Die Seite des SF / La page du FS

Big Data

Veranstaltung des Zentrums für Informations- und Kommunikationsrecht der Universität Zürich, der Forschungsstelle für Informationsrecht der Universität St. Gallen und des Schweizer Forum für Kommunikationsrecht SF•FS vom 31. Oktober 2013

AURELIA TAMÒ[*]

I. Einleitende Worte der Tagungsleiter

Die Tagung «Big Data» wurde von den Tagungsleitern Prof. Dr. ROLF H. WEBER, Ordinarius an der Universität Zürich und Leiter des Zentrums für Informations- und Kommunikationsrecht der Universität Zürich, und Prof. Dr. FLORENT THOUVENIN, Assistenzprofessor für Immaterialgüter- und Informationsrecht und Direktor der Forschungsstelle für Informationsrecht an der Universität St. Gallen, im Zürcher Zunfthaus zur Schmiden eröffnet.

WEBER und THOUVENIN verwiesen einleitend auf das Ziel der Tagung, vermehrt eine holistische Herangehensweise an das Thema «Big Data» zu fördern. Die Tagung wurde entsprechend in zwei Blöcke unterteilt: Während in der ersten Hälfte mehrheitlich die technischen Aspekte sowie das wirtschaftliche und soziale Potential von Big Data erläutert wurden, folgte im zweiten Teil die Auseinandersetzung mit den rechtlichen Spannungsfeldern.

II. Technische und rechtliche Grundlagen zur Thematik

1. Begriff und Bedeutung von Big Data

Dr. ANDREAS WESPI, Forscher am CTO Office IBM SWG Europe, stellte einleitend die drei charakteristischen Eigenschaften von Big Data vor: Volume, Velocity und Variety (3 V). So werden heute riesige Datenmengen (i.d.R. im Bereich von Zettabytes), die sich aus strukturierten (z.B. Zahlenreihen) und unstrukturierten Daten (z.B. Bildern oder Sensordaten) zusammensetzen, in hoher Geschwindigkeit fortlaufend ausgewertet. Zur Auswertung werde auf die Software Hadoop[1] zurückgegriffen, die in der Industrie als Standard anerkannt sei. Die Hauptstärke von Hadoop sei die Aufteilung der Daten in Teilprobleme, deren Bearbeitung zunächst einzeln erfolge, bevor sie schliesslich wieder zu einem Ganzen zusammengefügt würden. Zusammenfassend hielt WESPI fest, dass der Sprung von monolithischen technischen Systemen zu dynamischen und flexiblen Lösungen Big-Data-Analysen erst ermöglicht hätten.

---

[*] Doktorandin, Forschungsstelle für Informationsrecht (FIR-HSG), Universität St. Gallen.
[1] Apache Hadoop ist ein in Java programmiertes open source software framework, mit welchem grosse Datenmengen parallel verarbeitet und gespeichert werden können.

Als Hauptmotivation für das Sammeln von Daten und deren Auswertung nannte WESPI das Streben nach intelligenten Systemen. Doch berge eine «smarte Welt» unterschiedliche Spannungsfelder. Während beispielsweise Energieeinsparungen ein gesellschaftlich wünschenswertes Ziel seien, das durch den Einsatz von smart metering ermöglicht werde, entstünden gleichzeitig verfeinerte, personalisierte Profile von Haushalten und Einzelpersonen. Für Informatiker in den betroffenen Branchen bedeute dies, dass sie sich mit der (rechtlichen) Frage auseinandersetzen müssten, wie korrekt und vertrauenswürdig die Daten und somit die resultierenden Auswertungen seien.

Abschliessend hielt WESPI fest, dass die Schlussfolgerung «mehr Daten ist immer besser» unzutreffend sei. Die zentrale Frage laute vielmehr, welche Daten gebraucht würden, um sinnvolle Schlussfolgerungen ziehen zu können. Auch dürfe nicht vergessen werden, dass mit dem Aufbau von zentralen Systemen ein Sicherheitsrisiko einhergehe. Dementsprechend dürften rechtliche Grundlagen Unternehmen nicht daran hindern, einen angemessenen Schutz sicherzustellen. So sei es zum Aufbau eines sichereren Systems zuweilen durchaus sinnvoll, mehr Daten als unbedingt nötig zu sammeln, um auf diese Weise präzisere Monitoring-Systeme aufzubauen, die Cyberangriffe frühzeitig erkennen könnten. Das Sammeln von «mehr Daten als unbedingt nötig» widerspreche jedoch dem Datenminimierungsprinzip.

2. Paradoxe in der Datenschutzdiskussion und das Streben nach neuen Konzepten

Aufbauend auf und im Einklang mit WESPIs Erläuterungen hielt WEBER fest, dass neue Herausforderungen für das Recht insbesondere durch die vermehrte Bedeutung der Zweitverwendung von Daten entstünden.

WEBER hob hervor, dass die Datenschutzdiskussion von Paradoxen geprägt sei. So sei einerseits der Wunsch nach Transparenz vorhanden, indem viele Unternehmen und der Staat nach möglichst vollständigen Informationen über ihre Konsumenten respektive Bürger strebten, während gleichzeitig dem Wunsch nach einer Offenlegung der Auswertungsverfahren nur selten entsprochen werde. Nebst dem «Transparenzparadox» erkennt WEBER auch ein «Machtparadox», welches auf der Erstellung eines umfassenden Nutzerprofils durch Big Data basiere. Dabei seien die Gewinner von Big Data jene, welche den grössten Einfluss auf die Datenanalysen hätten. Dies sei aber nicht so zu verstehen, präzisierte WEBER, dass nur Unternehmen als Gewinner anzusehen seien; vielmehr seien die Datenbesitzer Gewinner im Big-Data-Zeitalter. So nütze es beispielsweise einem Konsumenten, dass er schnell in Erfahrung bringen könne, welche medizinische Behandlung für ihn persönlich erfolgversprechend und am verträglichsten sei.

Im Weiteren stellte WEBER die Frage, ob die heutigen Datenschutzansätze noch adäquat seien, ob konkret die Trennung zwischen Personen- und Sachdaten anzupassen sei und ob neue Einwilligungsregelungen sinnvoll wären. Ferner seien Diskussionen über die Verbesserung der Verfahrensgarantien für Nutzer notwendig, damit die im Datenschutzgesetz (DSG) verankerten Rechte auch durchgesetzt werden könnten. WEBER wies auf die Möglichkeit hin, Unternehmen durch Codes of Conduct zur Selbstverantwortung zu animieren und mahnte Regulatoren in Europa und der Schweiz, sektorspezifische Regulierungen nicht ausser Acht zu lassen.

WEBER beendete seine Ausführungen mit der provokativen Forderung eine Nutzergemeinschaft zwischen Unternehmen und Nutzern anzustreben, welche die aus den erhobenen Daten resultierenden Vorteile – auch monetärer Natur – unter den beiden Anspruchsgruppen aufteilen solle. Eine solche Nutzergemeinschaft wirft die Frage nach dem Eigentum an Daten auf – eine Diskussion, auf die WEBER nicht näher einging. Der Aufbau einer Nutzergemeinschaft würde die Schaffung einer Überwachungsorganisation bedingen. Diese müsste nicht zwangsläufig staatlicher Natur sein, jedoch trotzdem garantieren können, dass sich die Unternehmen an die Vertragsbedingungen halten.

III. Ausgewählte Unternehmungs- und Forschungsbereiche

1. Suchmaschinen und Social Media

JEAN-PIERRE KÖNIG, Head of Big Data Analytics bei YMC AG, eröffnete sein Referat mit der Anmerkung, dass das Phänomen Big Data an sich nichts Neues sei, sondern Unternehmen schon lange den Wert von Datenanalysen erkannt hätten. Gründe, weshalb Unternehmungen sich mit dem Thema Big Data auseinandersetzen würden, seien in der Umsatzsteigerung, in der verbesserten Entschei-

dungsfindung und strategischen Positionierung des Unternehmens sowie in dem verbesserten Risikomanagement zu suchen.

Als Treiber von Big Data nannte KÖNIG die vermehrte bewusste und unbewusste Generierung von Daten. Bewusste Daten (human-generated data) entstünden u.a. durch das tägliche Benützen von Social Media, Suchmaschinen oder durch das Erstellen von Dokumenten am Arbeitsplatz. Dabei präzisierte KÖNIG, dass das Adjektiv «bewusst» nicht immer zutreffend sein müsse, da beispielsweise bei der Benutzung von Suchmaschinen vielen Nutzern nicht bewusst sei, dass nebst dem Suchkriterium auch weitere Daten generiert würden. Unbewusste Daten dagegen (sensor-generated data) entstünden durch den Einsatz von verschiedenster Sensoren. KÖNIG prognostizierte, dass der grösste Teil unserer physischen Welt in naher Zukunft mit dem Internet verbunden sein wird (Stichwort «Internet of Things») und dass das gesamte Aufkommen an durch Maschinen generierten Daten bald das Aufkommen von menschlich generierten Daten übersteigen wird.

Sowohl bewusst wie unbewusst übertragene Daten würden von Unternehmen zur Erstellung von Nutzerprofilen verwendet, die sich aus den Interessen, Bewegungsprofilen etc. eines Kunden zusammensetzten. Aus diesen Profilen würden wiederum Produktempfehlungen generiert, die auf Ähnlichkeiten mit anderen Nutzerprofilen basierten. Für KÖNIG steht dabei fest, dass auch in Zukunft die Erhebung, Auswertung und Nutzung solcher Daten durch Unternehmen weiter zunehmen wird. Dies ergebe sich bereits aus der Tatsache, dass die Open-source-Software Hadoop immer billiger werde und somit vermehrt von kleinen Unternehmen eingesetzt werden könne. KÖNIG prognostizierte abschliessend, dass die Motivation der Unternehmen, 360°-Nutzerprofile herzustellen und Kunden sowohl offline als auch online zu erreichen, zu einer «Massenpersonalisierung» führen werde.

2. Neuer Trend: Personalisierte Medizin

Prof. Dr. MATTHIAS DEHMER vom Institute for Bioinformatics and Translational Research an der Privaten Universität für Gesundheitswissenschaften, Medizinische Informatik und Technik Tirol verschaffte den Teilnehmern einen Einblick in die Welt der personalisierten Medizin. Dabei ging er insbesondere auf personalisierte Gesundheitsbilder und die sich daraus ergebenden Behandlungsansätze ein. DEHMER führte aus, dass diese Patientenprofile auf der Analyse komplexer Netzwerke basierten, welche ihrerseits aus Datenbanken extrapoliert würden. Die Netzwerke beschrieb er als eine relationale Struktur, welche aus Knoten- und Kantenmengen bestehe. Zur Veranschaulichung erwähnte er das Internet, in welchem Websites sog. Knotenmengen darstellten und die einzelnen Links im World Wide Web als Kanten beschrieben werden könnten.

Sein Forschungsteam sei auf komplexe Krankheiten spezialisiert, wobei in diesem Bereich insbesondere die Interaktion der Gene in den Netzwerken interessant sei. Um solche Netze aussagekräftig zu analysieren, würden statistische Methoden angewendet. Obwohl immer noch Zukunftsmusik, legte DEHMER dem Publikum die Vorteile für den Patienten nahe, wenn dieser beim Arzt jeweils sein persönliches Netzwerk von Genen vorlegen und somit eine persönliche Behandlung erlangen könne. Dabei unterstrich DEHMER den Wert der Datenquantität. Grosse Mengen von Daten seien notwendig, um sinnvolle Schlüsse aus den Auswertungen zu ziehen.

Im Weiteren stellte DEHMER fest, dass es für Forscher im Bereich der personalisierten Medizin nicht immer einfach sei, an frei verfügbare Daten zu gelangen. Überdies spiele die Datenqualität eine wesentliche Rolle. Gleichzeitig aber würden grössere, qualitativ hochstehende Datensets zu höheren Kosten bei der Auswertung und einem langwierigen Prozess bei der jeweiligen Ethikkommission der Universität führen, welche Projekte, die Personendaten (auch anonymisierte Daten) und insbesondere sensitive Daten beinhalten, bewilligen müsse.

Abschliessend merkte DEHMER an, dass das enorme Potenzial im Bereich der personalisierten Medizin durch Big-Data-Analysen immer weiter vorangetrieben werde und eine solche Entwicklung wünschenswert sei. Gleichzeitig führten die explosionsartige Anhäufung, Generierung und automatische Auswertung von Daten aber dazu, dass Forscher im Bereich der Bioinformatik die angewendeten Methoden der Verarbeitung gut kennen müssten, damit Fehler in der Nutzung der Daten minimiert werden könnten.

IV. Deep Dives – Analyse von rechtlichen Spannungsfeldern

1. Anonymisierung vs. Re-Individualisierung

Der Datenschutzbeauftragte des Kantons Zürich, Dr. BRUNO BAERISWYL, ging auf die Frage ein, inwiefern bei Big-Data-Analysen Personendaten verarbeitet werden. Insbesondere konzentrierte er sich dabei auf die Rolle der Re-Identifikation im Rahmen des DSG. So sei zwar der Prozess des Anonymisierens datenschutzrechtlich relevant, das Resultat aber – anonyme Daten – nicht mehr. Demnach müssten beim Anonymisierungsprozess die Datenbearbeitungsgrundsätze des DSG eingehalten werden. Dagegen sei der Prozess des Re-Identifizierens dogmatisch nicht vorgesehen und somit datenschutzrechtlich nicht relevant. Das Ergebnis jedoch, die re-individualisierten Daten, stellten den notwendigen Personenbezug wieder her, womit sich die Anwendbarkeit des DSG erneut ergebe.

BAERISWYL unterstrich die Bedeutung von lernenden Algorithmen, welche die erfolgreiche Suche nach dem zum Zeitpunkt der Datensammlung unbekannten Mehrwert ermöglichten. Dabei sei für den Einzelnen oftmals die informationelle Integrität bei Big-Data-Analysen nicht gewährleistet. Bei der informationellen Integrität gehe es darum, wie eine Person auf-grund ihrer Daten präsentiert werde und was für Prognosen über sie gestellt würden. Big-Data-Analysen würden auch ein nicht vernachlässigbares Diskriminierungspotenzial bergen. Ferner sei in der Literatur bereits mehrfach aufgezeigt worden, dass durch das Anhäufen von anonymen Daten die Re-Identifizierung möglich ist und Personenprofile wiederhergestellt werden könnten.

Basierend auf der obigen Argumentation stellte BAERISWYL die Nutzung von Daten – auch von anonymen – in den Mittelpunkt der juristischen Debatte. Zentral seien dabei die Fragen, ob bereits die Übermittlung oder Veröffentlichung von anonymen Daten einen Rechtfertigungsgrund benötige und ob datenschutzrechtliche Grundsätze auch bei der Bearbeitung von anonymen Daten anwendbar sein sollten. Weiter stellte BAERISWYL die Frage, ob die Re-Identifikation einen Rechtfertigungsgrund benötige. So gebe es etwa im Verwaltungsrecht ein Konkordat zwischen den Kantonen, das die polizeiliche Datenauswertung regle und Rechtfertigungsgründe nenne. Im privaten Sektor jedoch stehe die Vertragsfreiheit im Zentrum, womit laut BAERISWYL weitere Probleme im Rahmen der informierten Einwilligung (Weiss ein Nutzer, welche Daten erhoben werden?), der Zweckbindung (Weiss ein Nutzer, für welchen Zweck die Daten erhoben werden?) und der individuellen Durchsetzung (Kann ein Nutzer die Erhebung und Nutzung der Daten nur für den eingewilligten Zweck durchsetzen?) entstünden.

2. Erkennbarkeit und Zweckbindung

THOUVENIN knüpfte an die Überleitung von BAERISWYL an und wies zunächst auf die beiden zentralen datenschutzrechtlichen Grundsätze hin: das Erkennbarkeits- und das Zweckbindungsprinzip. Diese Prinzipien könnten die zukünftigen Nutzungsmöglichkeiten von Datensets erheblich einschränken. Im Weiteren erläuterte THOUVENIN, dass sein Referat sich auf die Verwertung von Daten durch private Parteien konzentriere.

Das Prinzip der Erkennbarkeit gemäss der EU-Richtlinie 95/46/EG verlange, dass die jeweils betroffene Person (Datensubjekt) von der Datenverarbeitung Kenntnis habe. Dabei solle nicht nur das Beschaffen, sondern auch das Bearbeiten sowie die Weitergabe der Daten an Dritte für das Datensubjekt aus den Umständen – demzufolge nicht explizit – erkennbar sein. Dies solle schliesslich einen informierten Entscheid des Betroffenen ermöglichen. Mit steigender Komplexität der Datenverarbeitung wüchsen deshalb auch die Anforderungen an die Erkennbarkeit. Die grössten Herausforderungen in der Praxis seien aber mehrmalige Weitergaben der Daten an Dritte sowie die Frage, inwiefern die Information über die Datenverarbeitung dem Datensubjekt erkennbar gemacht werden müsse, und in welchem Rahmen Daten zu anderen Zwecken weiter verwendet werden könnten.

Basierend auf diesen praktischen Herausforderungen erörterte THOUVENIN das Prinzip der Zweckbindung anhand seiner historischen Entwicklung sowie dessen Zielsetzung und Bedeutung im heutigen Big-Data-Zeitalter. In der Schweiz hätten bei der Ausarbeitung des DSG zwei Vorentwürfe vorgelegen: einer für den öffentlichrechtlichen und einer für den privatrechtlichen Bereich. Unter der Leitung von Prof. MARIO PEDRAZZINI sollten die Entwürfe beider Kommissionen zu einem Gesetz verschmolzen werden. Dabei sei der Grundsatz der Zweckbindung im heutigen Sinn anfänglich nur im Vorentwurf für die Bundesverwaltung und nicht auch in jenem für den Privatsektor vorgesehen gewesen. Der Sinn und Zweck der Bestimmung sei entsprechend in der öffentlich-rechtlichen Sphäre zu suchen, argumentierte THOUVENIN. Ob und inwiefern die Zweckbindung im privaten Bereich sinnvoll sei, müsse deshalb hinterfragt werden. Bezeichnend sei, dass sich weder der Gesetzgeber noch die Lehre und die Rechtsprechung bisher vertieft mit der Legitimation der Zweckbindung auseinanderge-

setzt hätten. Führe man sich die Entstehungsgeschichte des Datenschutzgesetzes vor Augen, sei davon auszugehen, dass diese kaum zum Schutz der Privatsphäre der Betroffenen vorgesehen worden sei, sondern vielmehr als logische Folge der Beschränkung des staatlichen Handelns in einem grundrechtlich geschützten Bereich, mithin eine Konkretisierung der Voraussetzungen der gesetzlichen Grundlage, des öffentlichen Interesses und des Verhältnismässigkeitsprinzips zu verstehen sei. Davon ausgehend, dass das Prinzip auch künftig weiter bestehen wird, stellte THOUVENIN die Frage, ob das Zweckbindungsprinzip bei der Datenverarbeitung von Privatpersonen weniger strikt ausgelegt werden sollte oder ob durch Kombinationsmöglichkeiten von einem weitem Zweck und weiter Bindung oder durch das Zulassen von inhaltlich weit gefassten Einwilligungen übermässigen Einschränkungen entgegen gewirkt werden könnte.

3. Datensicherheit

Rechtsanwältin NICOLE BERANEK ZANON wies einleitend auf die typischen Charakteristiken von Big-Data-Analysen hin, die den Grundvorstellungen des Datenschutzes diametral entgegenstünden. So habe zum Beispiel Google Translate massgeblich verbessert werden können, indem sich das Unternehmen bei der Produktentwicklung auf enorme Datenmengen gestützt habe, mitunter auch auf «schlechte» bzw. nicht integre Übersetzungsdaten. Offensichtlich bestehe somit ein Anreiz für Unterneh-men, möglichst viele Daten zu sammeln und zu verarbeiten.

Das Referat von BERANEK ZANON widmete sich dem Thema der Datensicherheit. Sie erklärte, dass die Sicherheit bei der automatisierten Bearbeitung von Personendaten durch mehrere Kontrollschritte gewährleistet werde. Zum einen sei die Zu-gangskontrolle wesentlich. Im Zeitalter von Big-Data-Analysen sei bereits diese gefährdet, da sicheres Computing technisch immer anspruchsvoller werde und Fehler beispielsweise bei der Zusammensetzung von Datenpaketen in Hadoop entstehen könnten. Weiter nannte BERANEK ZANON Personendatenträgerkontrollen, Eingabe- und Zugriffskontrollen auf Datenbanken, Transportkontrollen, wie zum Beispiel der Einsatz von Verschlüsselungstechniken, Bekanntgabekontrollen (z.B. durch end-to-end validation) sowie Speicherkontrollen als Prüfziel.

Ferner zeigte BERANEK ZANON anhand aktueller Fälle auf, dass es gute Gründe gibt, die heutigen datenschutzrechtlichen Konzepte kritisch zu hinterfragen. So habe die Weiterleitung anonymisierter Daten durch TomTom (einem niederländischen GPS-Dienstleistungsanbieter) an die Verkehrspolizei, dieser das Anbringen von Radarattrappen an strategischen Standpunkten erleichtert. Die Kunden von TomTom hätten diese Praxis aber kritisiert und deren Rechtmässigkeit hinterfragt. Heute erkläre das niederländische Unternehmen seinen Kunden anhand eines instruktiven Videoclips, wie ihre GPS-Daten verarbeitet und weitergegeben werden. Unternehmen sollten, so BERANEK ZANON, ihre Kunden vermehrt gezielt darüber informieren, wie Daten verarbeitet werden und was anschliessend damit gemacht werde. So könne der gesellschaftliche Mehrwert von Big Data genutzt und gleichzeitig datenschutzrechtlichen Anliegen entsprochen werden.

V. Podiums- und Plenumsdiskussionen

Die Tagung wurde durch zwei Podiums- und Plenumsdiskussionen abgerundet. Die erste Debatte bezog sich auf die Ausführungen zum Grundlagenteil und die ausgewählten Unternehmens- und Forschungsbereiche. Die zweite abschliessende Diskussionsrunde widmete sich den ausgewählten rechtlichen Fragen. Neben den jeweiligen Referenten wurde die zweite Diskussionsrunde durch drei Gastredner bereichert, darunter der Europäische Datenschutzbeauftragte Dr. PETER HUSTINX, der Rechtsanwalt Dr. ROLF AUF DER MAUR und der Rechtskonsulent DAVID ROSENTHAL.

Die erste Diskussionsrunde befasste sich schwergewichtig mit der Betroffenheit der Nutzer bei Big-Data-Analysen. Die Referenten waren sich einig, dass die Nutzer nur zu einem geringen Grad von solchen Analysen betroffen sind. Einerseits würden diese nämlich stets über die Datenverarbeitung informiert (obwohl meistens über die Unterzeichnung von AGB), und andererseits würden Kunden von den Auswertungen direkt profitieren (z.B. durch effizientere Dienstleistungen). DEHMER präzisierte, dass die Frage nach der Betroffenheit im Bereich der medizinischen Forschung immer von einer Ethikkommission abgewogen werde. Bei der Frage nach Massnahmen zur Verbesserung des Schutzes für die Nutzer kamen die Referenten auf keinen gemeinsamen Nenner. So wurden die transparentere Kommunikation und Information der Betroffenen, die Suche nach datenschutzfreundlichen technischen Massnahmen, die Etablierung von internen organisatorischen Prozessen zur Einhaltung der

Datenschutzbestimmungen oder die Einführung von Zertifikaten als mögliche Strategien genannt und kontrovers zwischen WESPI, KÖNIG, THOUVENIN und WEBER diskutiert.

Die zweite Podiumsdiskussion wurde von HUSTINX eröffnet, welcher zum verantwortungsbewussten Umgang mit Daten aufforderte. Während das Potenzial von Big Data von allen Rednern hervorgehoben wurde, gingen die Meinungen hinsichtlich des Umgangs mit den Risiken auseinander. Diskutiert wurden mitunter die in der Europäischen Union angestrebte Vereinheitlichung der Datenschutzbestimmungen durch eine übergreifende Verordnung, Gründe, welche für sektorspezifische Regulierungen sprechen sowie die Möglichkeit der Selbstregulierung der Industrie. AUF DER MAUR, ROSENTHAL und BERANEK ZANON merkten an, dass Datenschutz nur ein Mechanismus zur Lenkung von Unternehmen sei. Auch die Vertrauensbildung und der öffentliche Druck hätten Einfluss darauf, wie Daten in Unternehmen gesammelt, verarbeitet und gelagert würden. Gleichzeitig mahnten aber BAERISWYL und HUSTINX zur fairen Balance widersprechender Interessen zwischen Unternehmen und Nutzern. Insbesondere die Frage nach der Legitimation einer Datenverarbeitung und nach den berechtigten Interessen für die Datenverarbeitung sollen in den Vordergrund gestellt und die Risiken von Persönlichkeitsverletzungen überprüft werden.