Erschienen in: Friedewald, M., Roßnagel, A., Lamla, J. (Hrsg.): *Informationelle Selbstbestimmung im digitalen Wandel.* Wiesbaden: Springer Vieweg, 2017.
http://dx.doi.org/10.1007/978-3-658-17662-4

Emission statt Transaktion

Weshalb das klassische Datenschutzparadigma nicht mehr funktioniert

Sven Türpe, Jürgen Geuter und Andreas Poller

1. Einleitung

Der Datenschutz hat angesichts von Big Data und Internet einen schweren Stand. Websites verfolgen ihre Nutzer, Werbung passt sich den Empfängern an, soziale Netze erkennen Gesichter in Fotos und Spielzeugpuppen schicken aufgezeichnete Gespräche zur Verarbeitung an einen Dienst im Internet.[1] Langatmige Datenschutzerklärungen und pauschale Einwilligungen sollen rechtliche Anforderungen erfüllen, schaffen tatsächlich aber wenig Transparenz und Kontrolle, sondern überschwemmen die Betroffenen mit einer solchen Menge an Text, dass eher das Gegenteil – totale Intransparenz – erreicht wird. Unterdessen werden Datenschutzbeauftragte und -aktivisten nicht müde, uns alle vor den Gefahren der sorglosen Internetnutzung zu warnen. Einerseits trägt fast jeder ein Smartphone in der Tasche, das »die Cloud« über sein Leben auf dem Laufenden hält, andererseits bestätigen wir artig jeder Website, dass wir wissen, dass sie Cookies einsetzt. Ist Datenschutz ein Relikt der Vergangenheit, eine Sammlung sinnentleerter Gesten?

In den 1970er und 1980er Jahren entstanden die bis heute geltenden Grundzüge des deutschen Datenschutzrechts. Damals waren Computer kaum mehr als automatisierte Karteischränke. Aus dieser Zeit und ihrer Technik stammt die Idee, jeder Person ein Verfügungsrecht über die sie betreffenden Daten zu gewähren und dieses Recht durch spezifische Entscheidungs- und Eingriffsbefugnisse der Betroffenen zu garantieren. Dazu diente eine Transaktionsschnittstelle zwischen Betroffenen und Verarbeitern: Daten sollen möglichst bei Betroffenen selbst erhoben werden; unabhängig von der Art der Erhebung ist in der Regel – sofern keine gesetzliche

* Sven Türpe, Andreas Poller | Fraunhofer-Institut für Sichere Informationstechnologie SIT | {sven.tuerpe;andreas.poller}@sit.fraunhofer.de
Jürgen Geuter | Unabhängiger politischer Informatiker | jg@juergengeuter.com

1 Barthélémy und Wilkens, »Reden und lauschen«.

Erlaubnis besteht – deren widerrufliche Einwilligung erforderlich; später erlauben Auskunfts- und Eingriffsrechte die Kontrolle; technische und organisatorische Maßnahmen stellen sicher, dass diese Vorgaben nicht unterlaufen werden.

Dieser Ansatz hat sich für die Informationstechnik seiner Entstehungszeit bewährt. Zu Datenerhebungen kam es nur gelegentlich und explizit, die Datenmengen waren überschaubar und in der Regel handelte es sich um Angaben, deren Inhalt und Bedeutung ohne Schwierigkeiten nachzuvollziehen war. Dazu passten die Unterschrift unter einen Hinweistext als Mittel der informierten Einwilligung und Regeln rund um die Speicherung von Daten.

Mangels Alternativen arbeiten wir bis heute mit solchen Mitteln, doch sie passen immer weniger. Die klassische Datenerhebung ist nur noch ein Spezialfall. Alltägliche Vorgänge wie der Besuch einer Website lösen im Hintergrund Interaktionen mit einem ganzen Netz von Akteuren aus, das sich mit jedem Klick verändern kann. Dabei entstehende Daten werden nicht mehr auf einer elektronischen Karteikarte zum späteren Abruf in Datenbanken abgelegt, sondern durch Data Mining und maschinelles Lernen zu statistischen Modellen und darauf gestützten Entscheidungen verarbeitet.

Die dem Transaktionsparadigma zugrundeliegenden Annahmen gelten oft nicht mehr. Stattdessen entstehen in unserer Umgebung fortlaufend Daten über jeden von uns, verbreiten sich und werden von teils unbekannten Empfängern mit komplexen Verfahren interpretiert. Die Ergebnisse und Auswirkungen lassen sich weder anhand der Daten noch anhand grober Verfahrensbeschreibungen einschätzen. Dies schwächt die herkömmlichen Kontrollmechanismen. Sie scheitern an der Häufigkeit der Erhebungsvorgänge, unklarer Dateninterpretation und inhärenter Intransparenz der Verarbeitung.

Daraus folgen Konflikte zwischen Vertretern des kodifizierten Datenschutzes, IT-Betreibern und Betroffenen, welche die öffentlichen Debatten zuweilen auf die Frage »Datenschutz oder Datenverarbeitung?« zuspitzen. Um dieser falschen Dichotomie zu entkommen, nehmen wir die Grundzüge der heutigen Informationstechnik ebenso als gegeben hin wie das Bedürfnis, Menschen und ihre Persönlichkeitsrechte vor Auswüchsen und unangemessenen Benachteiligungen zu schützen. Wir konzentrieren uns hier auf die technisch-organisatorische Dimension und stellen mit der Emissionsmetapher die Frage, auf welche Weise die Schutzziele[2] des Datenschutzes heute und morgen am effektivsten verfolgt werden können.

[2] Rost und Bock, »Privacy by Design und die neuen Schutzziele«.

2. Datenschutz als Informationskontrolle

Datenschutz übersetzt das abstrakte Recht auf informationelle Selbstbestimmung in Regeln, mittels derer Personen die Verarbeitung sie betreffender Informationen steuern und kontrollieren können. Datenschutz ist vor allem dort erforderlich, wo Interessen divergieren und die Auswirkungen von ungleichen Machtverhältnissen auszugleichen sind. Als Mittel bedient er sich einer Mischung aus sozialen Normen und Praktiken, rechtlichen Regelungen sowie technischen Maßnahmen.

Vorstellungen von Privatheit und die Mittel zu ihrem Schutz haben sich mit der Zeit gewandelt. Besonders der technische Fortschritt und damit verbundene Veränderungen der Gesellschaft verändern immer wieder Anforderungen, Randbedingungen und Lösungsansätze. Im Jahr 1890 motivierte die damals neue Technik der Fotografie Warren und Brandeis zu ihrem bis heute vielfach zitierten Aufsatz »The Right to Privacy«.[3] Ab den 1970er Jahren führte dann die heranreifende Informationstechnik zur Entwicklung des Datenschutzes, wie wir ihn bis heute kennen. In den USA entwickelte eine von Ware geleitete Kommission 1973 die Idee der Fair Information Practice.[4] In Deutschland bereitete ein Gutachten von Steinmüller et al.[5] bereits 1971 das Problem des Datenschutzes umfassend auf und zeichnete viele spätere Entwicklungen vor.

2.1. Transaktionale Datenkontrolle

Der klassische Datenschutz stellt Mittel bereit, mit denen Individuen Daten in fremder Hand unter Kontrolle behalten. Seine Begriffe und Regeln korrespondieren mit den zur Entstehungszeit vorherrschenden Mitteln der Datenverarbeitung, zentralisierten Datenbanken. Dazu passend hat sich ein transaktionales Paradigma des Datenschutzes entwickelt. Die transaktionale Kontrolle der Datenverarbeitung stützt sich auf die Idee, Personen ein Verfügungsrecht über die Verarbeitung aller sie betreffenden Daten zuzusprechen. Dieses Verfügungsrecht üben sie in Transaktionen mit den verarbeitenden Stellen aus. Jede Einwilligung und jede Datenerhebung beim Betroffenen, wie sie das Bundesdatenschutzgesetz (BDSG) als Regelfall vorsieht, ist eine Datentransaktion. Im weiteren Sinne üben Betroffene auch ihre Rechte auf Auskunft, Berichtigung, Löschung, usw. in Transaktionen mit den jeweiligen Verarbeitern aus.

Der transaktionale Datenschutz stützt sich also auf drei Säulen:

1. *Die informierte Einwilligung* gilt als Voraussetzung für die Datenerhebung und -nutzung. Damit werden Transaktionen explizit, so dass die Betroffenen sie zur

[3] Warren und Brandeis, »The Right to Privacy«.
[4] Ware, *Records, Computers and the Rights of Citizens*.
[5] Steinmüller u. a., *Grundfragen des Datenschutzes*.

Kenntnis nehmen, vollständig verstehen, ihre Konsequenzen abschätzen und sie ggf. auch verhindern können.

2. *Interventionsmöglichkeiten* erlauben den Betroffenen zum einen, Transaktionen rückgängig zu machen und zu einem früheren Zustand zurückzukehren, zum anderen auch die Korrektur von Fehlverhalten durch Datenentzug.

3. *Technische und organisatorische Maßnahmen* stellen sicher, dass zulässige Transaktionen möglich sind und dass Daten nur auf der Grundlage zulässiger Transaktionen verarbeitet werden können.

Diese Aspekte zeigen sich deutlich in einschlägigen Gesetzen wie dem BDSG.

2.2. Zugrundeliegende Annahmen

Die transaktionale Kontrolle über die Datenverarbeitung folgt einer einfachen Idee: Die Datenverarbeitung ist in der Regel nur gestattet, wenn die Daten das Ergebnis erlaubter und freiwilliger Transaktionen mit den Betroffenen sind. Auf diese Weise kann jeder den Überblick über die ihn betreffende Datenverarbeitung behalten und sie steuern. Damit das wirklich funktioniert, müssen jedoch einige Annahmen erfüllt sein:

- Die Bedeutung freigegebener Daten bzw. die Konsequenzen ihrer Verarbeitung lassen sich meist ohne Schwierigkeiten beurteilen.
- Risiken entstehen durch die Datenverarbeitung und verringern sich, wenn keine Daten vorliegen.
- Explizite Datentransaktionen sind in ihrem Umfang und ihrer Häufigkeit handhabbar.
- Zwischen Verarbeitern und Betroffenen existieren Beziehungen, die eine explizite Interaktion ermöglichen.
- Verschiedene Verarbeiter und Verarbeitungszwecke lassen sich ohne Schwierigkeiten voneinander trennen.

Diese Annahmen passen zur Informationstechnik und ihren Anwendungen zur Entstehungszeit des Datenschutzes in den 1970er bis in die 1990er Jahre. Daten repräsentierten vorwiegend Einzelangaben über Personen, Gruppen oder Sachen;[6] Computer bewahrten solche Daten zum späteren Abruf in strukturierten Datenbanken auf. Ware spricht demgemäß vom »computer-based record keeping«, das die

[6] Steinmüller u. a., *Grundfragen des Datenschutzes*.

lange Entwicklung der Verwaltung seit der Antike fortsetzt.[7] Daten zu verarbeiten bedeutete in erster Linie, sie nach verschiedenen Kriterien zu ordnen und sie anhand dieser Kriterien später abzurufen.

Steinmüller et al.[8] gingen von einem generischen Prozess der Informationsverarbeitung aus, der Vorgänge wie die Ermittlung, Erfassung, Speicherung, Veränderung, Ausgabe, Weitergabe und Löschung umfasste und damit den Verarbeitungsbegriff des BDSG vorwegnahm. Die Möglichkeiten und Risiken der eigentlichen Verarbeitung gespeicherter Daten durch Programme wurden zwar betrachtet und sie waren angesichts etwa der früh eingesetzten Rasterfahndung nicht zu übersehen. So betonten sowohl Ware als auch Steinmüller et al. neben dem einfacher werdenden Informationszugang auch die menschliches Vermögen übersteigende Verarbeitungskapazität des Computers als Gefahrenquelle. Jedoch erschien die weitergehende Verarbeitung als eine bloße Folge der Erfassung und Speicherung von Daten.

Für nicht in einer Datenbank enthaltene Datensätze entfielen auch die daran geknüpften Vorgänge. Erhoben wurden Daten oft durch das Ausfüllen von Formularen. Vernetzung bedeutete Fernzugriff auf die Daten anderer Organisationen. Insgesamt erschienen Computer noch vorwiegend als digitale Karteischränke, betrieben von Organisationen zur Unterstützung ihrer jeweiligen Aufgaben. Datentransaktionen waren oft eingebettet in transaktionale Alltagsvorgänge wie etwa eine Katalogbestellung bei einem Versandhaus.

2.3. Degenerierte Begriffe und Werkzeuge

Die wichtigsten Hilfsmittel des transaktionalen Datenschutzes, Erklärungen und Einwilligungen, begegnen uns heute allerorten, auf jeder Website. Dies ist formalen Anforderungen zu verdanken, die damit erfüllt werden. Jedoch wirken die traditionellen Mittel im Web kaum noch überzeugend. Man hat sie, weil man muss, aber sie bewirken wenig. Dies hat mehrere Gründe:

Vielfältige Empfänger: Man schaue sich etwa die *PayPal-Datenschutzgrundsätze*[9] an. Sie umfassen eine als eigenes Dokument bereitgestellte Tabelle[10] aller Firmen, an die PayPal personenbezogene Daten übermittelt. Dieses Dokument hat gegenwärtig 60 Seiten und ist damit im Alltag nutzlos. Weder vermittelt es einen Eindruck von den tatsächlichen Datenflüssen und Verarbeitungsvorgängen, noch

---

7 Ware, *Records, Computers and the Rights of Citizens*.

8 Steinmüller u. a., *Grundfragen des Datenschutzes*.

9 PayPal: PayPal-Datenschutzgrundsätze, https://www.paypal.com/de/webapps/mpp/ua/privacy-full

10 PayPal: Liste der Dritten (außer PayPal-Kunden), für die personenbezogene Daten freigegeben werden können, https://www.paypalobjects.com/webstatic/de_DE/ua/pdf/paypal_third_party_disclosure_list_as_of_feb_23_2016.pdf.

sind Eingriffe eines Betroffenen praktikabel, selbst wenn ein verantwortlicher Betreiber als einziger Ansprechpartner fungiert. Welche informierten Entscheidungen sollten Betroffene auf dieser Grundlage treffen? Welche wirksamen Schranken sind der Datenverarbeitung noch auferlegt?

Vernetzte Dienste: PayPal ist kein Ausnahmefall, überall im Web finden sich ähnliche Verhältnisse. Abb. 1 illustriert dies anhand einer Browsersitzung, in der einige Websites besucht wurden. Kreise stellen die besuchten Sites dar, Dreiecke die zusätzlich im Hintergrund aufgerufenen Adressen. Jeder Dienst in diesem Netz erhält Daten über den Nutzer, oft auch über mehrere besuchte Websites hinweg. Damit verbunden ist die Möglichkeit zum Nutzer-Tracking[11] mittels Cookies oder Browser-Fingerprinting.[12] Auch hier leisten förmliche Datenschutzerklärungen – und erst recht die allseits bekannten Cookie-Banner – wenig für die informationelle Selbstbestimmung. Browsererweiterungen wie der PrivacyBadger[13] der EFF versprechen zu einem gewissen Grad Abhilfe, aber ihr Nutzen ist angesichts intransparenter Ursache-Wirkung-Beziehungen schwer zu verdeutlichen.

Indirekte Datenflüsse: Beim Web-Tracking fließen die Daten zumindest noch direkt zu den einzelnen Empfängern, so dass technische Eingriffe des Nutzers möglich bleiben. Anders ist dies, wenn wir beispielsweise Kontaktdaten an eine bekannte Person weitergeben, die sie im Adressbuch ihres Smartphones speichert. Dort sind sie für andere Apps zugänglich, welche die Daten wiederum weitergeben können. Hier gehen Daten nacheinander durch mehrere Hände und geraten außer Kontrolle, ohne dass Betroffene viel dagegen tun könnten. Solche Datenflüsse haben ähnliche Auswirkungen wie die Verknüpfung von Datenbanken, sie kommen jedoch anders zustande.

Datenmischung: In einer Datenbank oder auch in einer Verknüpfung mehrerer Datenbanken ist der Personenbezug jedes Datensatzes klar. Anders in sozialen Netzen: Dort kann eine Nachricht zum Beispiel beim Check-in in einem Café zum Treffen mit Freunden entstehen und dann von weiteren Nutzern bewertet, geteilt und

[11] Unter http://datenblumen.wired.de und mit dem Browserplugin Lightbeam (https://www.mozilla.org/lightbeam/) kann man sich Visualisierungen der eingebundenen Fremdinhalte anschauen. Vgl. auch Roesner, Kohno und Wetherall, »Detecting and Defending Against Third-Party Tracking on the Web«.

[12] Beim Fingerprinting versucht man, Eigenheiten der Browserkonfiguration und des Browserverhaltens als implizites Pseudonym zu nutzen. Dies gelingt erstaunlich oft und vermeintlich die Privatsphäre schützende Handlungen können dabei sogar helfen, wenn sie den Fingerabdruck eindeutiger machen. Vgl. Eckersley, »How Unique Is Your Web Browser?«

[13] https://www.eff.org/privacybadger

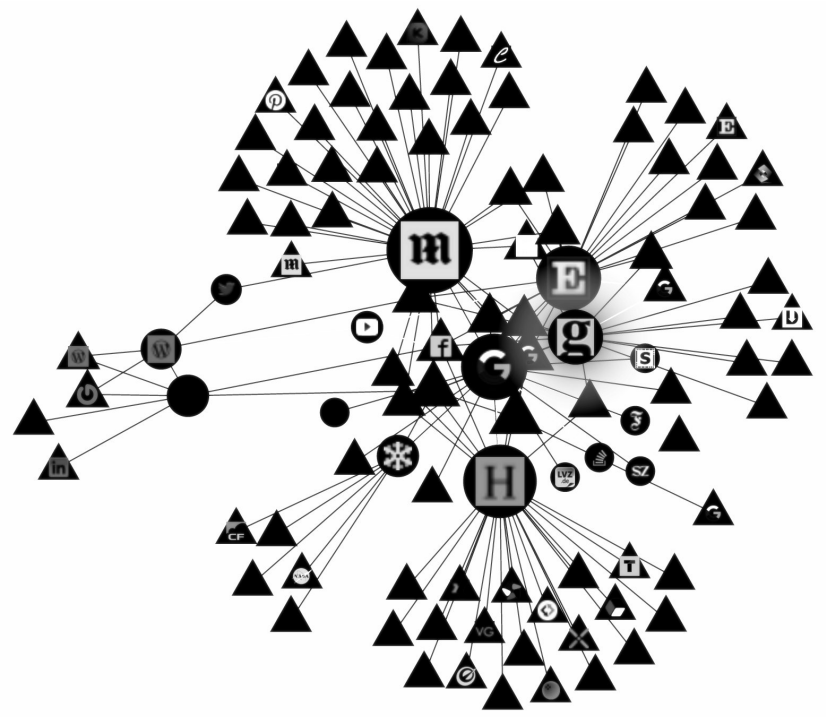

Abbildung 1.: Visualisierung einer Browsersitzung mit Lightbeam (https://www.mozilla.org/lightbeam/). Zu den besuchten Websites (Kreise) kommt ein Vielfaches an im Hintergrund aufgerufenen Adressen (Dreiecke).

kommentiert werden. Es entstehen Mischdaten mit Bezügen zu allen Beteiligten. Wer soll welche Verfügungsgewalt über diese Daten bekommen? Ist es angemessen, jeden Beteiligten aufgrund des jeweiligen Personenbezugs mit einem Vetorecht auch auf Kosten aller anderen auszustatten?

Latente, implizite und kontextuelle Personenbezüge: Personenbezüge entstehen nicht mehr nur durch explizite Zuordnung von Daten zu Identitäten. Ein Beispiel ist die Assoziation zwischen Suchbegriff und eingeblendeter Werbung in Suchmaschinen, wenn der Suchbegriff ein Name ist. Dokumentiert ist etwa ein Fall, in dem die Namen von Schwarzen in der Google-Suche systematisch zu ande-

ren Werbetexten führten als die von Weißen.[14] Die betreffende Werbung pries Background-Checks in Polizeidatenbanken an und war suggestiv formuliert. Das generische Verfahren, Werbung anhand der eingegebenen Suchbegriffe auszuwählen, wird durch die Eingabe eines Namens personalisiert. Ein weiteres Beispiel sind automatisch generierte Vorschläge zur Vervollständigung von Sucheingaben, die bei der Eingabe eines Namens Aussagen über die betreffende Person suggerieren können.[15]

3. Die digitale Sphäre heute

Zunehmende Datenmengen alleine stellen das Transaktionsparadigma nicht in Frage. Der technische Fortschritt hat jedoch nicht nur Datenmengen vergrößert, sondern er stellt die Annahmen in Frage, auf die sich der transaktionale Datenschutz stützt. Dazu tragen drei Entwicklungen bei: die Arbeitsteilung in vernetzten Systemen, die massenhafte Entstehung von Kollateraldaten sowie die Verfügbarkeit von Inferenz- und Lernalgorithmen für das Data Mining.

3.1. Arbeitsteilung in vernetzten Systemen

Von einem reinen Transportmechanismus für Daten hat sich das Internet zu einer Plattform für die Zusammenarbeit aller möglichen Akteure weiterentwickelt. Dies schlägt sich insbesondere in den technischen Architekturen von Anwendungen und den damit korrespondierenden Organisationsstrukturen nieder. Früher waren Anwendungsprogramme monolithische und statische Softwarepakete, die auf isolierten Systemen liefen oder höchstens einen Fernzugriff erlaubten. Heutige Anwendungen sind Cloud-Dienste, die im Hintergrund eine Reihe anderer Dienste einbinden und orchestrieren und die im laufenden Betrieb weiterentwickelt werden. Die einzelnen Dienste haben verschiedene Betreiber und begegnen Nutzern in unterschiedlichen Zusammenstellungen und Kontexten. Jede moderne Anwendung ist ein Fenster in ein Geflecht technischer Dienste und geschäftlicher Beziehungen.

Ein scheinbar einfacher Vorgang wie der Besuch einer Website umfasst in Wirklichkeit komplizierte Interaktionen mit einer Vielzahl von Akteuren und ihren technischen Artefakten, die größtenteils verborgen bleiben. Nutzungsvorgänge lassen bei allen Akteuren Daten entstehen. Explizite Interaktion und explizite Beziehungen bilden dabei nur die sichtbare Oberfläche, wie in Abb. 1 zu sehen. Analog verknüpft jedes Smartphone mit seinen Apps eine Sammlung von Diensten, die sich mit jeder Installation oder Deinstallation ändert.

---

[14] Sweeney, »Discrimination in online ad delivery«.
[15] Bager, »Rotlicht-Gerüchte«.

3.2. Kollateraldaten[16]

Explizite Angaben und Inhalte machen nur noch einen Teil der verfügbaren und verarbeiteten Daten aus. Daneben erzeugt die Informationstechnik vielerlei Daten als Nebeneffekt ihrer Funktionen, oft ohne besonderen Zweck. Die beim Browser-Fingerprinting[17] und Web-Tracking[18] verwendeten Daten fallen beispielsweise zu einem großen Teil in diese Kategorie: Cookies, Besonderheiten von Protokollabläufen, usw. Explizite Angaben existieren weiter, etwa wenn wir Formulare ausfüllen, Nachrichten posten oder Bilder hochladen. Die explizite Übermittlung, die »Datenerhebung« ist jedoch wiederum nur noch ein Oberflächenphänomen.

Kollateraldaten unterscheiden sich von expliziten Angaben zum einen dadurch, dass sie fortlaufend oder mit hoher Frequenz entstehen. Zum anderen haben sie für sich genommen oft keine erkennbare Bedeutung für diejenigen, die von der Verarbeitung dieser Daten betroffen sein können, während sie gleichwohl für andere interessant und nützlich sein können. Jeder Faktor für sich und erst recht die Kombination erschweren die Steuerung durch Datentransaktionen.

3.3. Data Mining und maschinelles Lernen

Eng verbunden mit der Verarbeitung großer Datenmengen sind Verfahren zum Data Mining und zum maschinellen Lernen. Sie erfassen statistische Regelmäßigkeiten einer großen Menge bekannter Datensätze und repräsentieren sie in einem Modell oder Entscheidungsalgorithmus, mit dem sich weitere, vorher unbekannte Datensätze interpretieren lassen. Insbesondere Lernverfahren sind heute eine Standardtechnologie, die in immer mehr IT-Systemen zu finden. Wir skizzieren hier nur einige Begriffe; für tiefere Erklärungen verweisen wir auf die umfangreiche Literatur zum Thema.[19]

Inferenz ist das Schließen aus Daten unabhängig vom angewandten Verfahren. Beispielsweise kann man von einer IP-Adresse häufig auf den ungefähren Standort des dazugehörigen Computers schließen. Welche Schlüsse aus Daten gezogen werden, hängt nicht nur von den Daten selbst ab, sondern auch von den angewandten Schlussregeln und der Wissensbasis, auf der diese Regeln angewendet werden.

---

[16] Wir meinen damit Daten, die aus Nutzersicht nebenbei und im Hintergrund anfallen. Zum Teil (z. B. Cebulla, »Umgang mit Kollateraldaten«) wird die Bezeichnung *Kollateraldaten* auch für Daten von Dritten gebraucht.

[17] Eckersley, »How Unique Is Your Web Browser?«

[18] Roesner, Kohno und Wetherall, »Detecting and Defending Against Third-Party Tracking on the Web«; Acar u. a., »The Web Never Forgets«.

[19] Besonderheiten des maschinellen Lernens fasst dieser Artikel gut zusammen: Domingos, »A few useful things to know about machine learning«.

Die *statistische Inferenz* gewinnt ihre Schlussregeln mit statistischen Mitteln. Die statistischen Eigenschaften einer Stichprobe werden verallgemeinert und auf ähnliche Daten angewandt, um (mutmaßlich) näherungsweise richtige Ergebnisse zu berechnen. Statistische Inferenzverfahren liefern im Idealfall die bestmögliche Schätzung anhand der vorliegenden lückenhaften Informationen.

Das *maschinelle Lernen* ist eine Spielart der statistischen Inferenz. Anstelle klassischer statistischer Verfahren verwendet man Algorithmen,[20] um anhand einer Stichprobe von Trainingsdaten die Parameter eines Modells anzupassen. Dazu werden Daten und Modelle in vieldimensionalen Räumen repräsentiert und das Modell so angepasst, dass es auf den Trainingsdaten möglichst gute Ergebnisse liefert. Auf diese Weise erhält man Algorithmen, die zu einem Teil vom Menschen entworfen, zu einem anderen Teil automatisch an Daten angepasst sind.

Das *Crowdsourcing von Problemlösungen* an eine Nutzerpopulation ist ein Anwendungsbereich des maschinellen Lernens, der die aus dem Verbraucherscoring[21] bekannten Ansätze verallgemeinert und in interaktiven Systemen einsetzt. Internetdienste wie zum Beispiel Suchmaschinen lernen zunehmend von ihren Nutzern: welche Eingaben wahrscheinlich Tippfehler enthalten und wie diese zu korrigieren sind, welche Fortsetzungen einer begonnenen Eingabe häufig vorkommen, welche Suchergebnisse am besten ankommen oder auch welche Werbung in welchem Kontext am häufigsten funktioniert. Das beobachtete Verhalten der Nutzerpopulation schlägt sich so im künftigen Verhalten des Dienstes nieder.

Der breite Einsatz statistischer Inferenzverfahren hat mehrere Konsequenzen für die Datenverarbeitung. Erstens lässt sich die Bedeutung und Aussagekraft von Daten nicht mehr verlässlich aus diesen Daten selbst ermitteln, sondern erst mit Bezug auf konkrete Modelle und Algorithmen zu ihrer Verarbeitung. Was ein Datenverarbeiter »weiß«, ist zu einem großen Teil in seinen Modellen kodiert und nicht in den ihm übermittelten Daten. Das Zurückhalten oder Verändern von Daten kann dabei unerwartete Wirkungen haben, die der damit verbundenen Absicht zuwider laufen.[22] Zweitens haben Inferenzergebnisse den Charakter einer Schätzung oder Vermutung; es handelt sich nicht um gesicherte Fakten, sondern um Wahrscheinlichkeitsaussagen unter gewissen Annahmen. Dies ist für viele Anwendungen ausreichend, jedoch lassen sich Wahrscheinlichkeitsaussagen zum Beispiel im Einzelfall schwer widerlegen.

---

[20] Breiman, »Statistical modeling«.

[21] Kamp und Weichert, *Scoringsysteme zur Beurteilung der Kreditwürdigkeit*.

[22] Webbrowser schicken bei jeder Interaktion mit einem Webserver eine Zeichenfolge mit, die Produkt und Version erkennen lässt. Eckersley, »How Unique Is Your Web Browser?« erläutert, dass das Verändern dieser Produktkennung auf Phantasiewerte in der Absicht, weniger Informationen preiszugeben, beim Fingerprinting ein weiteres Merkmal liefert und die Browserinstanz so leichter unterscheidbar macht.

Drittens arbeiten maschinelle Lernverfahren ohne strenge Theoriebildung. In gewissen Grenzen ermitteln und repräsentieren sie statistische Signale in den jeweiligen Trainingsdaten unabhängig davon, woher diese Signale rühren. Viertens repräsentieren trainierte Modelle anders als eine Datenbank nicht die verwendeten Trainingsdaten selbst, sondern Aussagen über die Gesamtheit ihrer Trainingsdaten. Die klassischen Datenmanipulationen – Einfügen, Ändern, Löschen – sind jedenfalls in ihrer herkömmlichen Form nicht möglich. Fünftens verallgemeinern Lernverfahren aus Stichproben und die Auswirkungen auf eine Person hängen nicht davon ab, ob diese Person zur Stichprobe beigetragen hat.

3.4. Schwächen des Transaktionsparadigmas

Nach diesen Ausführungen wird deutlich, woran transaktionsorientierte Datenschutzverfahren heute scheitern: Die zugrundeliegenden Annahmen sind nicht mehr erfüllt. Tabelle 1 stellt die IT-Ökosysteme von damals und heute gegenüber, die Welt der Datenbanken aus der Entstehungszeit unseres Datenschutzes sowie die Welt von heute.

	Annahmen des transaktionalen Datenschutzes	**Heutige Realität**
Typische Technologien	• Datenbanken • Formulare • Berichte	• Webanwendungen • Verteilte Systeme • Vernetzte Geräte überall • Maschinelles Lernen
Verarbeitungsparadigma	• Speichern strukturierter Daten • Auswahl und Abruf von Datensätzen nach einfachen Kriterien	• Statistische Modellierung und Inferenz • Optimierte Entscheidungen in Echtzeit • Verarbeitung von Datenströmen
Häufigkeit der Erhebungs- und Verarbeitungsvorgänge	• Gering bis moderat	• Hoch bis fortlaufend

	Annahmen des transaktionalen Datenschutzes	**Heutige Realität**
Wahrnehmbarkeit der Datenübertragung	• Explizit (z.B. Ausfüllen eines Formulars)	• Explizit oder implizit • Kollateraldaten
Dateninterpretation	• Wörtlich, direkt • Fakten, Einzelangaben • Aussagen über Personen	• Direkt oder gefolgert • Datenpunkte in Merkmalsräumen • Ähnlichkeit • Probabilistisch
Primäre Risiken	• Missbrauch gespeicherter Daten durch Verarbeiter • Unerlaubter Zugriff, unerlaubte Übermittlung	• Missbrauch von gespeicherten Daten oder Schlussfolgerungen • Fehlinterpretation • Diskriminierung • Folgen von Vorhersagen
IT-Ökosystem	• Wenige isolierte Verarbeiter	• Viele vernetzte Akteure
Datenverarbeiter	• Einzelne, bekannte Einrichtungen	• Vielfältige, oft unsichtbare Akteure
Beziehung zwischen Betroffenen und Verarbeitern	• Explizit • Direkt (first-party)	• Oft implizit • Indirekt (third-party)
Kardinalität der Beziehungen (Betroffene–Verarbeiter)	• Viele Betroffene, wenige Verarbeiter	• Alle betroffen, viele Verarbeiter

	Annahmen des transaktionalen Datenschutzes	**Heutige Realität**
Häufigkeit von Beziehungsänderungen	• Selten	• Häufig (z.B. bei jedem Website-Besuch)

Tabelle 1.: Datenverarbeitung damals und heute

4. Die Emissionsmetapher

Wollen wir Privatheit und Selbstbestimmung weiter schützen, so müssen wir die veränderten Bedingungen beachten und Mittel suchen, die unter diesen Bedingungen wirksam sind. Als ersten Schritt fassen wir die Verhältnisse in der Metapher der fortlaufenden Datenemission zusammen:

> *Jeder Einzelne sendet ähnlich einer Lichtquelle Daten aus, fortlaufend und in alle Richtungen. Über emittierte Daten hat der Sender keine Kontrolle mehr.*

Diese Metapher bringt Eigenschaften des eng vernetzten Internet-Ökosystems zum Ausdruck und erleichtert damit die Analyse und Diskussion ihrer Konsequenzen.

4.1. Vorbild: Optik

Eine Metapher überträgt Begriffe und Beziehungen ihres Herkunftsbereichs in ein anderes Gebiet.[23] Die Emissionsmetapher entstammt der Optik.[24] Lichtquellen unterschiedlicher Gestalt und Größe senden elektromagnetische Wellen aus, die sich im Raum in alle Richtungen ausbreiten. Strukturen im Raum – Objekte und Regionen aus unterschiedlichen Medien – beeinflussen die Fortpflanzung der Lichtwellen, indem sie Licht absorbieren, reflektieren oder brechen. Stellt man irgendwo im Raum einen Sensor auf, so empfängt dieser direkt und indirekt Lichtwellen aus verschiedenen Quellen, deren Aussendungen sich mischen; Licht aus derselben Quelle kann auch auf unterschiedlichen Wegen zum Sensor gelangen. Auch ohne explizite Aufmodulation von Signalen tragen die an einem Ort eintreffenden Wellen Informationen, nämlich über die Lichtquellen und die Struktur des umgebenden Raums. Optische Instrumente – Mikroskope, Kameras, Projektoren, usw.

[23] Lakoff und Johnson, *Metaphors we live by*.
[24] Siehe z. B. Pedrotti u. a., *Optik für Ingenieure*.

– nutzen dies aus, um gezielt optische Abbildungen zu erzeugen. Dabei werden Lichtquellen oft gezielt eingesetzt, um andere Gegenstände zu beleuchten. Die beleuchteten Gegenstände interagieren mit dem einfallenden Licht, indem sie es teils reflektieren, teils brechen und teils absorbieren; dies erlaubt Rückschlüsse auf ihre Eigenschaften.

Mit der Emissionsmetapher übertragen wir drei Ideen in den Datenschutz:

1. Die Idee der Quelle, die fortlaufend in alle Richtungen abstrahlt.
2. Die Idee eines strukturierten Raums, in dem sich Lichtwellen fortpflanzen und von Objekten und Medien beeinflusst werden.
3. Die Idee abbildender Systeme, die irgendwo im Raum Lichtwellen auffangen und in einer für ihren Benutzer hilfreichen Weise abbilden.

Die spezifischen Naturgesetze der Optik lassen sich natürlich nicht einfach auf Daten im Internet übertragen, dafür sind die Unterschiede zwischen dem physikalischen Raum einerseits und dem »Cyberspace« andererseits zu groß.

4.2. Übertragung auf Daten

In einer Welt voller vernetzter Geräte und Sensoren verhalten sich Daten im Netz ähnlich wie elektromagnetische Wellen im Raum:

1. Vielfältige Vorgänge erzeugen fortlaufend Daten, die Informationen über Personen tragen. Häufig handelt es sich um Kollateraldaten, deren Entstehung wir nicht bemerken und die wir nicht ohne weiteres interpretieren können. Als kanonisches Beispiel kann das Smartphone dienen, dessen Nutzung im Alltag alle möglichen Daten über seinen Besitzer wie auch über andere Personen entstehen lässt.
2. Technische und organisatorische Strukturen im Internet-Ökosystem bestimmen die Verbreitung entstandener Daten. In der Regel gelangen Daten zu einer Vielzahl von Akteuren, von denen nur wenige explizit in Erscheinung treten. Auf dem Weg zu einem Empfänger können Daten transformiert werden.
3. Jeder Empfänger erhält Daten von einem Nutzerkollektiv. Welche das sind und in welcher Form, hängt von seiner Position und den umgebenden Strukturen ab. Zur weiteren Verarbeitung setzen Datenempfänger verschiedene Instrumente ein, welche die eintreffenden Daten für ihre Zwecke aufbereiten. Welche Bedeutung Daten bekommen, hängt nicht zuletzt von diesen Instrumenten ab.

Diesen Grundzügen mag man weitere Übertragungen hinzufügen, etwa der Unterscheidung zwischen dem sichtbaren und dem nicht sichtbaren Spektrum, der Idee unterschiedlich hoher Energie und damit Reichweite oder auch der Vorstellung einer künstlichen »Beleuchtung« mit dem Ziel, auswertbare Datenemissionen hervorzurufen.

4.3. Datenemission als Normalzustand

Die Emissionsmetapher hilft uns, bekannte Phänomene der digitalen Welt zu fassen, etwa das Tracking und Profiling im Web oder das Teilen von Inhalten und Daten in sozialen Netzen.

Das Web-Tracking[25] kann in den herkömmlichen Begriffen als Erhebung, Verarbeitung und Übermittlung von Nutzungsdaten beschrieben werden. Daran knüpfen sich die sattsam bekannten Diskussionen etwa um die Einwilligung der Betroffenen, um Cookies und IP-Adressen sowie um die Zulässigkeit des Einsatzes zum Beispiel von Google Analytics. Häufig wird jedoch nur der vereinfachte Spezialfall betrachtet, dass ein einzelner Website-Betreiber einen einzelnen Tracking-Dienst einbindet. Tatsächlich – siehe Abb. 1 – bewegen sich Nutzer jedoch in einem Netz von Akteuren mit vielfältigen Beziehungen, dem notdürftige Konstrukte wie das der Auftragsdatenverarbeitung nicht gerecht werden.

Die Emissionsmetapher hingegen erklärt die Tracking-Möglichkeiten zum technischen und organisatorischen Normalfall und wirft so die Frage auf, welche Risiken sich ergeben und welcher Schutzbedarf daraus resultiert:

Die Nutzung eines Webbrowsers generiert fortlaufend Daten, bei jedem Klick und sogar ohne Nutzeraktivität beim bloßen Betrachten. Diese Daten, etwa in Form von HTTP-Requests, sind kleinteilig und zunächst vor allem technischer Natur. Aggregiert und transformiert mögen sie etwas über den Nutzer aussagen; in der Form, in der sie entstehen, würde sie der Nutzer jedoch selbst nie angeben. Auch entstehen diese Daten nicht an einem klar definierten Ort und werden dann weitergegeben, sondern sie werden von einem verteilten System erzeugt. Oft lassen sich Daten gar nicht mehr klar einem einzelnen Nutzer zuordnen. So kommen etwa beim Teilen und Kommentieren in sozialen Netzen explizite und Kollateraldaten aller beteiligten Nutzer zusammen und lassen sich nicht sauber trennen.

Eine unbestimmte Menge von Akteuren ist in der Lage, die entstehenden Daten oder Teile davon zu empfangen. Neben den Betreibern der sichtbaren und explizit besuchten Websites gehören dazu beispielsweise deren Dienstleister (z. B. Content Delivery Networks, Analytics, Sicherheit, Bezahlen), aber auch Akteure wie die sozialen Medien oder Werbenetze, deren Funktionen und Inhalte vielfach in Websites

[25] Roesner, Kohno und Wetherall, »Detecting and Defending Against Third-Party Tracking on the Web«; Acar u. a., »The Web Never Forgets«; Steidle und Pordesch, »Im Netz von Google«.

eingebettet sind. In der Regel handelt es sich nicht um einfache Hierarchiebeziehungen, sondern um komplizierte Kollaborationsnetze. Zusätzlich können auch Dritte wie Browserhersteller oder Nachrichtendienste entstehende Daten empfangen.

Datenempfänger setzen verschiedene Verfahren und Algorithmen als Instrumente ein, um die erhaltenen Daten aufzubereiten und zu nutzen, zum Beispiel Profiling-Verfahren.[26] Diese Instrumente unterscheiden sich unter anderem hinsichtlich der Persistenz oder Flüchtigkeit der Eingangs- und Ausgangsdaten, der Individualität oder Kollektivität ihrer Projektionen, sowie hinsichtlich der Rückwirkungen auf Personen oder Gruppen und ihrer Abhängigkeit von eigenem und fremdem Verhalten. So betrachtet zum Beispiel eine einfache Reichweitenmessung für Websites nur statistische Aggregate und hat keine direkten Auswirkungen auf die Nutzer, deren Nutzungsdaten in die Messung einfließen. Empfehlungssysteme und Mechanismen zur Preisdifferenzierung im Online-Handel andererseits wirken sich unmittelbar auf einzelne Nutzer und oft auch auf Nutzergruppen aus. Eine bloße Protokollierung des Nutzerverhaltens hat zunächst gar keinen Effekt, lässt jedoch die künftige Nutzung offen.

4.4. Privatheit und Selbstbestimmung trotz Datenemission?

Unsere Emissionsmetapher betont die Randbedingungen, unter denen die Ziele des Datenschutzes heute und in Zukunft verwirklicht werden müssen. Erstens wird die detaillierte Steuerung der Datenverarbeitung durch Betroffene immer weniger praktikabel. Oberflächliche technische Phänomene wie Cookies und IP-Adressen erlauben kaum noch Rückschlüsse auf die eingesetzten Interpretationsinstrumente und ihre Auswirkungen, und erst recht keine gezielten Eingriffe. Zweitens macht die Vielzahl von Datenempfängern empfängerspezifische Deklarationen und Einwilligungen impraktikabel. Jedoch besteht zwischen den Akteuren auch keine klare Hierarchiebeziehung, die alles an einem Punkt zusammenführen könnte. Drittens hängen Auswirkungen und Risiken der Datenverarbeitung wesentlich von den eingesetzten Verarbeitungsinstrumenten ab. Die erforderlichen Differenzierungen lassen sich nicht alleine aus den verwendeten Daten und einer groben Zweckbestimmung herleiten.

Während die Ziele des Datenschutzes gültig bleiben, müssen sich seine Mittel diesen Bedingungen anpassen. Als weiteres Ziel kommt noch die Diskriminierungsfreiheit hinzu. Nach dem Transaktionsparadigma folgt Transparenz aus expliziten Transaktionen, aus Benachrichtigungen und der erkennbaren Bedeutung von Daten. Interventionen richten sich auf gespeicherte Daten und beeinflussen deren Existenz,

[26] Schermer, »The limits of privacy in automated profiling and data mining«; Hildebrandt und Gutwirth, *Profiling the European Citizen*.

Umfang und Inhalt. Technische Maßnahmen wie Zugriffskontrolle und Protokollierung verhindern (oder dokumentieren) unerlaubte Zugriffe. Der Rechtsrahmen beschäftigt sich mit Daten, ihrer Speicherung und Verarbeitung.

Unter den Annahmen der Datenemission richtet sich die Forderung nach Transparenz auf Modelle, Verarbeitungsergebnisse und Auswirkungen statt auf die Inhalte von Speichermedien. Wirksame Interventionen müssen Verarbeitungsergebnisse im Einzelfall sowie die Systematik ihres Zustandekommens beeinflussen. Technische Maßnahmen unterstützen zum einen Aufsicht und Intervention, zum anderen steuern sie das Schlussfolgern aus Daten. Der Rechtsrahmen schützt Menschen statt Daten, zum Beispiel vor Diskriminierung.

5. Ausblick

Wie lassen sich die Ziele des Datenschutzes verwirklichen, wenn fortlaufend Daten entstehen und von einem großen Empfängerkreis mit vielfältigen, komplizierten Instrumenten genutzt werden? Wir zeigen abschließend einige Tendenzen auf.

5.1. Vermeidung von Kollateraldaten

Die etablierte Idee der Datensparsamkeit erhält im Emissionsparadigma eine neue Bedeutung. Um wirksam zu sein, muss sie die Entstehung und Aussendung von Kollateraldaten so umfassend beeinflussen, dass die Verwendung unerwünschter Beobachtungsinstrumente nicht mehr sinnvoll möglich ist. Grundsätzlich lassen sich Kollateraldaten durch sorgfältigen Protokollentwurf vermeiden, wie es die eID-Funktionen des Personalausweises demonstrieren.[27] Inwieweit die Übertragung solcher Ideen auf Universalprotokolle wie HTTP und TLS gelingen kann, bleibt abzuwarten. Ansätze mit begrenzter Wirkung zeigen sich in der Browsererweiterung PrivacyBadger[28] der EFF und im in Deutschland recht weit verbreiteten Programm Shariff.[29] Beide versuchen das Tracking-Potenzial eingebetteter Social-Media-Buttons auf Websites zu verringern, PrivacyBadger im Browser und Shariff auf der Serverseite. Lassen sich solche Ansätze verallgemeinern?

5.2. Regulierte Instrumente

Ist die Datenemission als solche nicht zu vermeiden, so müssen sich Schutzvorkehrungen auf die Instrumente der Datenbeobachtung und Datennutzung konzentrieren.

---

[27] Poller u. a., »Electronic Identity Cards for User Authentication«.
[28] https://www.eff.org/privacybadger
[29] https://github.com/heiseonline/shariff

Bereits jetzt richten sich die Vorgaben des Datenschutzes an die verarbeitenden Stellen. Mit Ausnahme spezifischer Regelungen zum Scoring sind sie jedoch fokussiert auf die Datenspeicherung als Kern der Verarbeitung. Instrumente wie Data Mining und maschinelles Lernen werfen neue Fragen auf und erfordern eigenständige Vorgaben, deren Entwicklung auf langjährige Arbeiten zum Profiling[30] und Scoring[31] aufbauen kann. Wie lassen sich Inferenzen und die zugrundeliegenden Modelle verständlich machen, sei es für Betroffene selbst oder für Auditoren? Wie kann man Wahrscheinlichkeitsaussagen prüfen und wie wird algorithmische Diskriminierung nachweisbar? Welche Risiken sind überhaupt typisch? Die Antworten auf solche Fragen werden häufig von der betrachteten Anwendung abhängen. Welche Taxonomie erfasst die wesentlichen Aspekte?

5.3. Personen- und gruppenbezogene Risiken

Mit den Instrumenten der Datenverarbeitung rücken ihre Auswirkungen auf Personen und Gruppen ins Blickfeld. Derselbe Datenstrom kann mit den verschiedensten Zielen und Verfahren interpretiert werden. Dabei können einerseits ursprünglich sensible Daten zu einem risikoarmen Ergebnis führen, etwa wenn die Eingaben von Suchmaschinennutzern detailliert aufgezeichnet werden, um daraus am Ende ein sprachspezifisches Modell zur Tippfehlerkorrektur zu gewinnen. Andererseits können anhand anonymer oder für sich belanglos scheinender Daten Entscheidungen[32] fallen, die sich aufgrund ihres Kontextes auf Personen oder Gruppen auswirken, zum Beispiel wenn ein Online-Shop anhand von Merkmalen eines Nutzungsvorgangs seine Produkte zu unterschiedlichen Preisen anbietet.

Gesucht sind differenzierte Kriterien zur Risikobewertung und risikoorientierte Schutzmaßnahmen im jeweiligen Anwendungskontext. Einfache Ansätze wie die Unterscheidung zwischen personenbezogenen und anonymisierten Daten scheitern an der Vielfalt und Komplexität der Verarbeitungsmöglichkeiten. Zu beantworten sind Fragen wie diese: Ein Werbenetz führt pseudonymisierte Nutzerprofile über viele Websites hinweg und verwendet sie, um jeweils die am wahrscheinlichsten erfolgreiche Anzeige aus einem Pool auszuwählen – welche Risiken folgen daraus und was ist dagegen zu tun?

5.4. Unterstützung von Privatheitspraktiken

Als Gegenentwurf zur Datenvermeidung können komplexe Datenformate so angereichert werden, dass sie Praktiken des Privatheitsschutzes unterstützen. So

[30] Hildebrandt und Gutwirth, *Profiling the European Citizen*.
[31] Kamp und Weichert, *Scoringsysteme zur Beurteilung der Kreditwürdigkeit*.
[32] Pohle, »PERSONAL DATA NOT FOUND«.

haben sich beispielsweise für das Fotografieren von Personen in öffentlichen und teilöffentlichen Räumen und die Verwendung solcher Fotografien Regeln etabliert. Allgegenwärtige Kameras, zum Beispiel in Smartphones, und ihre typische Nutzung erschweren jedoch das explizite, transaktionsartige Einholen von Zustimmungen. Andererseits sind weitreichende Pauschalverbote angesichts der einfachen Verbreitung von Fotos im Netz nicht realistisch.

Das Projekt Offlinetags[33] zeigt beispielhaft, wie sich bestehende Praktiken in der digitalen Welt unterstützen lassen. Ein Offlinetag ist ein Button zum Anstecken an die Kleidung, der die Einstellung einer Person zu Fotos von sich wiedergibt. Zur Auswahl stehen vier Policies: (1) keine Fotos bitte, (2) Person vor Weiterverbreitung unkenntlich machen, (3) Fotos im Netz sind OK, aber nur ohne expliziten Identitätsbezug, sowie (4) alles ist erlaubt. Offlinetags nehmen den Emissionscharakter der Datenerzeugung hin und stellen ein Mittel bereit, das damit kompatibel ist. Eine Person kann ihre Präferenzen jederzeit ausdrücken, aber auch ändern. Dieses Statement begleitet jedes gemachte Foto, sofern es nicht mutwillig zerstört wird. Damit verbundene Regeln und Erwartungen lassen sich in (teil-)öffentlichen Räumen durch soziale Kontrolle durchsetzen, wenn Verstöße leicht erkennbar sind und wirksame Sanktionsmöglichkeiten bestehen.

5.5. Aufsicht über Daten-Ökosysteme

Soziale Kontrolle funktioniert nur unter günstigen Bedingungen. Andernfalls führt das Signalisieren der eigenen Präferenzen zu nichts, wie das Beispiel *Do Not Track*[34] demonstriert. Technisch funktioniert *Do Not Track* ähnlich den eben behandelten Offlinetags, aber es fehlt an einem sozialen Kontext, der die Nichtbeachtung bemerken und ahnden könnte. In solchen Situationen sind wirksame Aufsichtsmechanismen nötig. Die klassische Kombination aus Betroffenenrechten und Aufsichtsbehörden steht dabei vor dem Problem, dass Betroffene die Vielfalt der Datenempfänger und Nutzungen nicht mehr überblicken können. Auch in dieser Hinsicht scheint ein Anknüpfen an Auswirkungen statt an Daten als solche sinnvoll. Wo sich die Datenverarbeitung auf Personen auswirkt, muss dies in irgendeiner Form kenntlich werden; daran lassen sich dann zum Beispiel Auskunfts- und Eingriffsrechte knüpfen.

---

[33] Pallas u. a., »Offlinetags«, die Website http://www.offlinetags.net/ erklärt die verschiedenen Tags und bietet ein Programm zum Download an, das die Funktionsweise demonstriert.

[34] http://donottrack.us/

6. Fazit

Herkömmliche Regeln und Techniken des Datenschutzes sind an ihre Grenzen gestoßen. Sie stammen aus einer Welt isolierter Datenbanken mit überschaubarem Inhalt und naheliegender Dateninterpretation. Auf die heutige Welt vielfältig vernetzter Anwendungen, fortwährender Datenproduktion und lernender Maschinen angewandt, führen die alten Rezepte nicht zum Erfolg. Entweder passen sie überhaupt nicht, oder sie führen nur zur formalen Regeleinhaltung ohne den eigentlich beabsichtigten Schutzeffekt. Als einen Schritt auf dem Weg zu wirksameren Mitteln haben wir die Metapher der Datenemission eingeführt, die wesentliche Aspekte der heutigen Informationstechnologie repräsentiert. Vor diesem Hintergrund rücken wieder die Ziele hinter dem Datenschutz in den Fokus: die Rechte und die Autonomie des Einzelnen zu schützen. Dafür brauchen wir neue Werkzeuge.

Literatur

Acar, Gunes u. a. »The Web Never Forgets: Persistent Tracking Mechanisms in the Wild«. In: *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, Arizona, USA*. New York: ACM Press, 2014, S. 674–689. DOI: 10.1145/2660267.2660347.

Bager, Jo. »Rotlicht-Gerüchte: Bettina Wulff verklagt Google«. In: *Heise online* (8. Sep. 2012). URL: http://www.heise.de/newsticker/meldung/Rotlicht-Geruechte-Bettina-Wulff-verklagt-Google-1703200.html (besucht am 08. 09. 2016).

Barthélémy, Andrea und Andreas Wilkens. »Reden und lauschen: Interaktive Barbie kommt in US-Läden«. In: *Heise online* (9. Nov. 2015). URL: http://www.heise.de/newsticker/meldung/Reden-und-lauschen-Interaktive-Barbie-kommt-in-US-Laeden-2911626.html (besucht am 08. 09. 2016).

Breiman, Leo. »Statistical modeling: The two cultures (with comments and a rejoinder by the author)«. In: *Statistical Science* 16.3 (2001), S. 199–231. DOI: 10.1214/ss/1009213726.

Cebulla, Manuel. »Umgang mit Kollateraldaten«. In: *ZD – Zeitschrift für Datenschutz* 5.11 (Nov. 2015), S. 507–512.

Domingos, Pedro. »A few useful things to know about machine learning«. In: *Communication of the ACM* 55.10 (Okt. 2012), S. 78–87. DOI: 10.1145/2347736.2347755.

Eckersley, Peter. »How Unique Is Your Web Browser?« English. In: *Privacy Enhancing Technologies*. Hrsg. von Mikhail J. Atallah und Nicholas J. Hopper. Bd. 6205. Lecture Notes in Computer Science. Berlin und Heidelberg: Springer, 2010, S. 1–18. DOI: 10.1007/978-3-642-14527-8_1.

Hildebrandt, Mireille und Serge Gutwirth, Hrsg. *Profiling the European Citizen: Cross-Disciplinary Perspectives*. Dordrecht: Springer, 2008. DOI: 10.1007/978-1-4020-6914-7.

Kamp, Meike und Thilo Weichert. *Scoringsysteme zur Beurteilung der Kreditwürdigkeit - Chancen und Risiken für Verbraucher*. Gutachten im Auftrag des BMVEL. Kiel: Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein, Feb. 2006. URL: https://www.datenschutzzentrum.de/scoring/.

Lakoff, G. und M. Johnson. *Metaphors we live by*. Chicago: University of Chicago Press, 1996.

Pallas, Frank u. a. »Offlinetags: A Novel Privacy Approach to Online Photo Sharing«. In: *CHI'14 – Conference on Human Factors in Computing Systems Toronto, ON, Canada – April 26 - May 01, 2014, Extended Abstracts*. New York: ACM Press, 2014, S. 2179–2184. DOI: 10.1145/2559206.2581195.

Pedrotti, Frank L. u. a. *Optik für Ingenieure: Grundlagen*. 4. Aufl. Berlin und Heidelberg: Springer, 2007.

Pohle, Jörg. »PERSONAL DATA NOT FOUND: Personenbezogene Entscheidungen als überfällige Neuausrichtung im Datenschutz«. In: *Datenschutz Nachrichten* 39.1 (2016), S. 14–19.

Poller, Andreas u. a. »Electronic Identity Cards for User Authentication – Promise and Practice«. In: *IEEE Security and Privacy Magazine* 10.1 (Jan. 2012), S. 46–54. DOI: 10.1109/MSP.2011.148.

Roesner, Franziska, Tadayoshi Kohno und David Wetherall. »Detecting and Defending Against Third-Party Tracking on the Web«. In: *NSDI'12 Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*. Hrsg. von Steven Gribble und Dina Katabi. Berkeley: USENIX Association, Apr. 2012, S. 12. URL: https://www.usenix.org/conference/nsdi12/detecting-and-defending-against-third-party-tracking-web.

Rost, Martin und Kirsten Bock. »Privacy by Design und die neuen Schutzziele«. In: *DuD - Datenschutz und Datensicherheit* 35.1 (2011), S. 30–35. DOI: 10.1007/s11623-011-0009-y.

Schermer, Bart W. »The limits of privacy in automated profiling and data mining«. In: *Computer Law and Security Review* 27.1 (2011), S. 45–52. DOI: 10.1016/j.clsr.2010.11.009.

Steidle, Roland und Ulrich Pordesch. »Im Netz von Google – Web-Tracking und Datenschutz«. In: *DuD - Datenschutz und Datensicherheit* 32.5 (2008), S. 324–329. DOI: 10.1007/s11623-008-0078-8.

Steinmüller, Wilhelm u. a. *Grundfragen des Datenschutzes - Gutachten im Auftrag des Bundesinnenministeriums*. Bundestagsdrucksache VI/2826. 1971.

Sweeney, Latanya. »Discrimination in online ad delivery«. In: *Communication of the ACM* 56.5 (Mai 2013), S. 44–54. DOI: 10.1145/2447976.2447990.

Ware, Willis H. *Records, Computers and the Rights of Citizens: Report of the Secretary's Advisory Committee on Automated Personal Data Systems*. Santa Monica, 1973.

Warren, Samuel D. und Louis D. Brandeis. »The Right to Privacy«. In: *Harvard Law Review* 4.5 (1890), S. 193–220.