

erschieden in der FfF-Kommunikation,
herausgegeben von FfF e.V. - ISSN 0938-3476
www.fiff.de

Martin Degeling

Datenspuren

Bericht von der Konferenz am 13./14. Oktober 2012 in Dresden

Voll verwandt war der Untertitel der diesjährigen, vom CCC organisierten, *Datenspuren-Konferenz in Dresden*. Aus dem umfangreichen Programm stellt dieser Beitrag zwei Themenstränge vor: Zunächst geht es um aktuelle datenschutzrechtliche Diskussionen am Beispiel des Meldegesetzes und der Europäischen Datenschutzreform, anschließend um unterschiedliche Perspektiven auf Formen und Folgen von Data Mining.

Die „Datenspuren“ finden seit 2004 jährlich in Dresden statt. Ausgerichtet vom c3d2 entwickelte sich die Veranstaltung¹ relativ schnell von einem kleineren Symposium zu einem zweitägigen Kongress mit dem inhaltlichen Schwerpunkt auf Datenschutz, aber auch mit Raum für andere Themen von IT-Security bis Hardware-Hacking. *Voll verwandt* war vor allem das Foyer, wo die gesamte Zeit über *Pentabugs* zusammengebaut wurden. Mikrocontrollerboards, die mit Lichtsensoren, Lautsprechern, LEDs und Vibrationsmotoren ausgestattet über die Tische kro-

chen – wenn das Lötten erfolgreich war. Das *Pentabugs*-Bauen war Teil der Angebote, die vor allem (jugendlichen) Laien die Basics von Software und Hardware näher bringen sollten; Einführungsvorträge zum sicheren Surfen im Internet, TOR für Anfänger_innen, oder ein Bericht aus dem lokalen Projekt ‚Chaos macht Schule‘ fielen auch in diese Kategorie. In den zwei parallelen Tracks fanden über 20 Vorträge Platz. Vorstellen will ich im folgenden zwei Themenstränge, die sich in etwa fünf Vorträgen wiederfanden.

Juristische Datenschutzaspekte

Vielen ist das neue Meldegesetz aus der Presse bekannt als das schnellste Gesetz aller Zeiten, das – während des WM-Spiels Deutschland–Italien – von den anwesenden Abgeordneten ohne Diskussion durchgewunken wurde. Erst nach Protesten gegen diverse Klauseln, die die Meldebehörden zu Adresshändler_innen umfunktionieren sollten, wird das Gesetz aktuell im Vermittlungsausschuss beraten. *Jürgen von der Ohe*, Mitarbeiter beim Datenschutzzentrum in Schleswig-Holstein, stellte in seinem Vortrag einige weniger bekannte Punkte heraus, die einer Überarbeitung bedürfen. Insbesondere die phonetische Ähnlichkeitssuche steht in der Kritik. Dabei enthalten Adresslisten nicht nur exakte Treffer, sondern auch, beispielsweise bei der Anfrage nach der Adresse einer „Rieke Mayer“, die Adressen aller Rieke Meiers und „phonetisch ähnlicher“ Namen. Die Notwendigkeit einer flexibleren Suche steht für von der Ohe außer Frage. Das jedoch nicht genau nachgehalten wird, dass die abfragende Stelle all die Namen und Adressen löscht, die nicht Ziel der Anfrage waren, ist aber ein Problem. Selbst das in das Gesetz aufgenommene Verbot des Adresspoolings löst dieses Problem nicht. Aktuell gängige Praxis ist es, dass Mittler_innen wie *regis24* für größere Unternehmen die Meldebehördenanfragen übernehmen, die Adressen erfragen und an den/die Auftraggeber_in übermitteln, aber gleichzeitig eine Kopie in der eigenen Datenbank behalten. Stück für Stück entsteht auf diese Weise bei den Mittler_innen eine zentrale Spiegeldatenbank der dezentralen Meldebehörden. Diese Form des Pooling ist zwar zukünftig untersagt, aber neben der nicht sehr konsequent durchgesetzten Revisionierbarkeit (also der Nachverfolgung wie die Mittler arbeiten, und ob Sie die Daten auch tatsächlich löschen) fehlen im Gesetz auch Bußgelder und Strafen für eben solche Anbieter.

Aus der juristisch/politischen Perspektive berichtete *Zora Siebert*, Mitarbeiterin des Grünen-MdEP Jan Philipp Albrecht. Thema ihres Vortrags war der aktuelle Stand der Verhandlungen um die EU-Datenschutzgrundverordnung, für welche ihr Arbeitgeber als Berichterstatter des LIBE-Ausschusses (Ausschuss für bürgerliche Freiheiten, Justiz und Inneres) im Parlament tätig ist. Die inhaltliche Kritik deckt sich dabei in großen Teilen mit denen der FfF-Arbeitsgruppe, die auf der Jahrestagung vorgestellt wurde. Interessante Inneneinsichten gab Siebert zum Stand des Gesetzgebungsverfahrens. Nach der Überarbeitung des Entwurfs im Ausschuss findet in den kommenden Monaten ein

Dialog zwischen Ausschuss, Kommission und Rat statt. Für den Europäischen Rat wird hier Irland an den Verhandlungen teilnehmen, das zu dieser Zeit die Ratspräsidentschaft inne hat und z. B. im Konflikt mit Facebook für seine *entspannte* Position zum Datenschutz bekannt geworden ist. Die im Dialog entstandene Fassung wird erst danach im Europäischen Parlament diskutiert und verabschiedet. Mit einer abschließenden Abstimmung rechnet Siebert gegen Ende dieser Legislaturperiode (also bis Mitte 2014). Ihrer Einschätzung nach, und das war auch der Grund ihres Besuches, ist es notwendig die Lobbyarbeit der Konzerne mit kontinuierlicher Arbeit aus der Bürgerrechtsbewegung zu kontern. Gerade bei vielen Abgeordneten (natürlich vor allen denen der anderen Parteien) sollte man sich auch als Einzelne_r mit E-Mails und Briefen für einen stärkeren Grundrechtsschutz einsetzen. Genauso müsse man das Bundesministerium des Inneren (BMI), das über den Rat an der Überarbeitung beteiligt ist, noch von datenschutzfreundlicheren statt wirtschaftsfreundlichen Regeln überzeugen. Insgesamt, und so sieht auch das Fazit der FfF-AG aus, sieht sie aber große Vorteile in der zentralen Regelung von Datenschutzfragen, auch wenn gerade die Art und Weise der Verarbeitung personenbezogener Daten durch EU-Institutionen und durch Polizei- und Sicherheitsbehörden nicht in der Verordnung geregelt ist.

Data-Mining und Post-Privacy

Eine weitere inhaltliche Linie ergab sich zwischen drei Vorträgen, die im weitesten Sinne mit dem Thema *Data Mining* zu tun hatten. Dass ich diese Linie ausmachen konnte, hat vermutlich auch damit zu tun, dass ich einen der drei Vorträge gehalten habe, die ich im Folgenden umreißen will. *Joachim Scharloth* sprach über die Möglichkeit mittels computerlinguistischer Methoden der Textanalyse die Autorschaft eines anonymen Textes zu ermitteln. Dabei bezog er sich auf den Fall von Andrej Holm², einem Berliner Sozialwissenschaftler, der vor einigen Jahren als mutmaßlicher Kopf der *militanten Gruppe (mg)* verhaftet wurde, die sich zu mehreren Brandanschlägen, insbesondere auf militärische Fuhrparks, bekannt hatte. Der Fall war auch deswegen in den Medien präsent, weil das ermittelnde BKA seinen Verdacht gegen Holm unter anderem damit begründet hat, dass der Stadtsoziologe in seinen Veröffentlichungen einen sehr ähnlichen Schreibstil pflegte wie die *mg*. Die automatisierte Analyse des *sprachlichen Fingerabdrucks* hatten die Ermittler_innen auf Holms Spur gebracht, der schon vor 2007 die mittlerweile



Martin Degeling

Martin Degeling hat auf den Datenspuren erstmals einen Vortrag unter seinem Online-Pseudonym gehalten und vollzieht mit diesem Bericht das Ende der Identitätstrennung. Er ist wissenschaftlicher Mitarbeiter und Doktorand am Institut für Informations- und Technikmanagement an der Uni Bochum und in der FfF-AG mit dem Mailinglistenkürzel [EU-DSGVO] aktiv.

gebräuchlicheren Worte wie „Gentrifizierung“ und „Prekariisierung“ benutzte. Holm konnte aber auch nach intensiven Ermittlungen mit absurden Auswüchsen staatlicher Überwachung keine Urheberchaft nachgewiesen werden. Joachim Scharloth, selbst Professor für Angewandte Linguistik und Blogger³, hatte nun eine Reihe computerlinguistischer Methoden nicht nur auf Texte der mg und von Holm angewendet, sondern auch auf solche der Revolutionären Zellen, der RAF, dem Bildkolumnisten Franz Josef Wagner und der NPD. Entscheidungsbaume und Clustering-Verfahren wiesen zwar häufig eine Nähe zwischen der NPD und Wagner nach, gleichzeitig hätte aber auch die seit mehreren Jahrzehnten aufgelösten RZ Urheber der mg-Texte sein können, ebenso wie Holm auch in den 70ern schon RAF Mitglied hätte sein können. Der unterhaltsame Vortrag zeigte, wie schnell sich auf Basis angeblich neutraler wissenschaftlicher Verfahren genehme Ergebnisse produzieren lassen, denen man gerne Glauben schenkt. Einer genaueren Analyse hielt aber keines der Ergebnisse stand, etwa dass der Entscheidungsbaum, der Holm als mg-Mitglied klassifizierte, dies auf Basis eines Bruchteils der zur Verfügung stehenden Analysekatoren tat.

Scharloths Vortrag ging zufällig meiner anschließenden Beschreibung des *Panoptic Sort*⁴ voraus. Ein Begriff, der, in den 1990er geprägt, die Wirkmächtigkeit von Data Mining auf den Einzelnen beschreibt. Wenn beispielsweise Adresshändler_innen, die die Kaufkraft und Interessen von Haushalten anhand der Anzahl von Dokortiteln in einem Straßenzug berechnen, findet anschließend eine Kategorisierung statt, um zielgerichtete Werbung verschicken zu können. Ähnlich wie beim Scoring funktionieren diese Mechanismen als Verhaltensvorhersage und schränken damit durch die Hintertür ein, welche (Kauf-)Optionen den Einzelnen auf Grund der Vorhersagen überhaupt angeboten werden. Dabei ist es in vielen Fällen egal, ob man selbst in eine Datenweitergabe oder Nutzung eingewilligt hat. Notfalls werden nicht vorhandene Daten durch statistische Approximationen ersetzt während gleichzeitig die Anwendungsfelder ausgeweitet werden.

Viel weniger skeptisch war am Sonntag Vormittag Florian André Unterburger. Der engagierte Piratenpartei-Pirat argumentierte mit dem Optimismus der *Post-Privacy-Bewegung*, dass die Verarbeitung von immer mehr Daten im Interesse aller zur Verbesserung von quasi allem (von Krankheitsbehandlung bis Wirtschaftssystem) beitragen könne. Kern seines Vortrags waren dann zwei Methodenansätze, die in einer (zukünftigen) mehrheitlich *post-privacy* geprägten Welt einzelnen Privatheitsliebenden erlauben sollen, weiterhin über die sie beschreibenden Informationen zu bestimmen. Unter dem Begriff Multonymität fasst er Praktiken zusammen, die Anonymität durch Multiple Identitäten gewährleisten sollen. Durch die Pflege von diversen, voneinander abgetrennten, (Online-)Identitäten solle man verhindern, dass alle Informationen zu einer Person aggregiert werden. Gleichzeitig könnten die Daten der einzelnen Identitäten weiterhin für das Wohl aller genutzt werden. Die Idee erinnerte mich an ein Interview mit Google CEO Eric Schmidt⁵, der prophezeite, dass es zukünftig vermutlich notwendig sei, öfter mal den Namen zu ändern, um den gespeicherten Teil der Vergangenheit loszuwerden. Darüber hinaus, so Unterburger, soll-

ten Techniken entwickelt werden, die das Streuen von gezielten Desinformationen vereinfachen und so verhindern, dass eindeutige Profile entstehen. Hier wird also versucht, an der Idee der *informationellen Selbstbestimmung* festzuhalten, wenn auch nicht im klassischen Sinne einer Kontrolle über die Daten, die preisgegeben werden, sondern durch die Beeinflussung der Informationsmenge, die kursiert. So verlagert sich die informationelle Selbstbestimmung auf die Einflussnahme des Bildes, das die Daten durch weitere Daten von einer Person erzeugen, statt nur die Begrenzung der Datenpreisgabe selbst in den Blick zu nehmen. Auch wenn ich die Initiative, Datenschutz und informationelle Selbstbestimmung anders zu denken, im Kern begrüße, ist einiges an Unterburgers Haltung zu kritisieren. Zuerst die Rede von der nicht aufzuhaltenden Naturgewalt „Kontrollverlust“ – ein Fatalismus den ich für ein grundsätzliches Problem der Post-Privacy-Szene halte. Gepaart mit der Euphorie um die Möglichkeiten von *Big Data* wird dabei „Datenschutz als Übergangstechnologie“ besprochen, auf dem Weg in die diskriminierungsfreie, fortschrittliche Zukunft. An Unterburgers Vortrag selbst störte mich auch die Haltung, die „Privatsphärenfans“ an den Rand der Gesellschaft rückt und als diejenigen darstellt, die mit ihrer Datenzurückhaltung dem Fortschritt entgegen stünden. Aus piratigen Toleranzgründen müsse man aber Übergangstechnologien entwickeln, die die Transformation in die (post-private) Mehrheitsgesellschaft vereinfachen: Da wird aus Fortschrittsgläubigkeit schnell ein Ausgrenzungsmechanismus und kritische Datenschützer_innen werden zu einer Gruppe degradiert, die nicht selbst überlebt. Nicht unüblich für CCC-nahe Veranstaltungen – 2008 hatte Constanze Kurz den Begriff „Post-privacy-Spacken“ geprägt – war die Kritik nach dem Vortrag aber auch relativ laut und teilweise unsachlich. Auch wenn gerade technikaffine Menschen Data Mining und automatisierten Analysen eher positiv gegenüberstehen, scheint der Schritt in die post-private Zukunft den meisten doch nicht so unumstößlich.

Die *Pentabarf-Gameshow* und die *Lightning Talks* gehören mit kistenweise Club Mate zum Rahmenprogramm eines entspannten Kongresses, wie es der C3 mit mehreren Tausend Teilnehmer_innen kaum noch sein kann.

Referenzen

Aufnahmen aller Vorträge finden sich auch online unter www.datenspuren.de.

Anmerkungen

- ¹ <http://www.datenspuren.de>
- ² <http://annalist.noblogs.org/mediathek/>
- ³ <http://www.security-informatics.de/blog/>
- ⁴ Gandy, O.H. (1993): *The Panoptic Sort – A Political Economy of Personal Information*. Westview Press, Boulder
- ⁵ <http://online.wsj.com/article/SB10001424052748704901104575423294099527212.html>