht aber wie es implementiert ist. So richtig und wichtig das aus Komplexitätsmanagementperspektive ist, so sehr fördert es auch, dass das Wie vergessen geht und in der Praxis viele Module eingesetzt werden, deren Algorithmen unbekannt sind.

2.3 Nutzungsbereiche

[Rz 28] Wichtige Anwendungsbereiche sind (angeordnet nach subjektiv beurteilter Relevanz und ohne Anspruch auf Vollständigkeit) Wirtschaft, Wissenschaft, Gesundheitswesen, Verwaltung, Energie und Verkehr, Justiz, Politik und Kunst. In der Wirtschaft wird Big Data vor allem dafür verwendet, attraktive Kunden zu identifizieren und Produkte und Dienstleistungen masszuschneiden, beispielsweise im Customer Relationship Management. Der Nutzen besteht (sofern das professionell gemacht wird) darin, dass der im Durchschnitt geschaffene Wert pro Kunde grösser wird und damit auch die Verdienstmöglichkeiten steigen — sei es dass Kunden gewonnen werden, die sonst nicht kaufen würden, oder sei es, dass mehr Kunden mit hoher Zahlungsbereitschaft gewonnen werden, oder sei es, dass an bisherigen Kunden mehr verdient wird. Daneben wird Big Data auch für Strategieentwicklung und Planungsaufgaben eingesetzt. Der ideelle Nutzen besteht dann darin, dass Strategien und Pläne faktenbasierter sind. Ob dies sich in einen realen Nutzen umwandeln lässt, hängt von der Qualität der Faktennutzung ab.

[Rz 29] Insgesamt steht zu erwarten, dass Big Data die Wirtschaft stark verändern wird, wobei seine Wirkung je nach Branche sehr unterschiedlich sein wird und neben neuen Chancen auch neue Abhängigkeiten entstehen werden, insbesondere Abhängigkeit von jenen Dienstleistern, die über grosse Datenmengen zu Kundenpräferenzen verfügen. Eine am E-Government Institut der BFH laufende Expertenbefragung hat als Zwischenergebnis sehr divergente Meinungen darüber ergeben, wer davon profitieren wird.

[Rz 30] In der Wissenschaft wird Big Data zum Erkenntnisgewinn eingesetzt, der ja das Kernziel von Wissenschaft ist und also per definitionem — also eo ipso — Nutzen schafft. Dabei gibt es die Hoffnung, Teile des wissenschaftlichen Arbeitens automatisieren und den wissenschaftlichen Fortschritt wesentlich beschleunigen zu können. Es gibt aber auch ganz andere Perspektiven: Wenn Daten das neue Öl sind, wie von vielen gesagt wird, so sind Wissenschaftler möglicherweise die neuen Hilfsarbeiter beim Ölfördern.

[Rz 31] Im Gesundheitswesen wird Big Data nicht nur für Forschung, Absatzsteigerung und strategische Ressourcenplanung eingesetzt (siehe oben), sondern auch für eine personalisierte Behandlung von Patienten. Hier besteht der Nutzen darin, dass Therapien massgeschneidert eingesetzt und Informationen aus den Krankheitsverläufen anderer Patienten genutzt werden. Der Nutzen besteht in wirksameren Behandlungen, wobei dies bei Therapien in einigen Bereichen auf Kosten der bisherigen wissenschaftlichen Standards gehen könnte und Qualitätsfragen aufwirft. In Zukunft werden Gesundheitsdaten jedenfalls nicht absolut sondern patientenindividuell interpretiert und Therapien individuell durchgeführt werden.

[Rz 32] In der Verwaltung wird Big Data bisher primär im Sicherheitsbereich eingesetzt. Dies wird

in Zukunft noch zunehmen. Im Prinzip kann es aber auch ganz ähnlich eingesetzt werden wie in der Wirtschaft, das heisst für Planung, Strategieentwicklung und weitere Steuerungsaufgaben. Dazu wird in Zukunft auch der Einsatz in der Katastrophenvorsorge kommen, beispielsweise bei der Vorhersage von Hangrutschen. Massgeschneiderte Verwaltungsdienstleistungen analog zu massgeschneiderten Dienstleistungen in der Wirtschaft sind ebenfalls vorstellbar, finden aber derzeit vor allem in menschlicher Interaktion und verhandlungsbasiert statt. Bei allen potentiellen Vorteilen von Big Data bestehen aber auch grosse Befürchtungen, dass ganze operative Verwaltungstätigkeiten automatisiert werden könnten, beispielsweise die Vergabe von Sozialhilfe.

[Rz 33] Im Bereich Energie und Verkehr spielt Big Data schon jetzt eine grosse Rolle, diese wird aber in Zukunft noch wesentlich zunehmen — sei es, dass innovative Formen der verteilten Energiespeicherung und Energieverbrauchsteuerung entstehen werden, sei es, dass zusammen mit dem Internet of Things die Verkehrssteuerung wesentlich verbessert werden wird.

[Rz 34] In der Justiz besitzt Big Data grosses Veränderungspotential, insofern die Wahrscheinlichkeit von Annahmen z.B. in Bezug auf Tatverläufe berechnet werden kann. Also beispielsweise: Gegeben dass die Fakten x_1 bis x_N zu berücksichtigen sind, wie gross ist die Wahrscheinlichkeit, dass sich in der Wirklichkeit eine Situation so abgespielt hat, wie das im Urteil angenommen wurde? Oder wie gross ist die Wahrscheinlichkeit eines Rückfalls? Et cetera. Wie man mit diesen zukünftigen Möglichkeiten am besten umgeht, ist eine offene Frage.

[Rz 35] In der Politik hat Big Data spätestens mit dem zweiten Obama Präsidentschaftswahlkampf eingezogen, als sein Wahlkampfteam mittels Konsumdaten Wechselwähler identifizierte, die dann persönlich angesprochen wurden. In Zukunft wird Big Data auch ganz analog zur Wirtschaft in der Politik eingesetzt werden können, um besonders attraktive Wahlprogramme und politische Positionen zusammenzustellen. Der Nutzen besteht, lapidar formuliert, in der Erhöhung von Wahlchancen. Für die Ausübung von Exekutivämtern bietet Big Data zudem neue Möglichkeit der Langfristplanung. Der Schweizer Bundesrat könnte Big Data beispielsweise als Ergänzung zur aktuell verwendeten Szenario-Technik einsetzen⁴, um prognostizierbare Entwicklungen als Eckpunkte für Szenarien zu verwenden.

[Rz 36] Ganz anders und zugleich ganz ähnlich ist die Lage in der Kunst: Hier hilft Big Data Zusammenhänge zu offenbaren und kann auch benutzt werden, um Adressaten zu manipulieren, was in der Kunst ein beabsichtigtes, moralisch erziehendes Programm sein kann. So ist zu hoffen, dass die Kunst die verschiedensten Big Data Phänomene reflektieren und ihrem Publikum verständlich näher bringen wird.

2.4 Praktische Herausforderungen

[Rz 37] Die praktischen Herausforderungen von Big Data sind je nach Fall sehr unterschiedlich. Generell kann man aber sagen, dass dort, wo das Potential wirklich genutzt wird, die Herausforderung sehr gross sind. Auch für Big Data gilt das No-Free-Lunch-Theorem.

[Rz 38] Die zu bewältigenden Probleme reichen von der Identifikation dessen, was man wissen will (klassisches Big Data) über die sehr anspruchsvolle «intelligente» Datenverarbeitung (die

⁴ Vortrag von Frau Bundeskanzlerin CORINA CASANOVA beim SGVW Netzwerkanlass am 23. Oktober 2014 auf dem Landgut Lohn

richtigen Datenquellen mit ausreichender und bekannte Datenqualität finden + richtige Modelle und richtige Algorithmen nutzen) oder die fast noch anspruchsvollere Mustersuche (Voraussetzungen schaffen, dass Muster entstehen [z.B. bei netzwerkbasierter Angriffsdetektion] + Muster sehen + Muster verstehen) bis zur bislang teilweise noch sehr unreifen Technologie (z.B. bei Linked Data Verarbeitungspipelines) und der tatsächlichen effektiven Nutzung der Big Data Resultate, gegen die es oft heftige, erklärte und nicht erklärte, Widerstände in den Organisationen gibt.

[Rz 39] Es geht also um Herausforderungen bei Analyse, Design und Implementation, wobei jeweils vielfältige Formen von Intelligenz, Wissen und Fertigkeiten benötigt werden. Dazu kommen die zwingende Notwendigkeit Datenschutzgesetze einzuhalten und das potentielle Reputationsrisiko im Fall, dass es öffentlichen Widerstand gegen die Datennutzung gibt. Bisher war freilich das Hauptproblem für die Unternehmen, überhaupt Nutzen aus den Daten zu ziehen.

[Rz 40] Eine vielleicht noch grössere Herausforderung als die effektive Durchführung des Big Data Prozesses stellt der strategische Umgang mit Big Data dar. Das Top-Management muss dazu u.a. folgenden Fragen beantworten: Welche Daten will ich in Zukunft nutzen? Wie komme ich zu diesen Daten? Welche Maschinerie will ich für ihre Verarbeitung nutzen? Wie komme ich zu nützlichen Interpretationen? Wie stelle ich sicher, dass die Ergebnisse genutzt werden? Welches Team benötige ich? Wie gestalte ich das organisatorische Steuerungssystem?

[Rz 41] Bei komplexen Anwendung braucht es für eine effektive Durchführung des Big Data Prozesses ein in Zusammenarbeit eingeübtes — sozusagen ein eingespieltes — multidisziplinäres Team mit Domänen-Experten, Datenspezialisten/Statistikern, Mathematikern, Informatikern, Qualitätsmanagern, Juristen und umsetzungstarken Führungskräften. Und es braucht die Unterstützung durch das Engagement und Sachverständnis der Geschäftsleitung. Das ist nicht gerade wenig.

3 Kritische Paradigmenwechsel und Schimären

[Rz 42] Mit Big Data sind verschiedene Paradigmenwechsel verbunden: tatsächliche, realistisch erhoffte und illusionäre. Der Umgang mit Daten und damit mit Fakten wird durch Big Data in mehrfacher Hinsicht revolutioniert. Wobei die dabei Anwendung findenden Ideen nicht neu sind, aber sie neu im grossen Stil eingesetzt werden. Das schafft Nutzen und Risiken — und zwar sowohl Risiken des falschen Gebrauchs, des grössenwahnsinnigen Gebrauchs und des böswilligen Missbrauchs als auch Risiken die aus dem erfolgreichen kompetenten und rechtmässigen Gebrauch folgen, weil dieser die Wirtschaft, die Politik und das soziale Zusammenleben stark verändert.

3.1 Paradigmenwechsel

[Rz 43] Sechs Paradigmenwechsel werden in den nächsten Jahren in vielen Bereichen keinen Stein auf dem anderen lassen: der Wechsel vom sorgfältigen zum unsauberen Umgang mit Daten, der Wechsel vom Denken in Kausalitäten zum Nutzen von Korrelationen, der Wechsel vom Klassifizieren (von Kunden, Partners etc.) zur radikalen Personalisierung basierend auf digitalen Alter Egos, der Wechsel vom bedarfsorientierten Datensammeln zum Datenfluten mit dem Internet of Things, die Veränderung der Wirtschaft durch mächtige von einem einzigen Datenherrn

beherrschte Universum und der Abschied vom monodisziplinär juristischen Denken im Datenschutz.

[Rz 44] Alle sechs Paradigmenwechsel finden im Kontext eines a priori von Big Data unabhängigen Bedeutungsverlusts des Schutzes der Privatsphäre statt. Menschen kommunizieren heute über sich selber wesentlich mehr wahre und erfundene Fakten an wesentlich mehr Menschen als früher, insbesondere über digitale soziale Netze. Das vergrößert das Nutzenpotential von Big Data und beschleunigt seine Einführung entscheidend. Damit einher geht auch eine grössere Gelassenheit gegen die hässlichen Details aus der Vergangenheit von Menschen, was die Notwendigkeit des Schutzes der Privatsphäre mindestens scheinbar reduziert. Scheinbar deshalb, weil in unserem eher liberalen Teil der Welt Werkzeuge entwickelt werden, die einerseits in anderen Teilen der Welt dazu verwendet werden um Menschen ausfindig zu machen, die gefoltert und/oder ermordet werden sollen und andererseits auch bei uns vom organisierten Verbrechen für verschiedenste illegale Zwecke genutzt werden. Diesem Problem mit Exportbeschränkungen und Verkaufskontrollen für Big Data Werkzeuge zu begegnen, scheint von vornherein zum Scheitern verurteilt.

(1) Laxes Qualitätsverständnis + Unsauberes Mischen

[Rz 45] Erstens entspannt sich das Verhältnis zur Datenqualität und Datensauberkeit. Neu werden Daten, die verschiedenen Datenmodellen entsprechen und unterschiedliche Qualität haben, zusammengemischt. Das wissenschaftliche Gebot, nicht Äpfel und Birnen zu mischen fällt in Bezug auf die Datenwissenschaft, Daten unterschiedlicher Art und Qualität werden zusammengeführt. Dabei gibt es (a) den Fall, dass Daten des gleichen Datenmodells mit unterschiedlicher Qualität zusammengeführt werden, (b) den Fall, dass semantisch gleiche Daten mit unterschiedlicher Syntax zusammengeführt werden, (c) den Fall, dass syntaktisch und semantisch ähnliche Daten integriert werden (was aus Nutzungssicht der aktuell interessanteste Fall ist, beispielsweise in der medizinischen Forschung) und (d) den Fall, dass gänzlich verschiedene Daten mit inhaltlichen Querbezügen integriert werden (durch Linked Data). Das laxer Qualitätsverständnis geht in der Praxis so weit, dass auch Daten genutzt werden, deren genaue Semantik unbekannt ist. Die im Schweizer Datenschutzgesetz geforderte Richtigkeit der besonders schützenswerten Personendaten (DSG, Art. 5) wirkt in der Big Data Welt wie ein Anachronismus.

(2) Korrelationen statt Kausalitäten

[Rz 46] Zweitens wird in vielen Bereichen das Denken in Kausalitäten durch ein Arbeiten mit Korrelationen ersetzt. Das ist ein grösserer Schritt als bei logischen Implikationen den Richtungspfeil abzuschaffen, was schwerwiegend genug wäre. Bedenkt man dass das Erlernen des Denkens in logischen Zusammenhängen der fundamentale Einstieg in die modernen Wissenschaften ist, so geht dieser zweite Paradigmenwechsel noch viel mehr an die kulturelle Substanz unserer wissenschaftsgeprägten Gesellschaft als der erste. Dass Computer uns dabei die Rechenarbeit abnehmen, ändert nichts am Problem, dass in unserem Denken beobachtete Korrelationen oft in scheinbar bewiesene Kausalitäten umgewandelt werden.

[Rz 47] Wie der Paradigmenwechsel zu bewerten ist, hängt entscheidend vom Kontext ab. Das Beispiel von Gapminder (www.gapminder.ch) zeigt, dass die in Korrelationen ausgewiesene Faktenlage oft krass den Intuitionen widerspricht und uns in unserem Weltverständnis einige wesentliche Schritte weiter bringt.

(3) Radikale Personalisierung / Digitales Alter Ego

[Rz 48] Drittens führt Big Data zu einer radikalen Personalisierung. Die traditionelle Segmen-

tierung in der Werbung und der Produkt- bzw. Dienstleistungsspezialisierung wird durch das Fokussieren auf den oder die Einzelne/n ersetzt. Wo bisher Adressaten klassifiziert und klassenspezifisch angesprochen oder behandelt wurden, werden sie neu als Individuum angesprochen und behandelt, wobei sie im ersten Schritt weitestgehend mit ihrer digital aufgezeichneten Vergangenheit gleichgesetzt werden und im zweiten Schritt künstlich als Interpolationen des Verhaltens von Personen mit ähnlicher Vergangenheit berechnet werden. So versucht man direkt auf die Person zu spielen, spielt aber aus praktischer Notwendigkeit auf ihr digital konstruiertes Alter Ego, das eine Mischung aus selektiver individueller Vergangenheit und Massenverhalten ist. Frei nach GOETHE «Bei Euch, Ihr Herren, kann man gewöhnlich das Wesen aus den Daten der anderen lesen.»

[Rz 49] Interessant dabei ist insbesondere, dass Daten über das Verhalten anderer zu meinen personenbezogenen Daten werden. In der Praxis kann das zu vielen Unannehmlichkeiten führen — nicht nur zu aus Literatur, Film und Fernsehen bekannten falschen Verdächtigungen in Kriminalfällen sondern zu ganz banalen Alltagserschwernissen wie andauernden Kontrollen oder der Verweigerung von Dienstleistungen. Viel weiter geht allerdings die Umsetzung des alten Ökonomen-Traums von der totalen Preisdiskriminierung, bei der jeder so viel zahlt, wie ihm ein Gut oder eine Leistung wert ist. In der Konsequenz wird es auch im Supermarkt nur personalisierte Preise geben und wir werden feststellen, dass Big Data Formen von Solidarität auflöst, die uns bisher gar nicht bewusst waren.

(4) Totale Vermessung der Welt / Datenflutung

[Rz 50] Big Data gibt auch dem neuen Trend zum Internet of Things (IoT) zusätzlichen Sinn und wird das Datensammeln mittels des IoT stark fördern. Insbesondere dort, wo es keinen direkten Datenbedarf gibt, macht das Big Data Potential, wie unbestimmt es auch immer sein mag, das Vermessen der Welt attraktiv. Sofern es sich um Sachdaten handelt ist das datenschutzkonform. Werden diese Sachdaten aber mit Personen in Verbindung gesetzt, so können je nach Situation neben harmlosen auch für die Betroffenen schädliche Schlüsse gezogen werden. Ganz abgesehen davon dass viele harmlose Lügen im Zusammenleben zum Problem werden, weil sie entweder direkt durch IoT-Daten oder indirekt durch Big Data entlarvt werden. Die Wirkung dieses vierten Paradimenwechsels ist vor allem in Zusammenspiel mit den ersten dreien bedenklich. Sei es, weil meine Versicherungsprämie wegen IoT-Messdaten steigt (erhöhtes Kollapsrisiko bei Menschen, die in 40-Grad-Zügen reisen ...). Sei es, dass ganz abseits von Big Data Informationen Menschen zugänglich gemacht werden, die ihnen Erkenntnisse ermöglichen, die sie ohne IoT nicht haben könnte (wie den Verdacht auf Ehebruch aufgrund der Wasserbrauchsaufzeichnungen). Hier stellt sich die Frage, ob nicht der Datenschutz in Zukunft die Bereitstellung von Sachdaten radikal einschränken sollte.

(5) Datenuniversen / Monolithische Diversität

[Rz 51] Big Data benötigt in der Regel grosse Datenmengen, die im Fall meist im grossen Stil anwendbare Resultate liefern, wodurch Economies of Scale gefördert werden. Allerdings ist es auch für ökonomische Nischen mit hochspezialisierten Produkten oder Leistungen einsetzbar, wie schon das Beispiel der personalisierten Medizin zeigt. Die erwartbare Wirkung auf die ökonomische Diversität ist bislang deshalb unklar. Allerdings fördert Big Data insbesondere die Entwicklung von Datenuniversen, die durch einen Datenherrn kontrolliert werden und deshalb eine Art monolithische Diversität besitzen. Jaron Lanier hat dafür den Begriff Sirenen-Server

eingeführt.⁵ Das Geschäftsmodell eines Sirenen-Servers besteht *cum grano salis* darin, Datenlieferanten das Gefühl zu geben, dass sie kostenlose Dienste bekommen, sie indirekt aber fürs Datenliefern bezahlen zu lassen — sei es dass der Sirenenserver Dritten Dienstleistungen verkauft, deren Kosten sich diese von den Datenlieferanten bezahlen lassen (beispielsweise Präferenzprofile von Kunden, um bei diesen den Absatz oder die Produktpreise zu erhöhen, siehe klassisches Big Data), sei es dass er Steuern verlangt (beispielsweise die Verwaltung, der sich von den Einwohnern deren Überwachung finanzieren lässt). Das ist an sich weder neu noch originell, und die Stabilität der Datenuniversen ist noch unklar, die Ausprägung hat aber eine neue Qualität und vor allem Wachstumsdynamik erreicht, so dass man auch hier von einem Paradigmenwechsel sprechen sollte. Bislang sind Datenuniversen rund um Gratis-Informations- und -Kommunikationsdienstleistungen entstanden, sowie im Wartungs- und Vermietungsgeschäft, beispielsweise in der Autobranche, und im Geheimdienstwesen entstanden. Weitere Bereiche werden folgen.

(6) Transdisziplinärer Datenschutz

[Rz 52] Mit der weitgehenden Infragestellung der rechtlichen Unterscheidung von personenbezogenen und anonymisierten Daten (aufgrund des Profilings) und von besonders schützenswerten und nicht besonders schützenswerten Daten (aufgrund der klassischen Big Data Methodik) wird die Schutzmacht von Datenschutzgesetzen grundsätzlich in Frage gestellt. In Zukunft kann nur ein Zusammenspiel von Recht, Organisation, Technik und Bildung den Schutz der Privatsphäre gewährleisten. Neben der Rechtssetzung werden sich an den Stand der Technik anpassende Normen und Empfehlungen benötigt, ebenso wie neue Formen der Supervision und des Wissensaustausches, sowie Bildungsaktivitäten. Dieses transdisziplinäre Verständnis ist allerdings erst im Entstehen. Zur Diskussion steht unter anderem die Wiederbelebung des Konzepts der vertrauenswürdigen nicht verfolgbaren Anonymität, das auf grundlegende Arbeiten von DAVID CHAUM in den 80ern zurückgeht⁶ und dessen Anwendbarkeit für E-Government bereits vor über 10 Jahren untersucht wurde⁷. Die Alternativen zum transdisziplinären Datenschutzverständnis wäre eine stark restriktive Gesetzgebung, die die Nutzung des Big Data Potentials zu verhindern sucht. Solch eine Gesetzgebung würde aber vermutlich durch die von vielen praktizierte freiwillige Datenbekanntgabe unterlaufen.

3.2 Resultierende Chancen und Risiken

[Rz 53] Zusammenfassend können also mehr Daten genutzt werden (Paradigmenwechsel 1), was dazu führt, dass mehr Daten produziert werden (Paradigmenwechsel 4) und Datenuniversen mit einem dominanten Akteur entstehen (Paradigmenwechsel 5). Durch die Fokussierung auf Korrelationen (Paradigmenwechsel 2) kann in Zukunft in Wirtschaft, Verwaltung, Gesundheitswesen und Politik etcpp. das Verhalten radikal auf den Adressaten zugeschnitten werden (Paradigmenwechsel 3). Das schafft Mehrwert bei Dienstleistungen und reduziert Risiken in der Geschäftstätigkeit, schafft aber auch neue Risiken wegen der Kompliziertheit der Methoden und der Komple-

⁵ Lanier, Jaron, *Who Owns the Future*, 53ff, Simon and Schuster, New York (2013)

⁶ CHAUM, DAVID, *Security Without Identification: Transaction Systems to Make Big Brother Obsolete*. *Communications of the ACM*, 28(10):1030—1044, October 1985.

⁷ AUERBACH, NIKLAS, *Trustworthy Anonymous Digital Identity in E-Government*, Dissertation, Universität Zürich (2004).

xität der Sachhalte auf die die Methoden angewendet werden. Zudem stellt es neu die Frage, wie viel Solidarität es in unserer Gesellschaft geben soll. Eine erste sehr begrüßenswerte Reaktion auf die neuen Chancen und Risiken ist der sich abzeichnende, aber noch nicht wirklich vollzogene, Wandel im Datenschutzdenken hin zu einem transdisziplinären Verständnis (Paradigmenwechsel 6).

3.3 Weitergehende Hoffnungen

[Rz 54] Eine mit Big Data verbundene grosse Hoffnung ist, in Zukunft in vielen Bereichen gesicherte Zukunftsprognosen zu machen und die strategische Planung zu industrialisieren, das heisst den Datenverarbeitungspipelines zu überlassen. Damit könnte ein neues Paradigma entstehen, dass nämlich aus der Vergangenheit die Zukunft vorausgesagt werden kann. Solch ein neu-deterministisches Weltbild ist für viele sehr attraktiv, weil es demjenigen, der die Zukunft besser voraussagen kann, riesige Vorteile verspricht. Insofern es disruptive Veränderungen gibt, Innovation und Fortschritt immer über kleine Statistiken verfügen, ist aber davon auszugehen, dass Big Data gerade bei wesentlichen Veränderungen versagen wird — nicht bei allen, aber bei einigen — und eine Big Data basierte Innovationstriage entstehen könnte: Big Data identifizierbare Innovationen würden stark gepuscht und Big Data nichtaffine Innovation blockiert, weil frei nach MORGENSTERN «Nicht sein kann, was Big Data nicht voraussagt».

[Rz 55] Vernünftiger als Big Data zur neuen Wahrsager-Glaskugel hochzustilisieren ist es, anzustreben, dass man aus den Daten ihre Vorhersagekraft für die Zukunft ableiten kann. Das geht — mit der Big Data grundsätzlich inhärenten Einschränkungen, dass die Zukunft sich eben nicht uneingeschränkt aus der Vergangenheit ableiten lässt — indem man Big Data auf Big Data anwendet, um die Zuverlässigkeit von Big Data Prognosen zu schätzen. So kann Big Data zu einem wichtigen Instrument für die Zukunftsforscher werden. Es liefert zwar nicht notwendigerweise immer bessere Resultate als beispielsweise Delphi-Analysen, hilft aber, sich von den Einschränkungen der vorherrschenden Expertenmeinungen teilweise zu befreien.

3.4 Schimären der Wissenschaft und Fiktionen der Praxis

[Rz 56] Im wissenschaftlichen Diskurs der letzten Jahre sind zahlreiche Schimären aufgetaucht, manchmal dreist und frech beschrieben, manchmal halb versteckt hinter auftrumpfenden Prognosen zum durch Big Data herbeigeführten Zeitenwandel. Sie sind meist nicht unbegründet entstanden, ihre Übertreibungen zeigen aber, welche grosse gesellschaftliche Risiken durch Big Data Illusionen entstehen.

[Rz 57] Zur den Schimären des wissenschaftlichen Diskurses zählen

- *Vorhanden sein und Verfügbarkeit der richtigen Daten*: Das heisst, die Daten sind in der Welt und man muss sie nur einsammeln. **Kommentar**: In der Praxis sind die vorhandenen Daten nach wie vor häufig äusserst unvollständig.⁸ Zudem wird der Zugriff stark eingeschränkt. Bekannt ist, dass Betreiber von digitalen sozialen Netzen der Wissenschaft nur kleine Auszüge zur

⁸ Beispielhaft wird dies erläutert in OFFENHUBER, DIETER / RATTI, CARLO, Drei Mythen über Smart Cities, In: Geiselberger, Heinrich und Moorstedt, Tobias, Big Data — Das neue Versprechen der Allwissenheit, Suhrkamp, edition unseld (2013).

Verfügung stellen, ausgenommen den eigenen Wissenschaftlern.

- *Objektive Gültigkeit der Rohdaten*: Das heisst, wer Rohdaten verwendet, soll sich um deren Qualität keine Sorgen machen müssen, weil die «echt» sind und keinen «Bias» besitzen. **Kommentar**: Das Erzeugen von Daten enthält immer eine Selektion. Datenwahrheit und Wahrheit sind zweierlei. Wer diesen Bias nicht kennt arbeitet oft besser mit gut aufbereiteten Daten.
- *Statistisches Ausmitteln von Ausreissern*: Das heisst, Datenfehler werden durch Datenmassen kompensiert. **Kommentar**: Ob dies geschieht oder nicht hängt vom Daten-Bias ab.
- *Befreiung vom Joch der Modelle (oder: Das Ende der Theorie⁹)*: Dies ist die vielleicht derzeit wichtigste Schimäre. Sie meint, aufgrund der grossen Datenmassen und von Big Data ist es in Zukunft in der Forschung möglich, auf Modelle der Wirklichkeit zu verzichten. **Kommentar**: Damit wird eine wissenschaftsfreie Wissenschaft angestrebt in der Sachkenntnis kaum mehr Bedeutung hat. Solch eine Reduktion auf das allein selig machende Wundermittel hat in der Geschichte im Kleinen wie im Grossen meist zum Desaster geführt. Zudem ist die Vorstellung ziemlich erschreckend, dass in Zukunft Exekutiventscheidung auf Big Data Empfehlungen basieren, beispielsweise bei Polizeieinsätzen, die keiner versteht.
- *Jeder profitiert*: Das heisst, von Big Data profitieren zwar manche mehr als andere, doch alle haben etwas davon. **Kommentar**: Die Ungleichheit beim Profitieren von Big Data ist sehr gross. Bei den dominierenden Diensten der Internetgesellschaft werden einige wenige Multi-milliardäre, während die Masse Gratisdienstleistungen bekommt. Für diese Gratisdienstleistungen muss aber beim Konsum anderer Güter bezahlt werden. Zusätzlich werden Arbeitsplätze vernichtet. So dass die Nutzenbilanz für die Gesellschaft als Ganzes sehr zwiespältig ausfällt.
- *Alles freiwillig (und implizit: alles ist lokal)*: Das heisst, dass jeder entscheiden kann, ob er Daten liefern will oder nicht und dass seine Entscheidungen nur Auswirkungen auf ihn haben. **Kommentar**: In vielen Bereichen werden Dienstleistungen verweigert oder stark eingeschränkt, wenn nicht in die Datenweitergabe eingewilligt wird. Zudem betreffen die eigenen Daten gerade bei Big Data auch immer andere. Und nicht zuletzt sind alle Big Data Werkzeuge, die entwickelt werden auch für kriminelle Aktivitäten und für staatlichen Terror nutzbar. So wie Banken und Versicherungen die Vertrauenswürdigkeit von Kunden prüfen, können auch kriminelle Organisation die Vertrauenswürdigkeit ihrer Mitglieder und Staaten die Angepasstheit ihrer Bürger prüfen.

[Rz 58] Dazu kommen in der Praxis noch viele weitere Fiktionen. Als Begleiter der wissenschaftlichen Schimäre des modellfreien Big Data tritt beispielsweise in der Praxis die Illusion auf, dass durch Big Data Domänenwissen überflüssig wird. Weitere Illusionen sind:

- *Transparente Datenquellen*: Das heisst, wir wissen in der Regel woher die verwendeten Daten kommen, was sie bedeuten und wie hoch ihre Qualität ist. **Kommentar**: Das ist ganz selten der Fall.
- *Transparente Big Data Anwendungen*: Das heisst, Big Data ist verstehbar und nachvollziehbar, auch wenn man die Algorithmen nicht kennt. **Kommentar**: Big Data verführt dazu, die dadurch produzierten Empfehlungen nicht zu hinterfragen. In vielen Fällen hat niemand den Überblick, was dabei wirklich geschieht — insbesondere dann wenn alte Software-Module oder Software-Module anderer als Plug-in verwendet werden und die verwendeten Algorithmen

⁹ WEICHERT, THILO, Das Ende der Theorie — die Datenschwemme macht wissenschaftliche Methoden obsolet, In: [FN 7].

men bereits in Vergessenheit geraten sind.

- *Schutz der Privatsphäre durch Anonymisierung*: Das heisst, wenn personenbezogene Daten anonymisiert werden, dann ist die Privatsphäre geschützt und das DSGVO nicht anzuwenden. **Kommentar**: Big Data Profiling strebt genau die De-Anonymisierung an.
- *Jeder kann ...*: Das heisst, dass es keine besonderen Fähigkeiten braucht, weil die Big Data Nutzung als Dienstleistung bezogen werden kann. **Kommentar**: Für einfache Anwendungen ist das richtig, für komplexe sicher nicht. Die Ansprüche an die verschiedenen fachlichen Fähigkeiten, die benötigt werden, sind so hoch, dass es für einige Anwendungsbereiche gar keine Dienstleister gibt, deren Qualifikation über jeden Zweifel erhaben wäre. Für die Richtigkeit der bereitgestellten Kontextinformationen und die Valorisierung der Resultate muss sowieso jeder Nutzer selber gerade stehen.
- *Big Data = umfassendes Wissen über die Welt (oder auch: Kontinuitätsfiktion)*: Das heisst, dass wir mit Big Data aus Daten der Vergangenheit nicht nur zuverlässige allgemeine Zukunftsprognosen machen können, sondern auch ganz konkret zukünftige Interessen von Menschen in Politik und Wirtschaft voraussagen können. Denn (so die Meinung) von der Vergangenheit kann man auf die Zukunft schliessen, wenn man nur genügend viele Aspekte berücksichtigt. **Kommentar**: Aus digitalen Daten über die Vergangenheit lässt sich selbst die Vergangenheit nur rudimentär konstruieren, geschweigen denn die Gegenwart oder gar die Zukunft. Zudem stellt alleine schon die Erwartung, dass alles prognostizierbar ist, eine Innovationsbremse dar, nicht nur weil sie den Mut für disruptive Innovationen nimmt, sondern weil kreativitätsfreie, billige» datenbasierte Pseudoinnovation als allzu verführerisch erscheinen.
- *Big Data = direkter Nutzen*: Das heisst, wenn die ersten drei Schritte des Big Data Prozesses gut durchgeführt werden, schafft das Nutzen. **Kommentar**: In der Praxis gibt es grosse Widerstände gegen die Nutzung zusätzlichen Wissens, die über die Widerstände aufgrund von fehlenden Fähigkeiten für den Umgang mit Big Data weit hinausgehen. Es gibt natürlich Fälle, wo Fakten nur akzeptiert werden, weil sie von Computern stammen (weshalb die siebte Form von Big Data die Erkenntnisfassade ist), oft aber scheitert umgekehrt Big Data in der Organisationspraxis entweder an der Ablehnung von Faktenbasierung oder an der Ablehnung der Computerprogramme als Faktenquelle.

[Rz 59] Geht man in konkrete Nutzungsbereiche hinein, könnte man die Auflistung von Paradigmenwechseln, Schimären und Fiktion problemlos erweitern. Denn Big Data verändert in vielerlei Hinsicht ganz grundsätzlich wie wir arbeiten und planen. Ein Beispiel ist die Schimäre der intelligenten Stadt. Sie tritt mit wachsender Popularität in der Stadtentwicklung auf, führt dort zu grossen Einflussverschiebungen und schafft in der Praxis neue grosse Risiken, weil auch hier das existierende Domänenwissen nicht mehr berücksichtigt wird¹⁰. Exemplum docet, exempla obscurant.

[Rz 60] Bei aller notwendig kritischen Einstellung gegenüber den hier zitierten und anderen Schimären und Fiktionen: In ihnen steckt jeweils ein grosses Körnchen Vernunft. Wir können aus den frei verfügbaren Daten viel mehr Wissen extrahieren als wir bisher tun — und das brächte oft tatsächlich auch Nutzen für alle und ist oft weniger schwierig und weniger riskant als viele annehmen. Modelle können auch ein Fetisch sein und sind nicht immer notwendig. Das Verbergen der Rohdaten zählt tatsächlich zu den grossen Sünden der Wissenschaft und sollte im Sinne ei-

¹⁰ Interview mit GREENFELD, ADAM, Smart Phone, Smart Bomb, Smart City, In: Das grosse Rauschen: Warum die Daten-gesellschaft mehr Menschenverstand braucht, Verlag Neue Zürcher Zeitung (2013).

ner transparenten Wissenschaft überwunden werden (aber eben genau deshalb, weil Rohdaten in aller Regel nicht bias-frei sind). Anonymisierung ist sehr wohl in vielen Bereichen ein taugliches Mittel zum Schutz der Privatsphäre, es muss nur gut angewendet werden und es darf nicht als unbedingt zuverlässige Festung betrachtet werden. Et cetera ...

3.5 Schlussfolgerungen

[Rz 61] Das Bewältigen von Paradigmenwechseln stellt fast immer eine grosse Herausforderung dar. Fehler und grober unbeabsichtigter Missbrauch sind ebenso zu erwarten wie die Unfähigkeit, mit absichtlichem, gar kriminellem Missbrauch sicher umzugehen. «Denn wir wissen noch nicht, was wir tun.» Vorsicht ist also geboten. Die skizzierten Schimären und Fiktionen illustrieren ein wenig die Gefahren die auf uns lauern.

[Rz 62] Grundsätzlich verlangen neue Paradigmen, die obendrein durch eine mächtige Technologie unterstützt werden, ein Überdenken bisheriger Regulierungen. Welchen Sinn macht das Konzept der personenbezogenen Daten, wenn De-Anonymisierung von anonymisierten Daten in Zukunft Alltag sein wird? Was heisst überhaupt personenbezogene Daten, wenn das Verhalten anderer durch Dritte auf mich bezogen wird? Was heisst richtige Daten, wenn es darauf nicht mehr ankommt, ob Daten richtig sind? Wie wollen wir angesichts der entstehenden Datenuniversen, denen man als unfreiwilliger Datenlieferant fast nicht entkommen kann, die informationelle Selbstbestimmung effektiv sicherstellen? Wie viel Diskriminierung wollen wir anhand der radikalen Personalisierung zulassen? Muss uns kümmern, wenn anderswo unsere Vertrauens-technologien für staatlichen Terror genutzt werden? Wo wollen wir in Regierung, Verwaltung, Infrastrukturbewirtschaftung, Medizin und Justiz das grosse Potential von Big Data nutzen? Und nicht zuletzt — wenn wir selber es sind, die darüber entscheiden, wie viel Datenspuren wir für die Big Data Nutzung produzieren, wäre es nicht sinnvoll, in den Schulen Wissen und Kenntnisse zu vermitteln, die eine echte Selbstbestimmung ermöglichen?

4 Regulierungsbereiche

[Rz 63] Um nochmals einige wichtige Beobachtungen — ohne Anspruch auf Vollständigkeit — zusammenzufassen, die häufig zutreffen: Die Rohdaten sind nicht objektiv. Das Gesetz der grossen Zahlen ist bei systemischen Fehler wirkungslos. Ohne gute Modelle und vielfältige Fachkompetenzen geht es nicht. Ohnehin gibt es die Daten, die es wirklich bräuchte, entweder gar nicht oder sie sind nicht verfügbar. Von denen, die wir nutzen können, kennen wir häufig weder ihre genaue Bedeutung noch ihre Qualität. Ohnehin sind die wesentlichen Aspekte der Zukunft nicht vorhersagbar und das Denken wird uns schon gar nicht abgenommen. Dafür wird der Datenschutz durch De-Anonymisieren verletzt. Und fair ist das alles nicht.

[Rz 64] Big Data klingt nach einer Sache, die echt Spass macht! Zu Recht. Aber nicht immer rechtmässig. Und schon gar nicht immer im Interesse vieler direkt oder indirekt Betroffener. Manchmal nicht einmal im Interesse derer, die es tun. Wir sehen in drei Bereichen den Bedarf, über neue Regulierungen politisch zu diskutieren: Datenschutz, Diskriminierungsverbot und faire Wirtschaftsförderung.

4.1 Datenschutz

[Rz 65] Die Grenzen zwischen anonymisierten und personenbezogenen Daten lösen sich auf ebenso wie die Grenzen zwischen normalen Daten und besonders schützenswerten Daten im Sinne des Schweizer DSG. Die Verstösse gegen das DSG in der Schweiz wie gegen die Datenschutzgesetze in EU-Mitgliedsstaaten wachsen dabei schleichend, wobei die Grenzziehung nicht einfach ist. Aus Sicht des Paradigmenwechsels 6, das heisst eines transdisziplinären Datenschutzverständnis, lassen sich folgende Empfehlungen geben und offene Fragen identifizieren:

4.1.1 Respektierung des Datenschutzes in Organisationen unter Schweizer Recht

[Rz 66] Aktuell bestehen die Hauptprobleme in der oben angesprochenen Auflösung der Abgrenzungen, die im DSG eine zentrale Rolle spielen. Unternehmen und andere Organisationen können sich von sich aus proaktiv damit auseinandersetzen — und wir empfehlen das vor allem in Bezug auf mögliche Reputationsrisiken — aber auch eine gesetzliche Verpflichtung zu dieser Auseinandersetzung ist denkbar und sollte diskutiert werden.

1. Der erste Schritt zur Selbst- und zur Fremderkenntnis ist **Transparenz** darüber, was im Big Data Prozess passiert. Eine öffentliche Transparenz kollidiert aber mit berechtigten wie auch mit unberechtigten Geschäftsinteressen. Einerseits soll die Konkurrenz nicht sehen, wie man vorgeht, andererseits soll den Kunden nicht bewusst werden, was mit ihnen und ihren Daten geschieht. Bei lediglich organisationsinterner Transparenz sind die Interessenskonflikte geringer. Aufgabe der Rechtsinformatik-Forschung ist es, ein Konzept für Transparenz des Big Data Prozesses zu entwickeln, das alle berechtigten Interessen berücksichtigt. Anschliessend ist zu diskutieren, inwieweit die Schaffung von Transparenz von einem neuen Datengesetz, das das Datenschutzgesetz ergänzt¹¹, verordnet werden soll und kann.
2. Der notwendige zweite Schritt kann nicht staatlich verordnet, höchstens durch Information beworben werden. Er besteht in der Etablierung von **Problembewusstsein** in der Unternehmenskultur¹². Wenn Vertreter von Big Data Technologieanbietern allen Ernstes noch immer Datenschutz als grosses Problem Europas anprangern¹³, zeigt das, wie hoch das Risiko ist, bei der Einführung von Big Data durch die Zusammenarbeit mit Anbietern eine datenschutzfeindliche Kultur sich einzuhandeln. Sinnvoll ist es darum, die Schaffung des Problembewusstseins mit einer Ausbildung der Mitarbeitenden zu verbinden.
3. Auf der Basis von Transparenz und Problembewusstsein macht die Durchführung einer **Problem- und Risikoanalyse** Sinn, nebst nachfolgender technisch-organisatorischer Problembeseitigung. Beides setzt nicht zwingend die ersten beiden Schritte voraus, wird durch diese aber wesentlich erleichtert.
4. Der vierte Schritt ist die Verankerung in den Führungsstrukturen der Organisation durch die Definition von **Good Practices**, die in den organisationsinternen Governance-Strukturen als

¹¹ EPINEY, ASTRID, Big Data und Datenschutzrecht: Gibt es einen gesetzgeberischen Handlungsbedarf?, in: Jusletter IT 21. Mai 2015.

¹² Eine gute Definition von Unternehmenskultur findet sich im Klassiker SCHEIN, EDGAR H., *Organizational Culture and Leadership*, 4th Edition, John Wiley & Sons (2010): «[...] ein Muster gemeinsamer Grundprämissen, das die Gruppe bei der Bewältigung ihrer Probleme externer Anpassung und interner Integration erlernt hat, das sich bewährt hat und somit als bindend gilt; und das daher an neue Mitglieder als rational und emotional korrekter Ansatz für den Umgang mit Problemen weitergegeben wird.»

¹³ Dies geschieht vor allem im Lobbying off records.

Vorgaben oder Richtlinien eingebunden werden. Derzeit gibt es noch kein klares Bild, was denn konkret Good Practices in Bezug auf Big Data und die Wahrung der Privatsphäre Betroffener sind (abgesehen vom Schaffen von Problembewusstsein, das immer eine zentrale Bedeutung hat). Auch dies ist eine Forschungsaufgabe für die Rechtsinformatik. Die Einführung von neuer Bürokratie ist sicher NICHT die Lösung, aber eine saubere Definition von Rollen, Aufgaben und Prozessen kann von grossem Nutzen sein. Im Fall der Verwaltung wäre eine Identifikation solcher Good Practices durch eCH¹⁴ angezeigt.

5. Sinnvoll ist es weiters, **Audits** ein- und durchzuführen, die prüfen, ob die noch zu definierenden Good Practice Standards eingehalten werden. Diskutierenswert ist es auch, diese für einige Akteure beziehungsweise in einigen Bereichen (wie dem Gesundheitswesen) verpflichtend vorzuschreiben. Allerdings sind solche Audits durchaus auch im Interesse der Betroffenen, weil sie sie vom Verdacht des vorsätzlichen Missbrauchs befreien.

[Rz 67] Zum ersten Schritt sei noch angemerkt, dass die Linked Data Technologie ergänzt um semantische Standards das Schaffen von und Arbeiten mit Transparenz wesentlich erleichtert. Sie ermöglicht es, das Identifizieren von Quellen und von Folgerungen mit vernünftigem Aufwand durchzuführen und so mehr Nutzen aus der ansonsten oft ungenutzt bleibenden Transparenz zu ziehen.

[Rz 68] Zum zweiten Schritt: Mit der Durchführung einer internen Ausbildung demonstrieren Organisationen, dass ihnen die Einhaltung des Datenschutzes und eventuell weitergehender ethisch motivierter Selbstbeschränkungen ernst ist. Dies kann im Missbrauchsfall durch einzelne Mitarbeitende als Beleg genutzt werden, dass der Missbrauch nicht teil der Unternehmenspolitik war.

[Rz 69] Zum dritten und vierten Schritt: Eine strategische Option ist, sich einer freiwilligen Selbstbeschränkung zu unterwerfen und dies allenfalls einem Audit — wie im fünften Schritt besprochen — zu unterwerfen, um dann gegen innen (Mitarbeitende) und gegen aussen (Kunden, Investoren, Regulatoren) diese ethischen Standards als Marketinginstrument zu nutzen.

[Rz 70] Die fünf Schritte haben keinen Anspruch auf Originalität. WEBER und OERTLY empfehlen konkret im gleichen Kontext wie hier die vier Schritte Risikoanalyse, Festlegung einer Strategie, Festlegung einer Bearbeitungspolitik und Einführung adäquater Verfahren¹⁵, was inhaltlich sehr nahe verwandt mit unseren Schritten 3 bis 5 ist. Obendrein betonen sie wie wir die grosse Bedeutung von Transparenz.

[Rz 71] Sinnvoll wäre ausserdem, ein dediziertes Risikomanagement im Schritt 3 zu starten, doch die Praxis zeigt leider, dass die Einführung eines zusätzlichen Risikomanagements (zusätzlich zu den jeweils bereits etablierten Formen von Risikomanagement) für Organisation fast immer sehr schwierig ist.

4.1.2 Wahrung der Rechte des Individuums

[Rz 72] Die Wirklichkeit untergräbt den Wert des Datenschutzes tagtäglich: Menschen geben über sich freiwillig vieles einer breiten Öffentlichkeit preis. Zudem akzeptieren sie grossmehr-

¹⁴ eCH ist die Schweizer Standardisierungsorganisation für E-Government, www.ech.ch.

¹⁵ WEBER, ROLF H. / OERTLY, DOMINIC, Aushöhlung des Datenschutzes durch De-Anonymisierung bei Big Data Analytics?, in: Jusletter IT 21. Mai 2015.

heitlich Geschäftsbedingung — oft mangels guter Alternativen — in denen sie Leistungsanbieter das Recht auf Weiternutzung ihrer Daten einräumen, beispielsweise im Kreditkartenzahlungsverkehr. Viele sind sich bei populären Internet-Dienstleistungen auch gar nicht bewusst, dass sie ein Tracking «einkaufen». Dies zu unterbinden ist in der Praxis sowieso äusserst schwierig, unter anderem weil es eine offene Frage ist, welche Gerichte jeweils zuständig sind.¹⁶ Es ist weiters bereits Folklore, dass die Überwachung durch Geheimdienste zwar nicht vollständig stattfindet, aber nicht vermieden werden kann und dass eine Weitergabe von geschäftsrelevanten Daten an Unternehmen und andere Organisationen nicht unwahrscheinlich ist. Anbieter im Internet identifizieren zudem einzelne Internetbrowser und legen zu diesen Historien an, die sämtliche Daten enthalten, die über den Browser bekanntgegeben werden. So können sie — mit einigen Ungewissheiten bezüglich der Zahl der Nutzer eines Browsers — ein Profiling durchführen und dafür oft sogar Daten von anderen Sites integrieren. Wir sind also bereits gläserne Bürger, Kunden, eventuell auch Patienten, Datenschutz hin oder her.

[Rz 73] Um die Situation etwas zu verbessern bieten sich unter anderem staatliche Massnahmen in der Bildung und Rechtssetzung, sowie nutzerorientierte Dienstleistungen und technische Lösungskonzepte an.

- **Öffentliche Information und öffentliche Bildungsangebote:** Aufklären über die Risiken und das Befähigen zur informationellen Selbstbestimmung sollte Teil der Schulbildung sein und sinnvollerweise schon in der Primarschule beginnen.
- **Schaffung von Informationsdienstleistern:** Diese würden unsere Daten in unserem Sinn verwalten ähnlich wie Finanzdienstleister unsere finanziellen Mittel verwalten und können entweder kommerzielle Dienstleister, staatliche Dienstleister oder Genossenschaften sein. Ein Beispiel für letzteres ist der Versuch des Vereins «Daten und Gesundheit» (www.datenundgesundheit.ch) eine Gesundheitsdaten-Genossenschaft aufzubauen.
- **Rechtliche Absicherung des Rechts auf Selbstnutzung der auf die eigene Person bezogenen Daten:** Dies würde beinhalten, dass Datensammler die gesammelten Daten an die Betroffenen bei Nachfrage herausgeben müssten und wird vom Verein Daten und Gesundheit zur Diskussion gestellt.¹⁷
- **Besserer Technischer Schutz:** Privacy-by-Design ist ein Designprinzip und Privacy-by-Default ist ein Konfigurationsprinzip für Software. Ihre Umsetzung sind aktuell Gegenstand von Forschung und Entwicklung im Horizon2020-Programm der EU. Im Kern geht es dabei darum, ein Konzept für elektronische Identität beziehungsweise elektronische Identitäts- und Zugriffsmanagement (eID bzw. IAM) zu entwickeln, das dem Betroffenen maximale Kontrolle über die Verteilung von auf seine Person bezogenen Daten erlaubt. Dabei spielt insbesondere das Konzept einer nicht-verfolgbaren vertrauenswürdigen anonymen elektronischen Identität eine zentrale Rolle. Diese ist viel weitergehend realisierbar als man meint, hat aber ihre Grenzen bei der Benutzung digitaler Geräte.¹⁸

[Rz 74] Bei der Leitung von insgesamt vier Experten-Workshops zum Thema elektronische Identität

¹⁶ <http://www.zeit.de/digital/datenschutz/2015-04/facebook-datenschutz-klage-schadensersatz-oesterreich>, alle Internetquellen zuletzt besucht am 17. Mai 2015.

¹⁷ GÄCHTER, THOMAS, Vortrag Das Recht auf Kopie — Grundvoraussetzung für die informationelle Selbstbestimmung am 26. Januar 2015 bei einer Veranstaltung des Vereins im Hotel Kreuz in Bern.

¹⁸ RODUNER, CHRISTOF, Citizen Controlled Data Protection in a Smart World, Master's Thesis, University of Zurich (2003) <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.202.456&rep=rep1&type=pdf>.

tität im letzten Jahr haben wir immer wieder festgestellt, dass die technisch natürlichste und aus Datenschutz-Perspektive korrekte Lösung eines durch den Betroffenen kontrollierten Verteilens von Eigenschaftszertifikaten als Schritt hin zum gläsernen Bürger aufgefasst wird — nicht von allen, aber auch von Spezialisten. Hier besteht ein krasses Missverhältnis zwischen der Akzeptanz des allgegenwärtigen Missbrauchs im E-Commerce und der Ablehnung sicherer Lösungen, die offensichtlich vielen Angst einflößen. Deshalb kommt der ersten Massnahme eine Schlüsselbedeutung zu. Neben einer breiten Aufklärung ist auch die Vermittlung eines multidisziplinären Fachverständnisses für Spezialisten notwendig.

4.2 Beschränkung der Datennutzung und Datenbereitstellung

[Rz 75] Eine bereichsübergreifende Diskussion dessen, was bei der Big Data Nutzung alles erlaubt sein soll und was nicht, sofern der Datenschutz eingehalten wird, erscheint wenig zweckmässig, da je nach Bereich unterschiedlich grosse Risiken existieren.

4.2.1 Beispiele möglicher Diskriminierungen

(1) Beispiel Verkauf

[Rz 76] Wie oben schon kurz skizziert, ist es möglich aus dem Profiling die Zahlungsbereitschaft eines Kunden abzuleiten und so den Verkaufspreis zu maximieren, ohne den Kunden zu verlieren. Bei einer umfassenden Umsetzung zahlt jeder einen anderen Preis. Vorstellbar wäre aber auch, ein Rabattsystem mit individuellen Rabatten und generell höheren Preisen so einzuführen, dass der Kunde dem Anschein nach durch das Profiling profitiert.

[Rz 77] Es sollte deshalb politisch diskutiert werden, ob diese De-facto-Abschaffung des Fixpreises im Handel uneingeschränkt zulässig sein soll.

(2) Beispiel Versicherungswirtschaft

[Rz 78] Während früher nur Risiken für alle bzw. für relativ grosse Gruppen berechnet werden konnten, mit Ausnahmen wie beispielsweise zukünftige Gesundheitsrisiken im Fall früherer Krankheiten, so können durch Big Data individuelle Risiken einer Person gut geschätzt werden und diese Schätzungen werden immer mehr zur Basis der Kostenfestlegung bei Versicherungen. Da die allgemein beschaffbaren Daten über eine Person oft nicht wirklich ausreichen, bieten Versicherungen mittlerweile Preisnachlässe für jene an, die sich von ihnen überwachen lassen, beispielsweise bei Autoversicherungen. Diese Entwicklung hat mehrere kritische Aspekte, die wir teilweise schon besprochen haben: die Auflösung des Solidaritätscharakters von Versicherungen, eine Preisdiskriminierung die auf dem Verhalten anderer beruht (weil selbst beim Fahrten-tracking bei Autoversicherungen ja nicht nur das Fahrverhalten beurteilt wird, sondern auch die Risikohäufigkeiten auf den befahrenen Strecken), die Infragestellung der Versicherungsbranche an sich (weil Versicherungen für ein berechnetes Schicksal wenig Sinn machen) und den Aufbau einer Überwachungswirtschaft als Pendant zum Überwachungsstaat.

[Rz 79] Es sollte deshalb politisch diskutiert werden, inwieweit alle Arten von Daten in der Versicherungswirtschaft tatsächlich genutzt werden dürfen und ob tatsächlich Versicherungsverträge mit Überwachung uneingeschränkt zulässig sind. Je nach Art der Versicherung wird man dabei möglicherweise zu unterschiedlichen Ergebnissen kommen.

(3) Beispiel Personalführung

[Rz 80] In Zukunft wird es möglich sein, aus den Daten über die Gene eines Menschen viele Schlussfolgerungen über seine Krankheitsrisiken abzuleiten. Wer gute Gene hat, hat deshalb einen Anreiz, genau dies bei einer Bewerbung anzugeben, in welcher Form auch immer. Doch auch wenn die Daten über die Gene von einigen freiwillig bereitgestellt werden, schafft das in den meisten Fällen Nachteile für andere Bewerber und Mitarbeiter des Arbeitgebers, die nicht ebenfalls Daten bereitstellen und ebenfalls gute Gene besitzen.

[Rz 81] Das Beispiel zeigt besonders drastisch, wohin es führt, wenn das freiwillige Bereitstellen von eigenen Daten uneingeschränkt erlaubt sein soll. Denn in diesem Beispiel ist eine Diskriminierung durch das Verbot Aussagen über Genome zu verwenden praktisch nicht kontrollierbar.

4.2.2 Schlussfolgerungen

[Rz 82] Es stellt sich die politisch zu beantwortende Frage, ob wir big-data-basierte Diskriminierungen zulassen wollen oder nicht. Im Fall dass wir dies in einigen Sektoren nicht wollen oder mindestens beschränken wollen, können wir dies teilweise durch entsprechende Anti-Diskriminierungsgesetze erreichen. In manchen Fällen werden gesetzliche Diskriminierungsbeschränkungen nicht ausreichen. In diesen Fällen besteht eine Option darin, die freiwillige Weitergabe der eigenen Daten zu beschränken.

[Rz 83] Die Zusammenhänge sind allerdings komplex. Es empfiehlt sich, regulierende Eingriffe ins System mit Bedacht durchzuführen. Die scheinbare Langsamkeit der politischen Prozesse ist hier ein Vorteil, sofern früh — also jetzt — mit der politischen Diskussion begonnen wird. Sinnvoll ist jedenfalls Konsumentenschutz und Preisüberwachung für das Thema zu sensibilisieren und eventuell auch von Seiten TA-Suisse¹⁹ die Entwicklung zu monitoren.

4.3 Staatliche Förderung von Big Data

[Rz 84] Big Data ist nicht nur ein grosses Thema für die Rechtspolitik, sondern auch eines für die Wirtschafts- und Bildungspolitik. Big Data bietet grosse Chancen und es ist — je nach politischer Einstellung — eine Aufgabe von Regierung und Verwaltung, diese Chancen zu stärken oder auch nicht. Bevor man dies diskutiert, sollte man sich bewusst machen, dass für Big Data zweierlei Infrastrukturen notwendig sind: Daten und Hochleistungsrechner. Bei beiden besteht ein Monopolisierungsrisiko der Art, dass einige wenige diese Infrastruktur kontrollieren. Potentiell ist das Risiko bei Daten grösser als bei Hochleistungsrechnern. Zugleich sind bei beiden die Investitionskosten sehr hoch und die ROI-Zeiten lang. Es ist also nicht selbstverständlich, dass Wirtschaftsunternehmen die nötigen Investitionen wirklich tätigen werden. Gerade in der Schweiz wäre zwar genügend Kapital vorhanden, um fünf bis zehn Milliarden privatwirtschaftlich in ein Big Data Center zu investieren, die Wahrscheinlichkeit, dass dies tatsächlich geschieht ist aber als gering einzuschätzen. Wenn aber die Chancen Schweizer Unternehmen — insbesondere von kleinen Unternehmen — schlecht sind, zu günstigen Konditionen und ohne Anbieter-Lock-in Big Data zu nutzen, so behindert dies das volkswirtschaftliche Wachstum. Deshalb stellt sich die

¹⁹ TA-Suisse ist das Kompetenzzentrum für Technologiefolgenabschätzung der Schweizer Akademien der Wissenschaften, <https://www.ta-swiss.ch/>.

Frage, wie der Staat zur Entwicklung eines Big Data Ökosystems²⁰ beitragen könnte.

[Rz 85] Neben den konventionellen wirtschaftspolitischen Überlegungen, die hier anzustellen sind, gilt es noch drei besondere Aspekte zu beachten

- Gerade aus Sicht von Privacy-by-Design Überlegungen, wird eine sichere, skalierende und günstige eID-Lösung benötigt, die das IAM optimal im Sinne aller rechtmässig Betroffenen unterstützt und gleichzeitig Missbrauch weitestmöglich einschränkt.
- Eine wesentliche Quelle für das Big Data Ökosystem sind Verwaltungsdaten, die als Open Government Data zur Sekundärnutzung veröffentlicht werden
- Big Data kann auch von der Verwaltung und in vielen anderen Bereichen des öffentlichen Sektors eingesetzt werden, der damit nicht nur Datenlieferant ist für das Ökosystem, sondern auch ein grosser Nutzer von Daten und Big-Data-Diensten des Ökosystems

[Rz 86] Zum ersten Punkt: Für das oben skizzierte Gesundheitsdatensammelbank-Projekt des Vereins Daten und Gesundheit wäre es zum Beispiel eine zusätzliche Erschwernis, wenn darin auch noch das eID-Problem gelöst werden müsste, abgesehen davon dass ein einzelner Player kaum die Nachhaltigkeit seiner Lösungen sichern kann.²¹

[Rz 87] Zum zweiten Punkt: Der Bundesrat hat am 2014 eine Big Data Strategie beschlossen²² und fördert seither aktiv die Bereitstellung von Big Data. Es empfiehlt sich, insbesondere bei der Veröffentlichung von Wirtschaftsdaten vorwärts zu machen.

[Rz 88] Zum dritten Punkt: Aus Sicht der Wirtschaftspolitik auf Bundesebene sollte man als die drei wichtigsten Stakeholder von Big Data die Unternehmen, die Forschungsinstitutionen und das WBF selber ansehen. Aber auch für alle anderen Politikbereiche wird Big Data zukünftig ein wichtiges strategisches Instrument sein. Bei der Energie- und Verkehrspolitik ist das offensichtlich, ebenso bei der Gesundheitspolitik und in der Stadtentwicklung, aber es gilt praktisch für alle Politikbereiche. Deshalb sollte Big Data auch bei der Weiterentwicklung der öffentlichen Verwaltung eine zentrale Rolle spielen — vor allem um die Verwaltung effektiver zu machen und Fehlplanungen zu vermeiden. Wie DUMERMUTH an der Big Data Governance Tagung 2014 in Bern ausführte²³, sollte die Verwaltung in die Ausbildung investieren, eine Strategie entwickeln und Governance-Leitlinien für den Einsatz von Big Data formulieren. Das Nutzenpotential ist sehr gross für die Verwaltung, wie die Risiken vielfältig sind, weshalb es wenig sinnvoll ist nur ein Laissez-faire zu praktizieren. Mindestens mittelfristig sollte es auch möglich sein, die szenario-basierte Langfristplanung des Bundesrats um Big Data Prognosen zu erweitern, beziehungsweise big-data-basiert die Szenarien zu fokussieren.

[Rz 89] Nicht in obiger Aufzählung aufgeführt sind weitere konventionelle Aspekte der Wirtschaftsförderung, die aber auch bei Big Data eine grosse Rolle spielen. Erstens: Ausbildung ist nicht nur notwendig, um die Menschen zum Schutz ihrer Privatsphäre zu befähigen, sondern auch um das Prosperieren der Big Data Nutzung zu fördern. Ausbildung für die Big Data Nut-

²⁰ Wie Big Data ist auch der Begriff des soziotechnischen Ökosystems nicht klar definiert. Wir verstehen darunter das Zusammenspiel der technischen Ressourcen und der sie nutzenden Akteure im politisch-rechtlichen, sozialen und ökonomischen Kontext.

²¹ Da dies auch für ein einzelnes Land kaum möglich ist, wenn die internationale Zusammenarbeit nicht blockiert werden soll, beteiligt sich die Schweiz am europäischen Large Scale Pilot STORK 2.0, <https://www.eid-stork2.eu/>, und leitet in diesem Projekt den e-Banking Piloten und das Service-Design.

²² <http://www.egovernment.ch/umsetzung/00881/00883/index.html?lang=de>.

²³ Siehe http://www.rechtswissenschaft.ch/Unterlagen_2014/Dumermuth_Tagung%202014.pdf. MARTIN DUMERMUTH, Chancen und Risiken von Big Data (Podcast), in: Jusletter IT 21. Mai 2015.

zung sollte deshalb nicht nur in den Unternehmen und in der Verwaltung stattfinden, sondern auch in den Ausbildungs- und Weiterbildungsprogrammen der Hochschulen angeboten werden. Zweitens: da wir erst am Anfang der Big Data Nutzung stehen, ist es sinnvoll in die (Angewandte) Grundlagenforschung zu Big Data zu investieren und nachfolgend die Angewandte Forschung dazu zu fördern. Überlegenswert ist weiter die Schaffung eines von einigen Hochschulen gemeinsam betriebenen öffentlichen Big Data Kompetenz- und Rechenzentrums, das die rechtskonforme, wirtschaftlich nützliche und gesellschaftlich nachhaltige Nutzung von Big Data fördert.

4.4 Die Rolle der Wissenschaftä

[Rz 90] Wir haben oben schon angesprochen, dass die Förderung von Wissenschaft sinnvoll ist. Doch wie so oft in der Geschichte ist die Rolle der Wissenschaft ambivalent und es besteht ein grosses Risiko, dass am Ende sie selber unter die Räder kommt.

[Rz 91] Ein grosser zukünftiger Problembrocken ist beispielsweise IoT, das Internet der Dinge. Einige diesbezügliche Pläne von Wissenschaftlern und Entwicklern — an Hochschulen wie in der Privatwirtschaft — würden gut zu einem James Bond Film-Bösewicht alten Stils passen. Das soll (um nur ein Beispiel aus mündlicher Quelle zu bringen) in Zukunft beispielsweise eine Gesundheitslandkarte für die Welt erstellt und Arbeitgebern zur Verfügung gestellt werden, damit diese wissen, wie gesund oder krank die Menschen in einem Stadtviertel oder Dorf sind, aus dem ein Stellenbewerber kommt. Das Ziel der Innovation: Förderung von Diskriminierung im Arbeitsmarkt (womit man viel Geld verdienen will), die vermutliche Wirkung fortschreitende Ghettoisierung. Leider ist es in der IoT-Community weit verbreitet, Ausnahmen bestätigen wie immer die Regel, staatliche Regulierung als Behinderung des Fortschritts abzulehnen. Dies geschieht unter anderem gern mit dem Argument, dass auch das Internet sich ohne staatliche Regulierung prächtig entwickelt hat. Vergessen wird dabei, wie gross die Geschwindigkeitsunterschiede zwischen beiden Fällen sind und wie sehr sich die Ziele unterscheiden. Das Internet brauchte lange und dahinter standen in den ersten zwanzig Jahren vor allem idealistische Ziele (von Zusammenarbeit in der Wissenschaft bis zum Kampf gegen die allgegenwärtige Werbung) und militärische Ziele (zuverlässige Kommunikationsnetze). Hinter Big Data stehen kommerzielle Ziele und Machtinteressen bis hin zum Versuch, die Welt zu kontrollieren. Darum ist bei Big Data ein Laissez-faire viel gefährlicher als es einst beim Entstehen des Internets war.

[Rz 92] Allerdings steht den oben skizzierten Umtrieben eine wissenschaftliche Forschung gegenüber, die genau das Gegenteil anstrebt, beispielweise die bereits stattfindende datenbasierte Diskriminierung in der Arbeitswelt aufzuzeigen.²⁴ Wir beobachten also ein grosses Spannungsfeld zwischen einer ohne ethischen Bedenken vorwärts strebenden, oft durch finanzielle Motive getriebenen, Big Data / IoT Forschung und Entwicklung auf der einen Seite und einer Nutzung der dabei entwickelten Technologien durch andere Forschende, um den Missbrauch aufzuzeigen. Hier wäre es dringend notwendig, dass wenigstens die Forschenden an den Hochschulen die ethischen Aspekte sehen und diskutieren.

[Rz 93] Das eigentliche Risiko für die Wissenschaft ist, dass sie angesichts ihrer Kommerzialis-

²⁴ Am 27. Bis 29. August 2015 findet die Konferenz «Discrimination in the Labor Market» in Bern statt. http://www.wirtschaft.bfh.ch/de/forschung/institut_unternehmensentwicklung/veranstaltungen/conference_lm_discrimination.html.

sierung und dem wachsenden Drittmitteldruck zu einem Dienstleistungssektor wird, dessen Arbeitnehmer die soziale Leiter immer weiter hinuntersteigen, weil in den Augen der Öffentlichkeit nicht das Erfinden genialer Algorithmen sondern das Geld Machen mit den Algorithmen wertvoll ist. Je weniger ethisch reflektiert Wissenschaftler arbeiten, desto grösser scheint uns das Risiko, dass dieser Abstieg stattfinden wird.

5 Zusammenfassung

[Rz 94] Big Data bringt grossen Nutzen, geht aber mit vielen nicht ganz unkritischen Paradigmenwechseln einher, die das Auftreten von verführerischen Schimären und Illusionen fördern. Zwei Hauptprobleme stellen das De-Anonymisierungspotential des Big Data Profiling und das Diskriminierungsrisiko dar. Ersteres lässt sich mit rein rechtlichen Massnahmen kaum befriedigend lösen, sondern benötigt einen transdisziplinären Ansatz, zweiteres verlangt nach einem breiten politischen Diskurs. Neben der Verhinderung des Missbrauchs sollte der Staat auch über positive Fördermassnahmen für das entstehende Big Data Ökosystem nachdenken. Für die Rechtsinformatik ergeben sich aus diesen Überlegungen neue Forschungsfelder, insbesondere in den Bereichen Transparenz²⁵, Good Practices bei der verantwortungsvollen Nutzung von Big Data in der Praxis und Nutzung des Internets der Dinge. Da wir erst am Anfang der breiten Big Data Nutzung stehen, müssen zwar viele Aussagen in diesem Beitrag als vorläufige Beurteilung angesehen werden, doch der politische Diskurs über Big Data ist schon jetzt möglich und sinnvoll — und in manchen Bereichen sogar dringend notwendig!

Prof. Dipl. Ing. Dr. phil. REINHARD RIEDL, Wissenschaftlicher Leiter des Fachbereichs Wirtschaft, Berner Fachhochschule (BFH), reinhard.riedl@bfh.ch; <http://www.wirtschaft.bfh.ch>.

²⁵ Transparenz war schon 2013 Generalthema der Rechtsinformatiktagung IRIS, <http://www.univie.ac.at/ri/IRIS2013/>.