Unsere Spuren im Netz

Teil 1

Wer sich im Internet bewegt, hinterlässt vielfältige Spuren. Spuren, die für die Nutzer erst einmal nicht sichtbar sind. Für die Betreiber der Angebote sind diese aber auf vielfältige Art und Weise von Bedeutung. Folgend werden einige der Möglichkeiten dargestellt, solche Spuren sichtbar zu machen. Diese Datenspuren werden nicht etwa durch die Inhalte unserer Kommunikation erzeugt also, durch das, was wir in eine Mail schreiben oder was wir posten, was wir für ein Foto machen, sondern anhand der Metadaten. Deswegen kurz eine Unterscheidung der Daten. Im Datenschutz wird zwischen Daten mit Personenbezug und ohne Personenbezug unterschieden. Und im Technischen haben wir dann noch andere Unterscheidungen wie Protokolldaten, Bewegungsdaten, Steuerdaten. Inhaltsdaten, Metadaten und einige mehr. Wesentlich dazu beitragen, dass das nachvollzogen werden kann, was wir im Internet machen, sind die Metadaten. Folgend wird das mal an einer E-Mail deutlich gemacht. Eine E-Mail besteht unter anderem aus dem Header mit den Metadaten, dem Mitteilungstext und gegebenenfalls aus dem Anhang. So in etwa sieht ein Mailheader aus (Abb. 1).

Abb. 1 Mailheader

Darin sind also der Absender, das Datum, wo die Mail versandt worden ist, das verwendete Mailprogramm, der Mail Provider über welchen der Service läuft, welches Betriebssystem man verwendet hat, die IP-Adresse, die verwendeten Schrifttypen, der Empfänger usw. Im Gegensatz zu den Informationen im Header können die Inhaltsinnformationen, also der Mailtext, und der Anhang aber verschlüsselt werden.
Es gibt nur wenige Möglichkeiten, unsere Spuren im Netz wirklich sichtbar zu machen, also zu visualisieren. Eine der Möglichkeiten ist das Programme Lightbeam. Ein sogenanntes Plugin für den Browser Firefox (bis Vers. 81.0.2) Wenn der installiert ist, kann sichtbar gemacht werden, wer einem beim Surfen im Internet zusieht. Hier einmal exemplarisch dargestellt (Abb. 2). Wird beim erstmaligen Aufrufen einer bekannten Suchmaschine dem „Setzen“ aller Cookies zugestimmt, und eine Suche gestartet wird, werden von den Betreibern der aufgerufenen Seite viele kleine Programme gestartet, die einem nun beim Surfen zusehen. Zu sehen ist in der Mitte ein rundes Symbol was hier diese Suchmaschine darstellt und davon sternförmig ausgehend, viele kleine Symbole.
Wenn man auf eines der Symbole klickt, werden auf der rechten Seite dazu weitere Details angezeigt. Also zum Beispiel, wo der Standort des Server von diesem kleinen Programm ist. Im unteren Teil des Hauptfeldes befinden sich dann einige Button. Damit können weitere Informationen über die einzelnen Tracker angezeigt werden. Auch gibt es weitere Möglichkeiten sich die Trackingrgramme etwas näher anzusehen. Wird beispielsweise „Liste“ gewählt, werde die Tracker als Liste dargestellt. Mit dem Markieren eines Trackers in der Liste, lässt sich dieser z.B. mit einer Funktion im unteren Feld dann entfernen.
Eine andere Darstellungsform ist die der Glocke bzw. des Halbkreises (Abb. 3). Am linken unteren Rand beginnt die Zählung mit 12 bzw. 1; also Mitternacht. Oben am höchsten Punkt ist dann mittags und am rechten unteren Rand dann die nächste Mitternacht. Hier ist ganz deutlich sehen, dass zwischen 10:00 Uhr und 13:00 Uhr und dann noch mal abends um 20:00 Uhr am meisten Aktivitäten stattgefunden haben. Auch hier kann man wieder auf eines die-

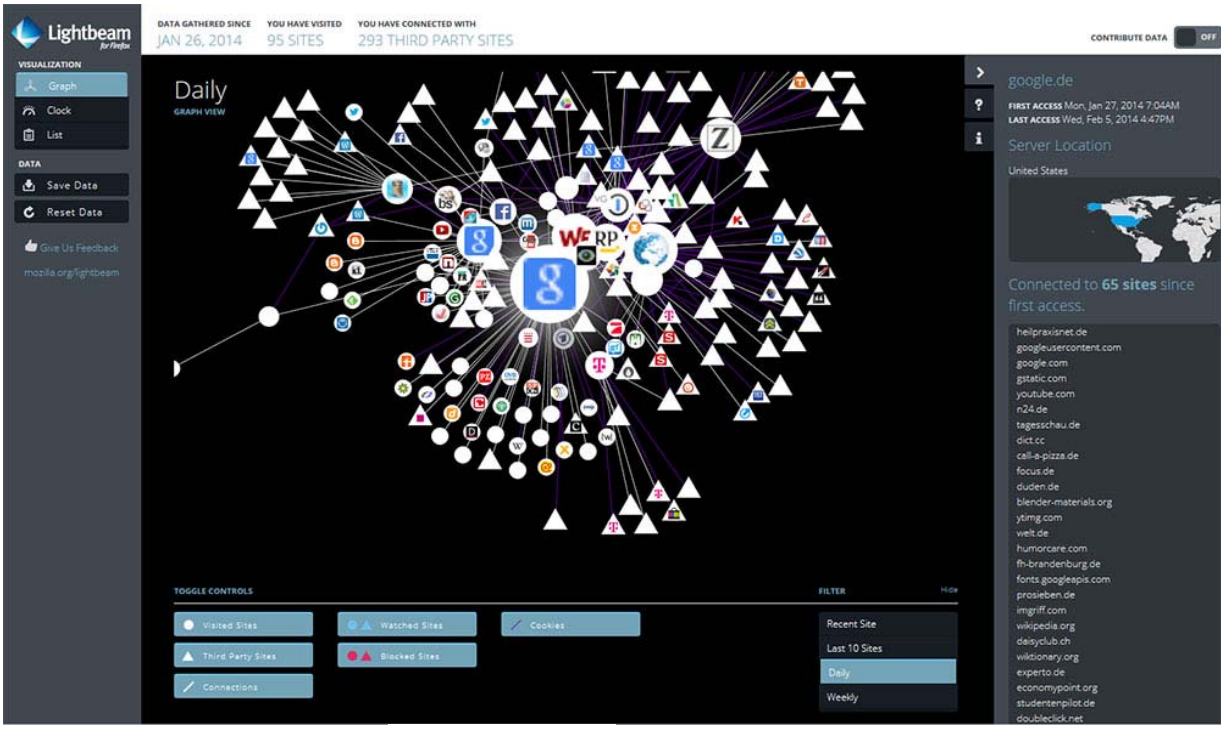


Abb. 2: Screenshot von Lightbeam (Kawalkowski 2021)

ser kleinen Symbole klicken und bekommt dann den Namen des Trackers angezeigt, also wer einen um diese Zeit dort verfolgt.
Nutzt man weitere Plugins, wie z.B. Ghostery, Privacy Badger, Blur, Web of Trust, uBlock etc. kann das Installieren vieler Tracker verhindert werden, und die „Zuschauer“ beim Surfen werden sehr reduziert auf die, die entweder von der Seite selber eingesetzt werden beziehungsweise notwendig sind, um die Funktionalität der Darstellung zu gewährleisten. Denn der Anbieter muss ja wissen, welche Auflösung mein Bildschirm hat, was für ein Betriebssystem ich habe etc. pp. Also diese Informationen sind zum Beispiel notwendig, um eine entsprechende Darstellung am jeweiligen Bildschirm zu ermöglichen.

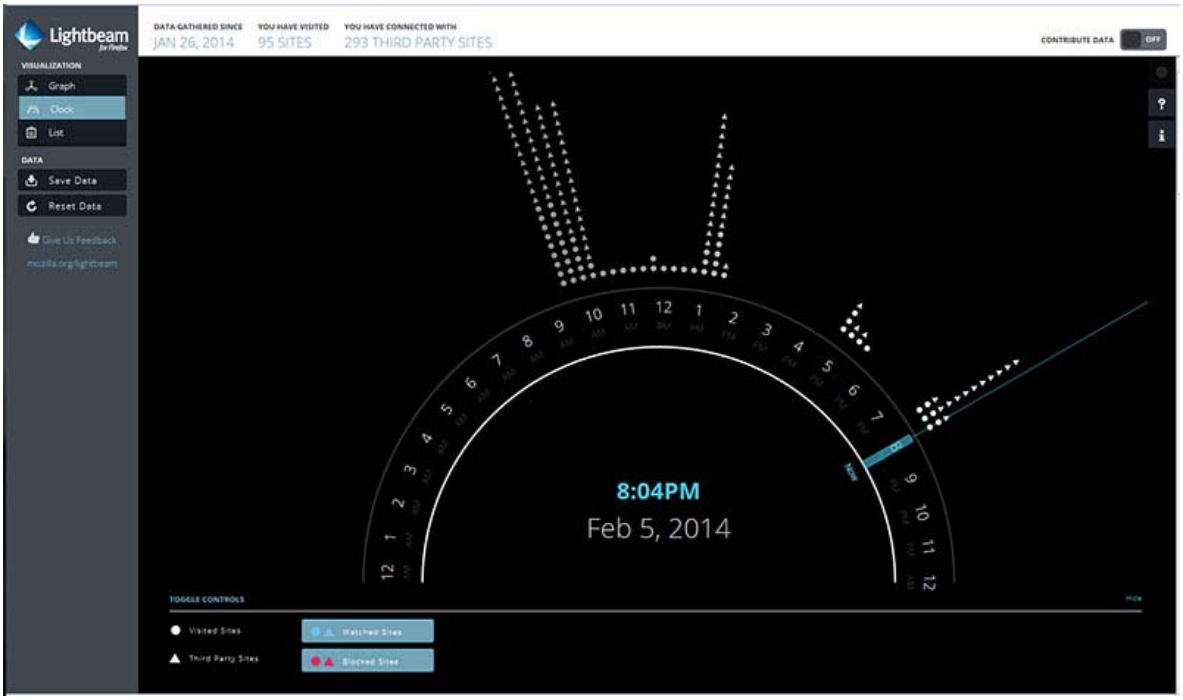
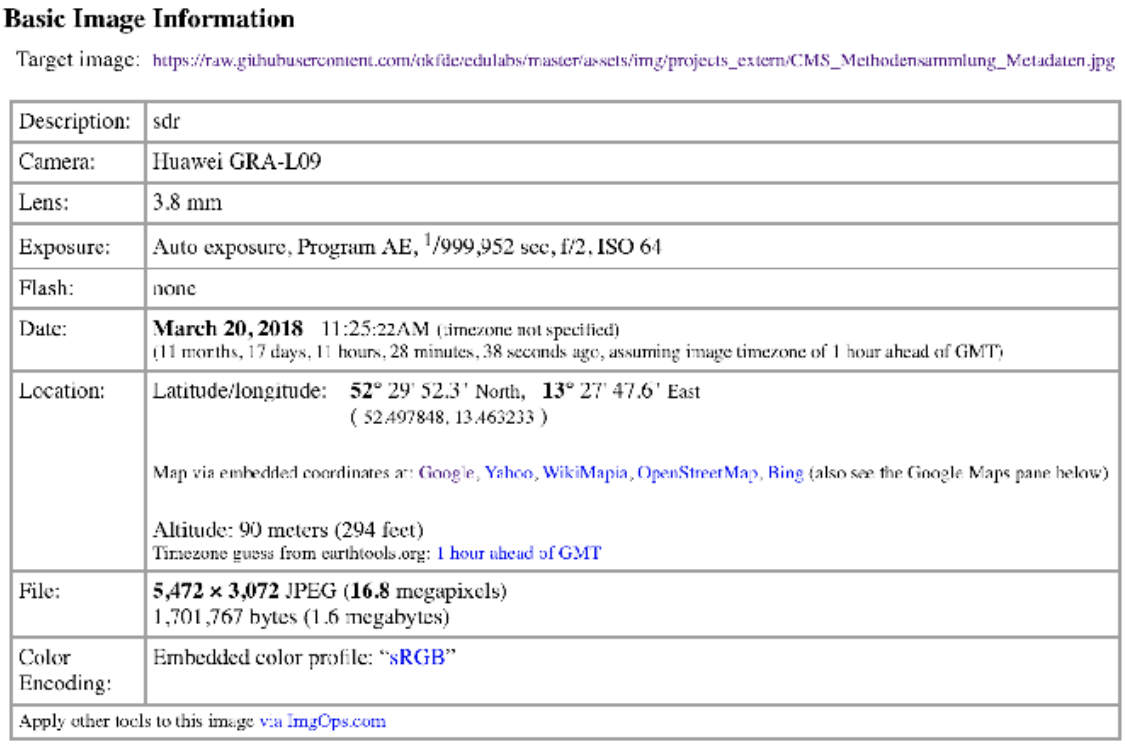


Abb.3 Screenshot von Lightbeam, Glockendarstellung (Kawalkowski 2021)

Es gibt aber auch weitere Möglichkeiten Spuren im Netz sichtbar zu machen. Zum Beispiel durch die Metadaten, die sich an den meisten digital gemachten Foto befinden, die von dem Gerät erzeugt werden mit dem das Foto gemacht wird und aus denen man dann viele Informationen erlangen kann, die man dem Bild selber gar nicht ansieht. Auf so einem Bild ist beispielsweise ein Schlafzimmer; vielleicht in einem Dachgeschoss. Und wenn man sich dann die Metadaten anschaut, (Abb.4) kann man da verschiedene Dinge daraus



Basic Image Information

Target image: https://raw.githubusercontent.com/okfde/edulabs/master/assets/img/projects_extern/CMS_Methodensammlung_Metadaten.jpg

Description:	sdr
Camera:	Huawei GRA-L09
Lens:	3.8 mm
Exposure:	Auto exposure, Program AE, 1/999,952 sec, f/2, ISO 64
Flash:	none
Date:	**March 20, 2018** 11:25:22AM (timezone not specified) (11 months, 17 days, 11 hours, 28 minutes, 38 seconds ago, assuming image timezone of 1 hour ahead of GMT)
Location:	Latitude/longitude: **52°** 29' 52.3" North, **13°** 27' 47.6" East (52.497848, 13.463233) Map via embedded coordinates at: Google, Yahoo, WikiMapia, OpenStreetMap, Bing (also see the Google Maps pane below) Altitude: 90 meters (294 feet) Timezone guess from earthtools.org: 1 hour ahead of GMT
File:	**5,472 × 3,072** JPEG (**16.8** megapixels) 1,701,767 bytes (1.6 megabytes)
Color Encoding:	Embedded color profile: "sRGB"
Apply other tools to this image via ImgOps.com	

Abb.4: Metadaten einer Digitalfotografie (Chaos macht Schule, Berlin)

ersehen. Also beispielsweise, mit welcher Kamera mit welcher Linse die Aufnahme gemacht wurde, an welchem Datum, und auch wo die Location ist. Die geographische Lage ist hier genau festgelegt und wenn man diese Daten dann wiederum zum Beispiel in ein Kartenprogramm eingibt, dann wird dieser Ort genau angezeigt und man kann sehen, wo diese Aufnahme erstellt wurde. Man weiß auch, wann das gewesen ist. Die Metadaten einer Fotografie verraten also recht viel über das Motiv hinaus. Es sei, denn man entfernt die Metadaten aus der Datei des Fotos.
Welche Spuren eine Person im Internet alles hinterlässt, kann auch gut an der US-Plattform MyLife.com (Abb.5) gezeigt werden.
Diese Plattform existiert seit 2002 und sammelt Informationen aus allen zugänglichen Quellen und verknüpft diese zu öffentlich einsehbaren Profilen. Inzwischen soll es davon mehr als 325 Millionen Profile geben. Dies ist also gut die US-amerikanische Bevölkerung. Dort sind Informationen über den richtigen Name, das Alter der Person, frühere und aktuelle Wohnadressen, Telefonnummern, E-Mail, Arbeitgeber, Ausbildung, Fotos. Verwandte, politische Zugehörigkeit und auch beispielsweise Informationen über sexuelle Orientierung, Straftaten, Hautfarbe, gefahrene Autos, und weitere Dinge. Diese Seiten sind für jeden US-Bürger in der Regel frei einsehbar. Eine Korrektur ist eigentlich nur dann möglich, wenn beispielsweise in den Bezahl-Modus gegangen wird und man nachweisen kann, dass eine Information über einen falsch ist, man also zum Beispiel gar keinen Kleinwagen fährt, sondern einen großen Wagen, oder dass man diese oder jene Straftat gar nicht begangen hat, was mit einem amtlichen Registerauszug geschehen kann. Oder dass man viel mehr oder viel weniger verdient, indem man den Konto-

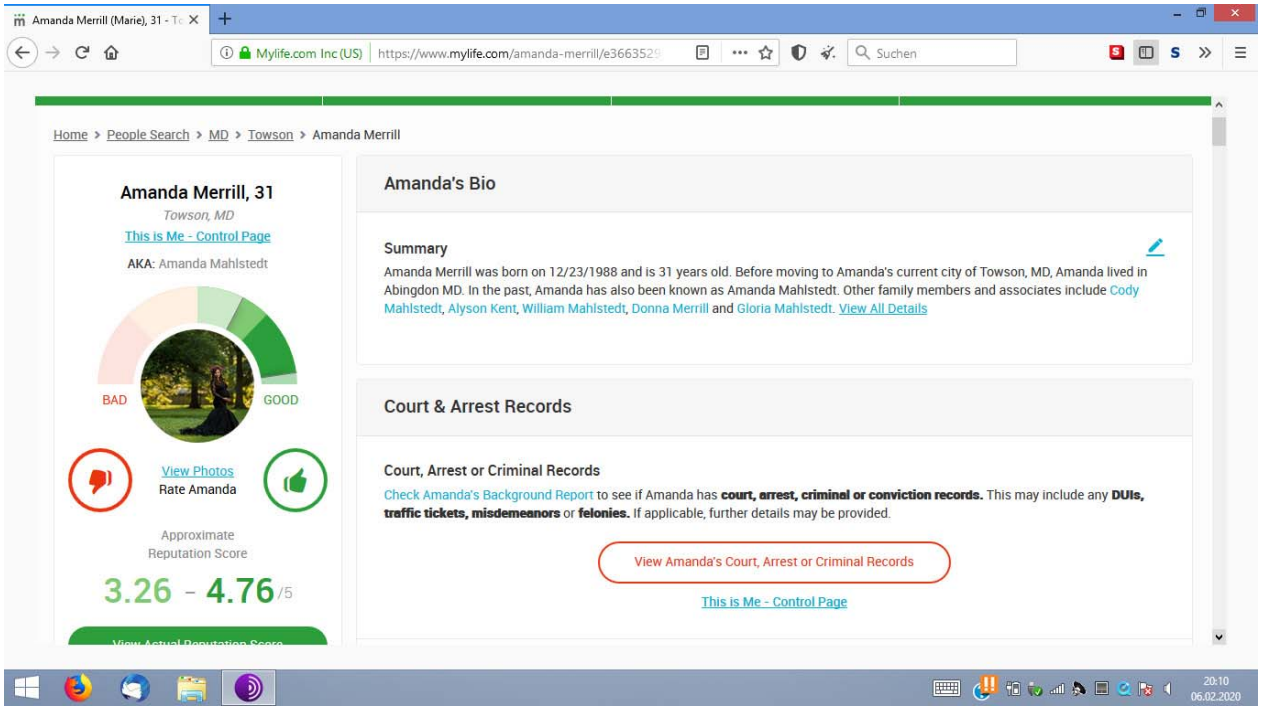


Abb.5: Screenshot einer Seite von mylife.com

auszug beibringt. Die Seiten sind nur von einem Internetanschluss aus den USA erreichbar. Man kann aber so tun, als wäre man in den USA, wie zum Beispiel mit dem Tor-Browser, mit dem man dann auf diese Seiten Zugang erhält. Derzeit ist, so ein Portal in Europa noch nicht. präsent. Es wird aber daran gearbeitet, ein solches Portal auch für Europa zu installieren.

Teil 2

Im zweiten Teil wird Googel als Beispiel dafür genommen, wie Daten beschafft werde und wofür sie verwendet werden. Sowohl Google (Alphabet[1]) als auch alle anderen großen Internetkonzerne, wie Facebook (Meta), Amazon, Alibaba legen eigentlich von jeder Person, von der sie auch nur einen Datenschnipsel bekommen, ein Profil an. Facebook macht das übrigens auch von Firmen. Und solange diese Personen selber sich dort nicht anmelden, wird von diesen aber ein Schattenprofil angelegt und gepflegt. Zu Google gehört ja nicht nur Google, also die Suchmaschine, sondern eine große Anzahl von Programmen und Anwendungen. Die meisten wissen, dass zum Beispiel Google Maps natürlich dazu gehört. Ebenso Googletranslation, YouTube, das Betriebsystem Android, Google Chat, Google Kontakte, Google Kalender, Google Docks, Google Präsentation, Google Drive und so weiter. Also eine große Anzahl von Anwendungen, die eigentlich alle unter dem großen Label Alphabet firmieren, weit verzweigt ist und alle untereinander unsere Daten munter austauschen. Von einigen davon sind hier die Icon abgebildet (Abb. 6).
Und selbstverständlich nutzen die meisten Menschen, die im Internet unterwegs sind, einige dieser Dienste. Und dann gibt es noch Google Analytics. Das ist ein Tool, was von Google für Webentwickler kostenlos zur Verfügung gestellt wird. Inzwischen wird es von 56% aller Websites verwendet. Damit wird jeder einzelne Klick auf einer Website aufgezeichnet und dem Seitenbetreiber mitgeteilt. Gleichzeitig erhält Google auch diese Information. Und da werden Informationen im Hintergrund dann beispielsweise benutzt, um zu schauen nach Geschlecht, Altersklasse, Kaufkraft, Konsumpräferenzen, möglicherweise weitere psychometrischer Kategorien, falls die über den

[1] https://abc.xyz/

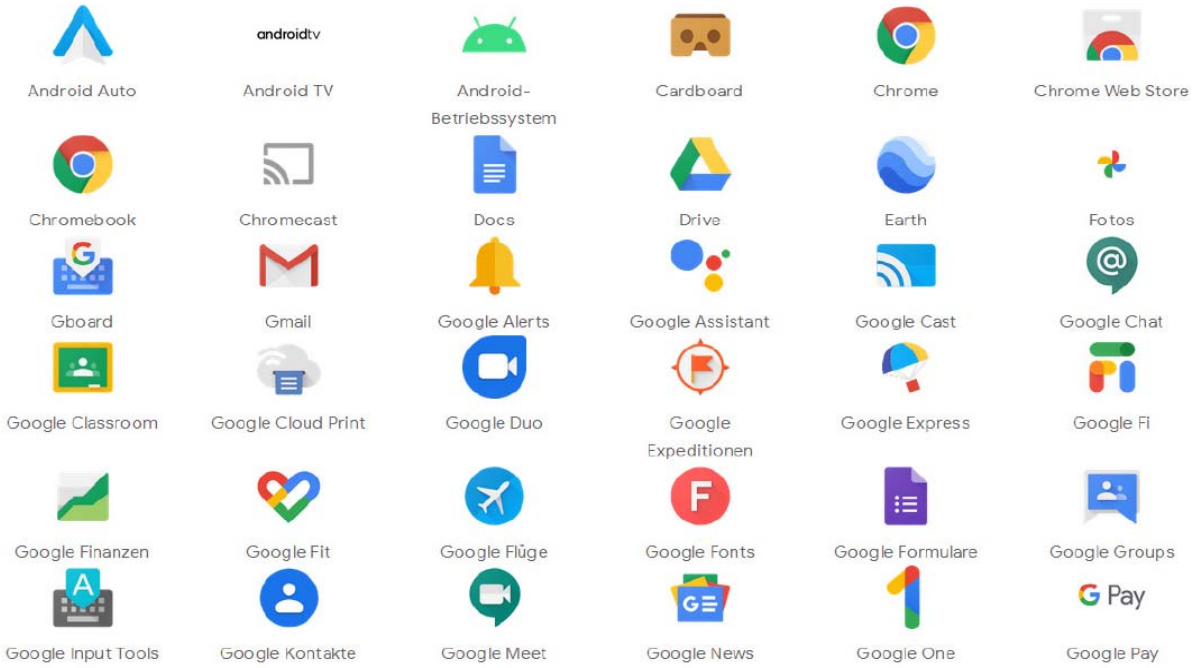


Abb.6 Auswahl einiger Google-Apps

Nutzer zur Verfügung stehen. Auch die Verkettung von Klicks also auf welche Seite man anschließend kommt, von welcher Seite man auf die eigene Seite kommt. Das wird alles an den Webseitenbetreiber (und Google!) mitgeteilt. Die Verläufe und Nutzungen können über Tage, Wochen, Jahre verfolgt werden. Was geschieht mit diesen Informationen? Der Seitenbetreiber kann selber bestimmen, was er mit den Daten macht. Und Google?

Ein weiteres Element dabei ist, dass jeder, der eine App entwickelt und diese App über den Google-App-Store vertreiben möchte, verpflichtet ist die Google-Werbe-ID zu verwenden. Also jeder, der so eine App nutzt, installiert oder nutzt auch die Google-Werbe-ID beisich installiert bekommt. Damit ist dann jedes Endgerät eindeutig identifiziert. Und demzufolge die Verknüpfung. Zwar kann man die Werbe-ID löschen, doch bei der Nutzung irgend einer dieser App wird wieder eine installiert und das Nutzerverhalten mit bisherigen IDs und deren Profil abgeglichen. Und immer wenn man das erste Mal auf einer Seite ist, wird man aufgefordert, dem Setzen von Cookies zuzustimmen. Meist erhält man eine bedingte Auswahl; manchmal soll man anstatt den Cookies zuzustimmen auch in den Bezahl-Modus (Paywall) wechseln. Hier wird deutlich, dass die Daten einen Geldwert besitzen. Stimmt man dem Setzen aller Cookies zu, weil man schnell weiter recherchieren möchte, lässt man ganz viele dieser Programme auf seinen Rechner bzw. seinem mobilen Endgerät zu. Es sei denn, man stimmt nur den absolut notwendigen Cookies zu.
Nun ein genauerer Blick auf diese Kekse. Gucken wir uns mal an, wie das mit diesen Cookies aussieht. Das heißt also, man kann bei diesen meisten der Aufforderungen Cookies zuzulassen, sich die Details zeigen lassen, wer hier alles zugucken soll und bekommt so eine Übersichtsliste, die im Einzelfall auch schon über 80 Nennungen enthält. Da ist dann der Dateiname, der Host, Laufzeit, Cookietyp, Kategorie und Kurzbeschreibung (Abb. 7). Es gibt, wie gesagt, eine ganze Menge unterschiedliche Arten von Cookies. Hier sollen einige kurz genannt werden: 3rd-Party-Cookies (Cookies, die zu setzen, Dritten erlaubt wird), HTTP-Cookies (das „gewöhnliche“ und meist unbedenkliche Cookie), Supercookie (Diese ermöglichen es zusammengetragene Informationen auf den Server zu übertragen. Sie bleiben auch beim Schließen des Browsers auf dem Rechner und sind nicht ohne weiteres zu finden und zu löschen), Flash-Cookies (Diese nutzen aus, dass der installierte

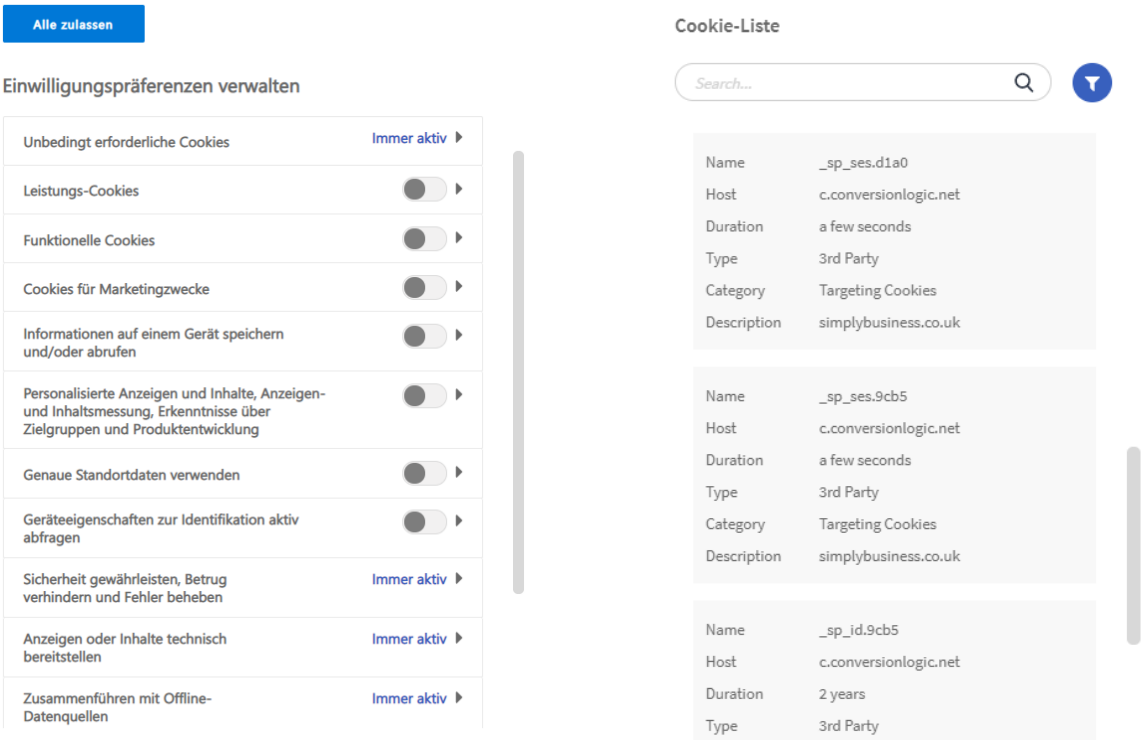
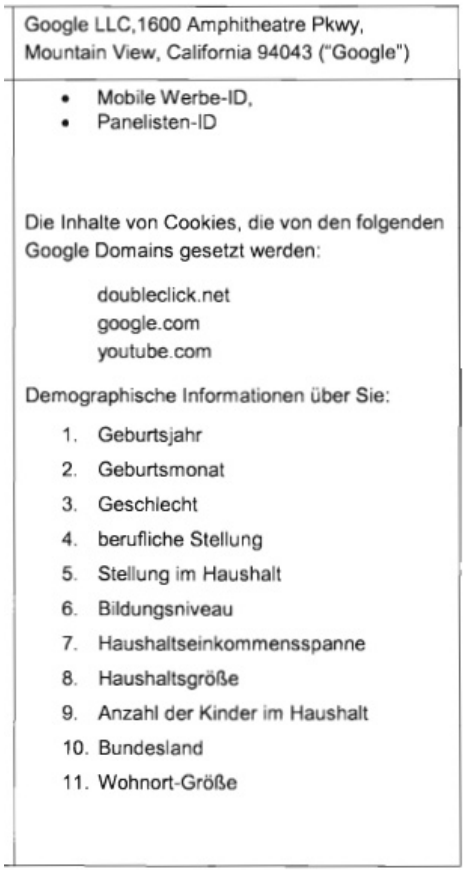


Abb. 7 Cookie-Einwilligungsfenster beim Aufruf einer neuen Internetseite

Flash-Player ein Speichern großer Mengen an Informationen erlaubt). Es gibt auch Cookies, die Informationen aus anderen Quellen zusammen tragen, oder sich nach anderen Geräten in der Nähe „umsehen" mit denen sie sich austauschen. Da wird dann halt geschaut, ob diese Geräte demselben Benutzer oder nur zum selben Haushalt gehören. Oder die Geräteeigenschaften werden abgefragt und abgeglichen. Erlaubt ist dies alles, weil der Nutzer dem ja mit der Einwilligung zugestimmt hat. Das heißt also, mit solchen Kennungen kann ein Gerät. Identifiziert werden auch, wenn die Werbe-ID zum Beispiel gelöscht wird. Und wenn man sich irgendwo mit einem persönlichen Account anmeldet, also bei Facebook, bei Amazon oder WhatsApp, Youtube etc., dann ist man als Person identifizierbar, weil oft eine Klarnamenpflicht besteht. Und beim Aufrufen einer neuen Seite werden diese Infos dann gleich weitergegeben (es sei denn, man hat die Cookies gelöscht). So hat Google ein Instrument geschaffen, was Webdesigner dazu bringt ein „Social Hacking" zu betreiben So wird auch ohne ein Eindringen in fremde Systeme auf diesen Geräten eine Software zum Einsatz gebracht, welche der Nutzer ausdrücklich duldet. So können deren Daten umfassend abgeschöpft werden. Und diese Nutzerprofile sind das eigentliche Geschäftsmodell von Google (Alphabet).
Wo sind nun diese ominösen Cookies zu finden? Unter „Einstellungen" geht man auf „Cookies und Websiteberechtigungen". Dann auf „Alle Cookies und Websitedaten anzeigen". Es öffnet sich ein Fenster, was alle „Cookies und gespeicherte Daten für Websites, die Sie besucht haben" zeigt. Geht man dann auf eine der angezeigten Dateien (Cookies), welche so Namen wie „_ga_Y8798Z0549" haben. Mittels der Suchfunktion kann man die so gefundene Datei auf dem Rechner lokalisieren und sich mit einem Editor anzeigen.

Was macht Google nun mit solchen Information? Natürlich Werbung, also vor allem personalisierte Werbung. Das ist uns allen klar, das heißt, wir bekommen aufgrund von unserem Profil, was aufgrund der vielfältigen Informationen, die so gesammelt werden, die Werbung eingeblendet, die auf uns passen soll. Dafür ist unser Profil zuvor meistbietend versteigert worden.

Je mehr Informationen dabei über uns verfügbar sind, desto teurer wird unser Profil dann verkauft.
Eine weitere Funktion ist die der Preisdiskriminierung. Will man beispielsweise verreisen und bucht erst einen Flug und dann die Unterkunft, werden einem für den Flug niedrige Preise und dann höhere Preise für die Unterkunft angezeigt. Wählt man den umgekehrten Weg, ist die Preisgestaltung umgekehrt. Denn durch das Tracking „weiß" das Reiseportal, dass z.B. schon ein Flug gebucht ist, und man nun auf der Suche nach einer (günstigen) Unterkunft ist. Hinzu kommt noch, dass das Portal möglicherweise aus meinem Profil weiß, wie viel Geld ich für Reise auszugeben bereit bin. Und wenn dann anschließend noch ein Mietwagen und ein Surfkurs gebucht werden, soll geht es in dieser Weise weiter.
Doch auch, wenn man beispielsweise eine Wohnung sucht, oder sich beruflich verändern möchte, werden unsere Datenspuren immer häufiger für eine solche Entscheidung der Wohnungsinhaber bzw. potenzieller Arbeitgeben herangezogen. Dies geschieht mit Hilfe eines aus unseren Datenspuren errechneten Scorewertes.
Eine Anfrage, was Google über mich weiß und woher dieses Wissen stammt, ergab recht Erstaunliches (Abb.8). Die Infos stammten aus der Auswertung meiner Nutzungen der Suchmaschine Google, doublecklick.net und YouTube. Benutzt wurden dafür die mir zugewiesenen Mobile Werbe-ID und die Panelisten-ID. Und dadurch hat Google von mit erfahren, welches Geburtsjahr ich habe, den Geburtsmonat, Geschlecht, berufliche Stellung, Stellung im Haushalt, Bildungsniveau, Haushaltseinkommensspanne, Haushaltsgröße, Anzahl der Kinder im Haushalt, Bundesland und Wohnortgröße. Je nachdem, welche Wertigkeit diesen Einzelinformationen zugewiesen werden, wird daraus ein Scorewert errechnet. Dies machen dann solche Firmen wie SCHUFA, GfK, GroupM, Nugg.ad GmbH, Annalect Group Germany GmbH, Emetriq GmbH, candidate select GmbH, Creditreform, Forteil GmbH, Z-Score Deutschland GmbH oder infoscore consumer Data GmbH. Anfragen dort ergaben verschiedene Scorewerte (Abb. 9), doch was keine der Firmen heraus rückte, war der Algorithmus, nach der der Scorewert berechnet wird.


Google LLC,1600 Amphitheatre Pkwy,
Mountain View, California 94043 ("Google")

- Mobile Werbe-ID,
- Panelisten-ID

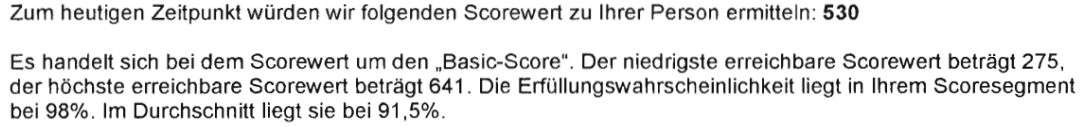
Die Inhalte von Cookies, die von den folgenden Google Domains gesetzt werden:

doubleclick.net
google.com
youtube.com

Demographische Informationen über Sie:

1. Geburtsjahr
2. Geburtsmonat
3. Geschlecht
4. berufliche Stellung
5. Stellung im Haushalt
6. Bildungsniveau
7. Haushaltseinkommensspanne
8. Haushaltsgröße
9. Anzahl der Kinder im Haushalt
10. Bundesland
11. Wohnort-Größe

Abb. 8. Googel-Auskunft (Ausschnitt)


Zum heutigen Zeitpunkt würden wir folgenden Scorewert zu Ihrer Person ermitteln: **530**

Es handelt sich bei dem Scorewert um den „Basic-Score". Der niedrigste erreichbare Scorewert beträgt 275, der höchste erreichbare Scorewert beträgt 641. Die Erfüllungswahrscheinlichkeit liegt in Ihrem Scoresegment bei 98%. Im Durchschnitt liegt sie bei 91,5%.

Abb. 9: Ausschnitt der Antwort einer Auskunftei zum eigenen Scorewert

Als nächstes informierte ich mich, was dies für den Abschluss einer privaten Krankenzusatzversicherung bedeutet. Auf Nachfrage wurde mir mitgeteilt, dass sie für meine Einstufung meine Scorewerte der SCHUFA und der Infoscore Consumer Data GmbH heranziehen.

Teil 3

Auch beim meistverbreiteten Messenger, WhatsApp werden Metadaten vom Anbieter ausgelesen und genutzt. Zum Beispiel Standortbestimmung über GPS

und WLAN, Telefonanrufe, Gerätenummer, Betriebssystem, Batteriestand, Signalstärke, App-Versionen, verwendeter Browser, genutztes Mobilfunknetz, Telefonnummer, Internetanbieter, verwendete Sprache, Zeitzone, IP Adresse, Ort und Zeitpunkt, an dem eine Nachricht geschickt wurde. All diese Informationen werden erhoben und sind von WhatsApp jederzeit vom Endgerät abrufbar. Ohne dass WhatsApp in die Inhalte reinschauen muss, weiß der Dienstanbieter also (fast) alles über den Tagesablauf der Nutzer. Da hilft es dann auch wenig, dass die eigentlichen Nachrichten sicher verschlüsselt übertragen werden.
Wichtigster Aspekt sind aber die eigenen Kontakte. Nach der Registrierung liest WhatsApp automatisch alle Kontakte des Telefons und synchronisiert diese mit der Datenbank des Anbieters. (Wenn dies unterbunden wird, werden nur die Telefonnummern der Kontakte angezeigt.) Diese werden zur Synchronisation auf die Server von WhatsApp übertragen und dort gespeichert. Die Metadaten werden an Dritte (Partner) weiter gegeben. Dazu steht in den Geschäftsbedingungen: „wir können deine Informationen an jedwedes unserer verbundenen Unternehmen oder Nachfolgeunternehmen bzw. jeden neuen Eigentümer übertragen.“[2]
Die Verknüpfung der so gewonnenen Kontaktdaten ermöglicht es WhatsApp ein genaues Bild unseres Netzwerkes, meist genauer als wir es kennen, zu erstellen und das als Teil unseres Profils für sich zu nutzen.
Die Nachrichten via WhatsApp selber werden mittels SP Ende-zu-Ende-verschlüsselt. Dies geschieht mit Hilfe des Curve-25519-Schlüssels. Die Verbindung wird mit dem Elliptic Curve Diffie-Hellman-Protokolls (ECDH) realisiert. Die Nachrichten unter Nutzung der Schlüsselableitfunktion HKDF generiert. Jede Nachricht wird mit einem eigenen sog. MessageKey mittels AES-256-Verfahren im CBC-Modus verschlüsselt. Die Signierung erfolgt mit dem HMAC-SHA-Schlüssel. Für jede Nachricht wird ein öffentlicher Curve25519-Schlüssel bereitgestellt[3]. Damit ist die Übertragung der Inhalte per WhastApp sehr sicher.
Da die Kommunikation aber auf den Servern von WhatsApp gespeichert wird, und WhatsApp ein US-Unternehmen ist, gilt für diese Daten US-Recht. Abschnitt 702 des US-Geheimdienstgesetzes Foreign Intelligence Surveillance Gesetzes (FISA) erlaubt Diensten wie der NSA | CIA | FBI | NRO | NGA | OOI | INR | TFI | I&A[4] ohne konkreten Verdacht und ohne entsprechenden Gerichtsbeschluss, weit reichende Zugriffe auf Kommunikationsdaten von Ausländern („non-US persons“), die von US-amerikanischen Unternehmen erhoben werden, auszuwerten – egal, wo auf der Welt die Server physisch stehen. Dies widerspricht aber europäischem Datenschutzrecht.

Teil 4

Das, was im Folgenden vorgestellt wird, geht aus einer Arbeit der Gruppe um Clemens Stachl und Quay Au hervor.[5]
Die Gruppe hat sich angeschaut. Welche Informationen und Rückschlüsse man aus Smartphonedaten herauslesen kann, ohne die Inhalte der Kommunikation dafür heranzuziehen; zu kennen. Also das, was WhatsApp macht, das was Goooogle macht.
Anhand von Verhaltensinformationen, die aus Smartphone-Daten entnommen werden, wird geprüft, ob mittels dieser Informationen Persönlichkeitspro-

2 https://www.whatsapp.com/legal/updates/terms-of-service/?lang=de
3 Ullrich, Elias (2018): Analyse des Messengerdienstes WhatsApp aus forensischer Sicht. Hochschule Mittweida, Fakultät Angewandte Computer- und Biowissenschaften. Bachelorarbeit.
https://monami.hs-mittweida.de/frontdoor/deliver/index/docId/10799/file/BA_Ullrich.pdf
4 Das sind alles US-Geheimdienste
5 Stachl, Clemens. et al. (2020): Predicting personality from patterns of behavior collected with smartphones. Proceedings of the National Academy of Science of the United States of America. 17680–17687 | PNAS | July 28, 2020 | vol. 117 | no. 30
Im Internet unter: https://www.pnas.org/content/pnas/117/30/17680.full.pdf

file erstellt werden können. Als Datenmaterial werden Sensor- und Protokolldaten (keine Inhaltsdaten) verwendet. Dies sind: Durchschnittliche Dauer von App-Nutzungen, Art und Umfang von Musikkonsum, das Kommunikationsverhalten anhand der Anzahl ausgehender Telefonate pro Tag, der mittlere Bewegungsradius anhand der GPS-Daten, allgemeine Nutzung des Mobilgerätes anhand der Anzahl täglicher Entsperrungen des Gerätes und das Verhältnis von Nacht- zu Tagaktivitäten. Parallel wurden die Persönlichkeitsmerkmale der Probanden mittels einer standardisierten Befragung erfasst. Dazu dient die deutsche Version des BFSI (60) mit 300 Items[6]. Zusätzlich wurden Alter, Geschlecht, höchster Bildungsabschluss etc. erhoben.

- *Emotionale Stabilität:* Maß für die persönliche Tendenz im Umgang mit eigenen Emotionen und potentiellen Belastungen (bspw. Selbstbeherrschung, positive Grundstimmung).
- *Extraversion:* Maß für die individuelle Tendenz im zwischenmenschlichen Verhalten (bspw. Durchsetzungsfähigkeit, Fröhlichkeit).
- *Offenheit:* Maß für Tendenzen im Umgang mit neuen Erlebnissen, Eindrücken, Ideen und Werthaltungen (bspw. Offenheit für Ästhetik oder Ideen).
- *Gewissenhaftigkeit:* Maß für persönliche Tendenzen im Arbeits- und Leistungsverhalten (bspw. Kompetenz, Disziplin).
- *Verträglichkeit:* Maß für Tendenzen in der Art des zwischenmenschlichen Umgangs (bspw. Vertrauensbereitschaft, Bescheidenheit).[7]

Mit den Smartphone-Daten wurden dann nichtlineare Regressionsmodelle (elastische Netze), anhand eines Out-of-Sample-Modells mit Hilfe eines (verschachtelten) kreuzvalidierten Ansatzes trainiert.
Es wird geschaut, welche Aktivitäten mit welchen Ergebnissen der psychologischen Persönlichkeitsanalyse korrelieren. In der Analyse wurden mittels der Kombination sämtlicher Aktivitäten mit sämtlichen Prädikatoren der Persönlichkeitsdimensionen (20) in Beziehung gesetzt.
Dazu standen 15.659 Variablen für das Training des Systems zur Verfügung.
Im Einzelnen wurde

- für Offenheit für Ästhetik die höchste,
- für Ordnungsliebe eine sehr hohe,
- für Ehrgeiz eine hohe,
- für Disziplin eine hohe,
- für Unternehmungslust eine gute,
- für Offenheit für Gefühle eine gute,
- für Selbstvertrauen eine gute,
- für Durchsetzungsvermögen eine gute,
- für Umsichtigkeit eine mäßig hohe,
- für Sorglosigkeit eine gute mittlere,
- für Offenheit für Gefühle eine mittlere,
- für Offenheit für Ideen eine mittlere,
- für Offenheit für Handlungen die niedrigste

Vorhersagegenauigkeit errechnet.

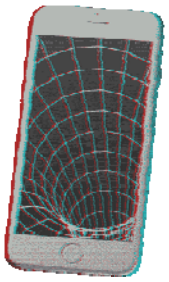


Abb. 10 Symbolbild Smartphone (CCC)

Es wurde also herausgefunden, dass es möglich ist die verschiedenen psychologischen Merkmalsdimensionen anhand einer Smartphonenutzung unterschiedlich gut zu bestimmen. (Das eingesetzte Random-Forest-Modell erweist sich dabei als besonders genau.) Mit Ausnahme der Dimension „Offenheit für Handlungen“ können aus den Daten demzufolge alle Persönlichkeitsaspekte gut bestimmt werden.
Das heißt, mittels der Metadaten, die von den Anbietern vieler Dienste aus dem Smartphone ausgelesen werden können, ist eine ziemlich gute Persönlichkeitsanalyse (analog der, wie sie von psychologischen Tests durchgeführt werden) möglich. So ist es völlig unerheblich, ob die Inhalte, welche wir versenden für Dritte nicht einsehbar sind, da mit den Informationen aus unserem Nutzungsverhalten unsere Persönlichkeit offen liegt.

6 https://psyexpert.de/wp-content/uploads/2019/04/BFSI.pdf
7 https://psyexpert.de/wp-content/uploads/2019/04/BFSI.pdf

Teil 5

Die Informationen für Teil 5 gehen primär aus den Veröffentlichungen von Mühlhoff[8] und Wachter[9] zurück. Hier geht es um Prädictive Privatheit. Diese Technik basiert auf KI-Anwendung großer Datenmengen; also Big-Data-Analytic in einem System-Learning-Modell. Diese Anwendungen entziehen sich der Sichtbarkeit. Das heißt, hier ist das auch nur irgendwie sichtbarmachen von Spuren (derzeit noch) nicht möglich.
Dabei werden eine große Menge unstrukturierter Hilfsdaten verwendet, also letztendlich solche wie zuvor beschrieben. Also diese klassischen Informationen wie die genannt wurden. Art des Browsers und Standortverläufe, App-Nutzung, Musikkonsum, Kommunikations- und Sozialverhalten, Mobilität, Tagesaktivitäten am Gerät sowie Likes, Posting und Kontakte. Also Kontakte im Social-Media-Bereich, wobei auch da vor allen Dingen die Frequenzen und die Anzahl etc. eine Rolle spielen und nicht die Inhalte. Interessieren tun sich die Anwender der Prädiktive Analyse, wie etwa unsere Bonität, also Zahlungsfähigkeit aussieht, welche Krankheiten wir haben, wir möglicherweise eine Suchterkrankung haben, wie wir politisch eingestellt sind, wie wir unser Geschlecht definieren sowie psychologische und emotionale Disposition. Also alles Informationen welche die meisten selten von sich preisgeben und die auch dem besonderen Datenschutz für eine mögliche Verwendung unterliegen. Das heißt, Prädiktive Analyse hat das Ziel, aus den einfachen Daten aus unserer Smartphonenutzung die Informationen zu „errechnen“, die wir gerade nicht preisgegeben haben.
Nun geben eine Reihe von Personen aber auch viele sensible Informationen von sich preis; nach dem Motto: „Ich habe ja nichts zu verbergen.“ Dies geschieht bei der Social-Media-Nutung, beim Einkaufen und sonstigen freiwilligen Angeboten wie Gewinn- und Mitmachspielen, deren Hauptzweck häufig allein die Datenabschöpfung ist, wo die Datenfreigabe aber meist freiwillig geschieht.
Dann werden aus sehr großen Mengen dieser Datensätze analysiert wie aus den in den meisten Datensätzen frei verfügbaren Daten, die auch von diesen Dateninhabern preisgegebenen sensiblen Informationen errechnet werden können. Sobald diese Berechnung hinreichend genau ist, werden diese so generierten Algorithmen auf Datensätze angewandt, in denen wenige oder keine sensiblen Informationen enthalten sind und beobachten diese Datensätze, ob z.B. die Dateninhaber von sich aus sensible Daten (z.B. bei einem Einkauf) ergänzen. Dann wird geschaut, inwieweit die KI dies vorherberechnet hat. Ist dies mit genügend großer Genauigkeit geschehen, wir dieser Vorgang in die Standardanalyse aufgenommen etc. Dies geschieht so lange, bis die Berechnung der sensiblen Daten einen genügend sicheren Schwellenwert erreicht hat. Usw. Auf diese Weise können sensible „Informationen“ auch von Personen „errechnet“ werden, die solche Information von sich gar nicht preisgegeben haben.
Andere Studien haben zum Beispiel auch ergeben, dass sich anhand von Facebook-Nutzerverhalten Krankheiten wie Depressionen, Psychosen und Diabetes oder Bluthochdruck sehr gut „errechnen lassen[10]. Aus der Nutzung verschiedener Facebook-Aktivitäten konnten persönliche Attribute wie sexu-

---

8 Mühlhoff, R (2020): Automatisierte Ungleichheit: Ethik der Künstlichen Intelligenz in der biopolitischen Wende des Digitalen Kapitalismus. In: Deutsche Zeitschrift für Philosophie, 68(6), S. 867–890. DOI: 10.1515/ dzph-2020-0059
Mühlhoff, R. (2020): Prädiktive Privatheit: Warum wir alle »etwas zu verbergen haben«. In: #VerantwortungKI – Künstliche Intelligenz und gesellschaftliche Folgen, (Hg.) C. Markschies & I. Hermann. Bd. 3/2020. Berlin-Brandenburgische Akademie der Wissenschaften.
Mühlhoff, R. (2020): Predictive Privacy: Towards an Applied Ethics of Data Analytics. SSRN preprint, https://ssrn.com/abstract=3724185

9 Wachter, S. (2019): Data Protection in the Age of Big Data. In: Nature Electronics 2 (1); S. 6-7

10 Duhigg, C (2012): How Companies Learn Your Secrets. In: The New York Times, 16. Februar 2012, https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html [20.11.2020].

elle Orientierung, ethnische Zugehörigkeit, religiöse und politische Ausrichtung, Intelligenz, Suchtverhalten, Trennung von den Eltern, das Alter, Geschlecht usw. ermittelt werden.
Solche Informationen werden dann zum Beispiel von Kreditfirmen wie Kreditec[11] verwendet. Diese Nutzen dann solche Analysen, um potenzielle Kunden zu finden. Was heißt, sie suchen Kunden, die eine schlechte Bonität haben, dringend Geld benötigen und wo trotzdem eine gewisse Wahrscheinlichkeit besteht, das Geld von den Schuldnern zurück zu erhalten. Wird dann ein Kredit angeboten, ist es bis zu einem Jahreszins von 300%. Es wird aber angegeben der Zinssatz beträgt nur 5,7% (und irgendwo steht dann, dass sich dies auf eine Woche bezieht.) Aber 5,7% hört sich erst einmal nicht viel an, wenn normale Zinssätze 3,5% betragen. Und wenn jemandem „das Wasser bis zum Hals steht“ liest der möglicherweise nicht alles und schon steckt dieser Mensch noch tiefer in der Schuldenfalle.
Das heißt, durch diese prädiktive Analyse wird die Gesellschaft durch diese Algorithmen in unsichtbare soziale Klassen eingeteilt. Es wird auch errechnet wie der Gesundheitszustand ist, das Lernverhalten in der Schule oder ob es demnächst Probleme mit dem Jugendamt geben könnte.
Durch diese prädiktive Analyse können privateste Aspekte erkannt werden, ohne dass sensible Daten faktisch bei oder über diese Person erhoben oder sonst wie beschafft wurden und die man auch nirgendwo preisgegeben hat.

Das heißt, hier greift das bisherige Verständnis von Datenschutz gar nicht mehr. Hier kommen sensible Informationen nicht durch ein Datenleck, heimliches Abschöpfen von Daten, Entschlüsselung von sorgfältig verwahrten Daten oder unerlaubte Weitergabe zur Auswertung. Sondern diese Informationsgewinnung geschieht durch Berechnung von Wahrscheinlichkeiten. Unsere Bonität, unsere politische Einstellung, unser Kauferhalten wird berechnet, obwohl wir sie nirgendwo preisgegeben haben. Diese Prognosen sind nur möglich, weil andere Personen überhaupt keine Bedenken hatten, freiwillig all ihre sehr sensiblen Daten (zum Trainieren einer KI) zur Verfügung zu stellen, nach dem Motto: „Ich hab ja nichts zu verbergen“ Und wenn das genügend Menschen tun, können wir, die wir möglicherweise mit Daten sparsam umgehen, weitgehend „berechnet“ werden.
Es reicht für die Art der Berechnung, dass eine Minderheit auf ihre Privatheit verzichtet. Denn anhand dieser Daten durchschnittlicher Nutzer werden Datensätze verfügbar gemacht, die dann für solch Prädiktive Analysen herangezogen werden können. Das heißt, der Datenschutz muss eigentlich neu vermessen werden. Denn. wenn wir unsere Daten schützen, schützen wir damit auch andere Personen vor Nachteilen. Man könnte sagen, ein verantwortlicher und sparsamer Umgang mit eigenen Daten ist auch eine zivilgesellschaftliche Aufgabe. Das heißt, der Datenschutz so verstanden, geht über die derzeit rechtliche Normung hinaus.

Quellen und Herkunft

Cookiepedia https://cookiepedia.co.uk/ (21.06.2021)

Duhigg, C (2012): How Companies Learn Your Secrets. In: The New York Times, 16. Februar 2012, https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html (20.11.2020).

Fanta, A. (2018): Ob Nutzer oder nicht: Facebook legt Schattenprofile über alle an. In: netzpolitik.org/2018/ob-nutzer-oder-nicht-facebook-legt-schattenprofile-ueber-alle-an/ (Abfrage: 18.08.2020)

Hawalkowski, B. (2021): Erschreckend: Das alles weiß WhatsApp über dich https://www.inside-digital.de/ratgeber/erschreckend-das-alles-weiss-whatsapp-ueber-dich

Link, M./Rehberg, A. I. (2020): Trackerentdecker. Tracker in Android Apps finden. In: c't Daten schützen

[11] Mühlhoff, Rainer (2020): Automatisierte Ungleichheit. Ethik der Künstlichen Intelligenz in der biopolitischen Wende des Digitalen Kapitalismus. S. 874
https://doi.org/10.1515/dzph-2020-0059

Lippert, J (2014): ZestFinance Issues Small, High-Rate Loans, Uses Big Data to Weed out Deadbeats. In: Washington Post, 11. Oktober 2014.

Mayer, J. R./Mitchell, J. C. (2012): Third-party web tracking. Policy and technology. In: Security and Privacy (SP), IEEE Symposium on, S. 413–427. IEEE.

Mühlhoff, R (2020): Automatisierte Ungleichheit: Ethik der Künstlichen Intelligenz in der biopolitischen Wende des Digitalen Kapitalismus. In: Deutsche Zeitschrift für Philosophie, 68(6), S. 867–890. DOI: 10.1515/ dzph-2020-0059

Mühlhoff, R. (2020): Prädiktive Privatheit: Warum wir alle »etwas zu verbergen haben«. In: #VerantwortungKI – Künstliche Intelligenz und gesellschaftliche Folgen, (Hg.) C. Markschies & I. Hermann. Bd. 3/2020. Berlin-Brandenburgische Akademie der Wissenschaften.

Mühlhoff, R. (2020): Predictive Privacy: Towards an Applied Ethics of Data Analytics. SSRN preprint, https://ssrn.com/abstract=3724185

Mühlroth, R./Deutschbein, R. (2020): Diese Daten greift WhatsApp von Ihrem Handy ab In: www.techbook.de/apps/messenger/whatsapp-zugriff-facebook-datenschutz (Abfrage: 18.08.2020)

Quinn, R. R./Baur, R. Bile, T./Bremert, B./Büttner, B./Grigorjew, O./ Hagendorff, T./Heesen, J./Krämer, N./Meier, Y./Nebel, M./Neubaum, G./Ochs, C./Rossnagel, R./Thom, H. S./Weiler, S. (2018): White Paper Tracking. Beschreibung und Bewertung neuer Methoden (Hg.) Fraunhofer Institut für System- und Innovationsforschung. Schriftreihe Forum Privatheit und selbstbestimmtes Leben in der digitalen Welt. Karlsruhe

Stach, C. et al. (2020): Predicting personality from patterns of behavior collected with smartphones. Proceedings of the National Academy of Science of the United States of America. 17680–17687 | PNAS | July 28, 2020 | vol. 117 | no. 30
Im Internet unter: https://www.pnas.org/content/pnas/117/30/17680.full.pdf

ULD - Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein (2010): Verbraucher-Scoring. Wie bewertet mich die Wirtschaft? Kiel

Ullrich, E. (2018): Analyse des Messengerdienstes WhatsApp aus forensischer Sicht. Hochschule Mittweida, Fakultät Angewandte Computer- und Biowissenschaften. Bachelorarbeit. https://monami.hs-mittweida.de/frontdoor/deliver/index/docId/10799/file/BA_Ullrich.pdf

Virtuelles Datenschutzbüro (2019): https://www.datenschutz.de/wp-content/uploads/kalins-pdf/singles/datenspuren-im-internet-vermeiden.pdf https://www.datenschutz.de

Wachter, S. (2019): Data Protection in the Age of Big Data. In: Nature Electronics 2 (1), S. 6-7

Witmer-Goßner, E. (2021): Big Data – Datenspuren im Netz https://www.storage-insider.de/big-data-datenspuren-im-netz-a-1001669/

derThomas
cyber4EDU