Soziotechnische Einflussfaktoren auf die »digitale Souveränität« des Individuums

Zinaida Benenson, Felix Freiling, Klaus Meyer-Wegener

Abstract *Dieses Kapitel nimmt eine eher technische Perspektive auf das Problem der »digitalen Souveränität« des Subjekts ein. Wir betrachten drei verschiedene soziotechnische Entwicklungen, die die Möglichkeit beeinflussen, in einer digitalisierten Welt ein selbstbestimmtes und autonomes Leben zu führen. Die erste Entwicklung betrifft die zunehmende Abhängigkeit von wenigen großen Diensteanbietern, die zwar für die Sicherheit der Nutzenden sorgen, aber gleichzeitig auch über die uneingeschränkte Macht über Daten und Geräte verfügen. Die zweite Entwicklung betrifft den Mangel an menschenzentrierter Gestaltung von Schutzmechanismen, der die realistische Wahrnehmung möglicher Bedrohungen erschwert. Die dritte und letzte Entwicklung bezieht sich auf unklare Vorstellungen über die Genauigkeit und die Nützlichkeit von gesammelten Daten. In der Gesamttendenz führen diese Entwicklungen zu starken Einschränkungen – sowohl im Hinblick auf die faktische als auch auf die selbst wahrgenommene Souveränität.*

1. Einleitung

Mehr als die Hälfte der Weltbevölkerung nutzt heute das Internet. Vielen ist dabei nicht bewusst, dass sie ein Netzwerk von verbundenen Computern nutzen, und viele weitere wissen nicht, dass Regierungen und internationale Konzerne bei der Benutzung des Internets ihre persönlichen Daten für eine Vielzahl von Zwecken sammeln und verarbeiten. Die digitale Transformation führt zunehmend zu einer Gesellschaft, in der eine aktive Teilhabe ohne Computer- und Internetnutzung kaum noch möglich ist. In diesem Kapitel legen wir den Fokus auf die »digitale Souveränität« des Subjekts, also die Möglichkeit, unter den oben beschriebenen Umständen ein selbstbestimmtes und autonomes Leben zu führen – sowohl im Hinblick auf die faktische als auch auf die selbst

wahrgenommene Souveränität des Individuums. Aus einer soziotechnischen Perspektive stellen wir allgemeine Beobachtungen zu drei relevanten Phänomenen an, die die »digitale Souveränität« stark beeinflussen und die mit den technischen Entwicklungen der Digitalisierung zusammenfallen.

Die erste Entwicklung wird durch den von Bruce Schneier (2012) geprägten Begriff der »feudalen Sicherheit« charakterisiert, der besagt, dass die Nutzenden heute vollständig von einem einzigen großen Diensteanbieter (oder einigen wenigen Diensteanbietern) abhängig sind. Diese Anbieter sorgen zwar für die Sicherheit der Personen, die die Dienste nutzen, haben aber auch uneingeschränkte Macht sowohl über ihre Daten als auch über die Funktionalität ihrer Geräte.

Die zweite Entwicklung betrifft die Probleme der Nutzenden, mögliche Bedrohungen für die Sicherheit und den Schutz der Privatsphäre realistisch wahrzunehmen, Risiken richtig einzuschätzen und angemessene Schutzmechanismen anzuwenden. Sicherheitsfachleute fordern häufig, dass das Verhalten der Nutzenden kontrolliert und eingeschränkt werden sollte und dass Verhaltensänderungen erzwungen werden müssen, um Sicherheitsprobleme einzudämmen. Diese Versuche, die Nutzenden zu »reparieren«, stehen jedoch in krassem Gegensatz zu Erkenntnissen aus der Forschung zur menschenzentrierten Sicherheit, die auf einen Mangel an menschenzentrierter Gestaltung von Schutzmechanismen hinweisen.

Die dritte und letzte Entwicklung bezieht sich auf eigentümliche Vorstellungen vom Wert personenbezogener Daten, dem »Öl des 21. Jahrhunderts« (vgl. Spitz 2017). Zweifellos können personenbezogene Daten verwendet werden, um Personen hinsichtlich ihrer politischen Orientierung, ihrer Persönlichkeitsmerkmale oder anderer gewünschter Eigenschaften zu klassifizieren. Geschäftsmodelle, die auf derartigen Klassifikationen basieren, sind breit etabliert. Die Genauigkeit und die Nützlichkeit dieser Klassifizierungen hängen jedoch vom angewandten Algorithmus und der Qualität der zugrunde liegenden Daten ab. Beide Aspekte führen zu Unsicherheiten, die schwer zu quantifizieren sind und weder durch die derzeitige Tendenz, alle Nutzungsdaten zu speichern, noch durch den Grundsatz der Zweckbindung bei der Nutzung personenbezogener Daten gemildert werden.

Die eher technisch gelagerte und demnach etwas eingeschränkte Perspektive dieses Kapitels hat ihren Ursprung in der wissenschaftlichen Sozialisation der Autorin und der Autoren, die aus der Informatik stammen – einer Disziplin, die noch vergleichsweise jung ist. Dies mag auch eine Erklärung dafür sein, dass der historische Betrachtungshorizont recht eng ist und vor allem

den Status quo betrachtet, also im Wesentlichen ausschließlich die Zeit der Expansion des kommerziellen Internets in den vergangenen 30 Jahren. Wir blicken auch nicht in die Zukunft, sondern beschränken uns auf die Nennung von Problembereichen, die maßgeblich durch »Errungenschaften« der Informatik entstanden sind. Gleichzeitig glauben wir, dass die Problembereiche nicht ohne eine kundige Kommentierung aus der Informatik voll verstanden werden können. Denn wie zu zeigen sein wird, beeinflussen die drei im Verlauf dieses Kapitels skizzierten Phänomene die »digitale Souveränität« weder vollständig positiv noch ausnahmslos negativ. Kühne Lösungsvorschläge sind demnach anderen Kapiteln dieses Bandes vorenthalten. Zur Einordnung existierender Phänomene sind die folgenden Betrachtungen möglicherweise trotzdem hilfreich.

Was sind Digitalisierung und Datafizierung?

Der Übergang von mechanischen und analogen elektronischen Systemen zur digitalen Elektronik begann in der zweiten Hälfte des 20. Jahrhunderts und hatte weltweit tiefgreifende wirtschaftliche, ökologische und gesellschaftliche Auswirkungen. Die zentrale Beobachtung, die u.a. von Shannon (1938) und Turing (1937) gemacht wurde, ist, dass Informationen in eine Folge von binären Ziffern (Bits) kodiert werden können, die ohne Qualitätsverlust dauerhaft gespeichert und auf denen allgemeine Berechnungen mithilfe digitaler Schaltkreise durchgeführt werden können. Eine solche digitale Informationsverarbeitung ist in gewissem Sinne universell: Jede Information kann in binären Ziffern kodiert werden, und binäre Ziffern können von universell programmierbaren digitalen Computern verarbeitet werden, die nur durch die Gesetze der Berechenbarkeit eingeschränkt sind (Hopcroft/Ullman/Motwani 2006). Während z.B. früher verschiedene Technologien verwendet wurden, um Text, Audio und Video zu speichern, zu übertragen und mit ihnen zu interagieren, wird heute alles auf kleinen universellen Computergeräten (Smartphones) erledigt, die anerkannten Regeln der Informationskodierung folgen (wie etwa mp3).

Der Begriff der *Digitalisierung* bezieht sich traditionell auf die Umwandlung von Informationen in eine digitale Form unter Verwendung bestimmter Kodierungsregeln. Heute wird der Begriff oft allgemeiner verwendet und verweist auf die zunehmende Nutzung digitaler Kommunikations- und Datenverarbeitungsgeräte in der Gesellschaft sowie auf die Übertragung einer wachsenden Zahl von Aufgaben an digitale Computer. Vor allem seit dem Aufkommen der nahtlosen globalen Vernetzung (»Internet«) zum Ende des

20. Jahrhunderts sind die Auswirkungen insbesondere auf die Unternehmen beträchtlich, zwingen diese zu einer »digitalen Transformation« (Matt/Hess/Benlian 2015) und haben einen ganz neuen Zweig datenorientierter Dienstleistungsunternehmen (wie Google und Facebook) hervorgebracht. Die Entwicklung wird durch die zunehmende Verfügbarkeit billiger Sensoren wie Kameras sowie lokaler und ferngesteuerter Erfassungsinfrastrukturen vorangetrieben, aber auch durch die allgegenwärtige Verbreitung persönlicher und verhaltensbezogener Daten, die von Einzelpersonen mithilfe digitaler Geräte erzeugt werden. Die Ära der digitalen Datenverarbeitung hat die Produktion digitaler Daten katalysiert und die anfänglichen Schwierigkeiten der Digitalisierung überwunden. Heute werden viele Daten »digital geboren« oder bei ihrer Entstehung nahtlos digitalisiert.

Die Datenproduktion bezieht sich nicht nur auf »Primärdaten«, die von Menschen erzeugt werden oder Sensormessungen widerspiegeln; auch digitale Berechnungen selbst erzeugen Daten, entweder als Ergebnisse der Berechnung von (primären) Eingabedaten oder als Daten, die den Berechnungsprozess aufzeichnen oder anderweitig beschreiben. Solche Daten werden gewöhnlich als »Sekundärdaten« oder »Metadaten« bezeichnet. Zu den Metadaten einer Berechnung gehören beispielsweise die Uhrzeit und das Datum der Berechnung, ein Verweis auf die Quelle der verarbeiteten Daten oder der Name der nutzenden Person, die für die Berechnung verantwortlich ist. Wenn einige Daten zur Beantwortung einer bestimmten Frage benötigt werden (in einem Produktionssystem oder in Bezug auf vergangenes persönliches Verhalten), ist es in der Regel kein Problem, das System so anzupassen, dass diese Daten für künftige Entscheidungen auch noch mitgespeichert werden. Die zunehmenden Rechengeschwindigkeiten, Bandbreiten der Kommunikationsnetze und Speicherkapazitäten haben zu dem Glauben an die allgegenwärtige Verfügbarkeit nützlicher und präziser Daten geführt, auf denen »gute Vorhersagen« basieren können (Dhar 2013). Folglich bezieht sich der Begriff der *Datafizierung* auf das Unterfangen, jede Frage in eine Frage nach Daten zu verwandeln.

Einerseits scheint die allgegenwärtige Verfügbarkeit von scheinbar fundierten Informationen auf Knopfdruck das Versprechen zu erfüllen, dass die Welt transparenter wird. Darüber hinaus können die vermeintliche Objektivität von Daten und ihre allgemeine Verfügbarkeit zu der Überzeugung führen, dass ihre Nutzung zu einer Verbesserung der Entscheidungsfindung von Organisationen und Einzelpersonen führen wird. Andererseits beruhen Informationen immer auf der Interpretation von Daten (bis hin zu den Kodierungsregeln der Datenspeicherung), und viele Antworten werden durch die Auswahl

der bereitgestellten Daten (und damit der gestellten Fragen) bestimmt. Überdies schafft die breite Verfügbarkeit und die gefühlte Transparenz personenbezogener Daten ein umgekehrtes Problem: Ein Mehr an Daten führt zu einem potenziellen Weniger an Privatsphäre.

Datafizierung, das Sammeln von persönlichen Daten und detaillierten Personenprofilen

In der Standardterminologie der Computersicherheit (vgl. Gollmann 2011) bedeutet das Sicherheitsziel der *Vertraulichkeit*, dass bestimmte Informationen nur befugten Personen zugänglich sind. Der Begriff der *Privatsphäre* bezieht sich in der Regel auf den Schutz der Vertraulichkeit von persönlichen Daten (im Gegensatz zu Daten, die einer Organisation gehören). Während der Begriff der Privatsphäre auch schon vor dem Aufkommen von Computern eine Bedeutung hatte, haben die mit der Digitalisierung einhergehenden Umstände der Datenproduktion und -verarbeitung zu einer komplexen Diskussion geführt, die über den technischen Bereich hinausgeht. Ein Grundstein für unser modernes Verständnis von Privatsphäre ist ein Urteil des Bundesverfassungsgerichts aus dem Jahr 1983. Darin wurde das Grundrecht auf informationelle Selbstbestimmung als die grundsätzliche Befugnis der einzelnen Person definiert, selbst zu entscheiden, welche Aspekte des eigenen Privatlebens anderen mitgeteilt werden sollen (vgl. BVerfG 1983). Dieses Urteil hat die von Westin (1970) entwickelten Ideen rechtlich kodifiziert und bildet heute die Grundlage des modernen europäischen Datenschutzrechts, einschließlich der Datenschutz-Grundverordnung (DSGVO bzw. General Data Protection Regulation, GDPR; vgl. European Parliament/Council of the European Union 2016).

Beim Schutz der Privatsphäre spielen personenbezogene Daten eine zentrale Rolle. Gemäß Artikel 4 der Datenschutz-Grundverordnung sind personenbezogene Daten definiert als »alle Informationen über eine bestimmte oder bestimmbare natürliche Person«. Die Definition ist recht weit gefasst und bezieht sich auf alle Daten, durch die eine natürliche Person (direkt oder indirekt) identifiziert werden kann. Dies umfasst offensichtliche »langfristige« Daten (wie Name, Geschlecht, Fingerabdruck, Sozialversicherungsnummer), mittelfristige Daten (Körpergröße, Handynummer) und auch kurzfristige Daten (aktueller Standort, persönliche Stimmung). Auch wenn einige dieser Daten subjektiv datenschutzrelevanter sind als andere, ist es durch die universelle Erfassung solcher Daten möglich, sich der persönlichen Identität einer Person anzunähern. Im Zusammenhang mit der Digitali-

sierung ist dies besonders relevant, da personenbezogene Daten häufig zur Identifizierung von Personen im Internet verwendet werden, was die Gefahr des Identitätsdiebstahls birgt (vgl. Hoofnagle 2007).

Angetrieben durch den Drang zur Datafizierung, ist der Trend zur Datenproduktion in vielerlei Hinsicht auch ein Trend zur Produktion von personenbezogenen Daten (vgl. Wylie 2019). Während bis etwa zum Jahr 2000 Computer oft von mehreren Personen gemeinsam genutzt wurden, sind Computer heute in den meisten Fällen persönliche Geräte, die nur von einer Person verwendet werden. Daher werden allein durch persönliche Interaktionen mit Computern bereits personenbezogene Daten erzeugt. Diese Daten lassen sich heute auch viel leichter einer Person zuordnen als in früheren Jahren, was die Erhebung personenbezogener Daten in einem Maße vereinfacht, dass Unternehmen vollständig individualisierte Dienste anbieten können. Dienste also, die eine einzigartige, auf bestimmte Personen zugeschnittene Zusammensetzung aufweisen, einschließlich ihrer thematischen und emotionalen Vorlieben bis hin zu aktuellen Meinungen und Stimmungen. Datenökosysteme, wie sie um soziale Netzwerke herum entstanden sind, zielen darauf ab, möglichst viele persönliche Daten zu sammeln, um detaillierte persönliche Profile zusammenzustellen, die manchmal sogar detaillierter erscheinen als die eigene persönliche Erinnerung (vgl. ebd.), aber dennoch nicht immer korrekt sein müssen (vgl. Garfinkel 2001).

Die Kluft zwischen faktischer und erlebter Souveränität eines Subjekts

Insgesamt ist die informationelle Selbstbestimmung also ein Grundrecht, das durch modernes Datenschutzrecht wie die DSGVO gestärkt wird. Aber die Digitalisierung und individualisierte Dienste stellen die Wahrnehmung und Gewährleistung dieses Rechts vor Herausforderungen, entweder durch die richtige Konfiguration von Datenschutzeinstellungen und Technologien zum Schutz der Privatsphäre oder durch die Rechenschaftspflicht von Diensteanbietern gegenüber Datenschutzbehörden (vgl. Kranig/Sachs/Gierschmann 2019). Wenn Daten bekannt werden, können sie außerdem nicht einfach vergessen werden. Die Kluft zwischen dem, was man können sollte, und dem, was man kann, wird immer größer, und es hat den Anschein, dass sich diese Kluft durch den zusätzlichen Einsatz von Informatikmethoden zur Datenanalyse vergrößert, der zu rasanten Fortschritten bei der Gesichts- und Spracherkennung, der Audio- und Videoproduktion sowie der Vorhersage von Aktivitäten, Meinungen und Stimmungen geführt hat.

2. Zunehmende Abhängigkeit von großen technischen Ökosystemen

Lock-in-Effekte, fehlende interoperable Standards und »feudale Sicherheit«

Die Digitalisierung hat ein wirtschaftliches Phänomen ermöglicht, das Forschende heute als Plattformökonomie bezeichnen. Vereinfacht gesagt, ist eine Plattform eine Umgebung, in der zwei Marktteilnehmende (Kaufende und Verkaufende) zusammenkommen, um kommerziell zu interagieren (vgl. Rochet/Tirole 2003). Während die Plattform-Metapher für klassische Plattformen (wie Einkaufszentren oder Kredit- und Debitkarten) eher unverständlich und konstruiert erscheint, hat die Digitalisierung (und insbesondere das Internet) die Rolle von Plattformen verstärkt, zunächst in Form von Internet-Medienportalen, wo kostenlose Inhalte Zuschauende anziehen, die wiederum Werbetreibende anziehen, und in letzter Zeit als digitale Ökosysteme, die oft um ein Betriebssystem (wie Microsoft Windows oder Apple iOS) oder um vernetzte Online-Umgebungen wie einen Marktplatz (Amazon), eine Suchmaschine (Google) oder ein soziales Netzwerk (Facebook) herum aufgebaut sind. Auch wenn noch viele Fragen rund um die Entstehung digitaler Plattformen, ihre *Governance*, Geschäftsmodelle und Auswirkungen auf Märkte und Gesellschaft unerforscht sind (vgl. Gawer 2010), spielen sie heute zweifellos eine beherrschende Rolle und sind bei der Teilnahme an vielen gesellschaftlichen Aktivitäten kaum zu vermeiden.

Eines der ältesten und am besten dokumentierten Geschäftsmodelle digitaler Plattformen, das auf den frühen Versuchen beruht, Nutzende für Medien- und Informationsportale wie Yahoo! zu gewinnen, ist Werbung. Erfolgreiche Plattformen ziehen die »Augäpfel« (engl. *eyeballs*) der Nutzenden an (Rochet/Tirole 2003) und verkaufen diese Ansichten (engl. *views*) an Werbekundschaft. Die Werbetreibenden entlohnen die Plattformen für den nachgewiesenen Konsum von Werbung, etwa auf Grundlage der Anzahl der Aufrufe oder Klicks. Moderne Internetwerbung erfolgt sogar auf individueller Basis (vgl. Anderson/Moore 2006): Auf der Grundlage dessen, was die Plattform über die oder den Nutzenden weiß, führt sie in Echtzeit eine Auktion für jede einzelne Ansicht durch. Bei dieser Auktion können Werbetreibende Gebote abgeben, und die höchstbietende Partei erhält den Zuschlag für ihre Anzeige. Dieser Ansatz ermöglicht es, Werbung mit fast beliebig kleinen Budgets zu schalten, was ihn für kleine Unternehmen und Privatpersonen attraktiv macht. Es wird je-

doch auch allgemein angenommen, dass die Konversionsrate von der Betrachtung bis zum Kauf bei individualisierter Werbung viel höher ist, da speziell zugeschnittene Anzeigen eine höhere Chance haben, das Interesse der Betrachtenden zu wecken. Untersuchungen deuten zwar darauf hin, dass dies zutreffen könnte (vgl. ebd.), aber es fehlen verlässliche Zahlen, wie viel effektiver dieser Mechanismus wirklich ist.

Insgesamt hängt das Versprechen personalisierter Werbung mit der Menge an Wissen zusammen, die eine Plattform über eine bestimmte Person, die sie nutzt, zur Verfügung hat. Um Werbeeinnahmen zu generieren, haben Plattformen ein weit verbreitetes und zunehmend aggressives Nutzendentracking betrieben, das auf Cookies oder anderen Browserfunktionen (vgl. Nikiforakis et al. 2013; Pugliese et al. 2020) bis hin zu personalisierten Gerätekennungen oder Geräte-Fingerabdrücken (vgl. Kurtz et al. 2016) basiert. Auch über Methoden der geräteübergreifenden Verfolgung und Verknüpfung von Informationsquellen wurde berichtet (vgl. Arp et al. 2017). Aufgrund ihrer Verwendung sind die meisten der erhobenen Daten eindeutig als personenbezogene Daten zu qualifizieren.

Natürlich steht es den Nutzenden frei, mehrere Plattformen sowohl als kaufende als auch verkaufende Person zu nutzen. In der Wirtschaftsliteratur wird dieses Phänomen als *Multihoming* bezeichnet. Multihoming auf einer Seite des Marktes führt in der Regel zu einer Verschärfung des Wettbewerbs auf der anderen Seite. In solchen Situationen versuchen die Plattformen, Anreize zu schaffen, damit die Nutzenden eine »exklusivere Beziehung« (Rochet/Tirole 2003: 993) zur Plattform eingehen. Dieses Verhalten lässt sich heute bei digitalen Plattformen beobachten. Einzelpersonen sind an personalisierte Konten gebunden, an die viele Annehmlichkeiten wie Cloud-Speicher, Backup, E-Mail und automatische Gerätekonfiguration geknüpft sind. Darüber hinaus werden der nahtlose Austausch von Dokumenten, die Nachrichtenübermittlung und die Verwaltung persönlicher Kontakte und Einstellungen nur innerhalb der Plattform ermöglicht. Plattformen fungieren als digitale Ökosysteme, in denen persönliche Daten in Silos mit proprietärer Software, Kommunikationsprotokollen und Datenformaten gesammelt werden. Dies steht in krassem Gegensatz zur offenen und interoperablen Natur der frühen Internetprotokolle (wie E-Mail und Internet Relay Chat) und führt dazu, dass die Nutzenden – langfristig gesehen – eine Plattform wählen müssen, die sie hauptsächlich nutzen wollen. Die Folge ist eine technologisch erzwungene Bindung an eine einzige Plattform. Trotz der Versuche, Datenportabilität in der Datenschutz-Grundverordnung zu regeln, ist es heute immer noch

schwierig, zwischen digitalen Ökosystemen zu migrieren, und das wird sicherlich auch noch einige Zeit so bleiben (vgl. Syrmoudis et al. 2021).

Obwohl es kontraintuitiv erscheinen mag, ist eine besondere Annehmlichkeit der digitalen Ökosysteme die Datensicherheit. Dies ist eine direkte Folge von zwei Entwicklungen. Die erste Entwicklung bezieht sich auf das Ausmaß der Kontrolle, die die Nutzenden über ihre eigenen Geräte haben. Bei klassischen PCs war es üblich, dass die Nutzenden ihre Sicherheit selbst verwalteten (und dafür verantwortlich waren): Sie konnten ein bestimmtes Betriebssystem wählen und ihr System fast ohne Einschränkungen installieren und konfigurieren. Den Anbietern moderner Smartphones und Tablets ist es gelungen, Hardware und Software so zu integrieren, dass es für die Nutzenden immer schwieriger wird, die volle Kontrolle über ihre eigenen Geräte zu erlangen. Die Kontrolle verbleibt bei den Herstellenden des Geräts, die Sicherheitsrichtlinien wie »installiere nur Software, die in Apples App Store erhältlich ist« durchsetzen können. Die zweite Entwicklung fällt mit dem allgemeinen Trend zusammen, Daten »in der Cloud« zu speichern, d.h. auf Servern, die von Unternehmen wie Dropbox kontrolliert werden, die diesen Dienst anbieten. Bei diesen meist von digitalen Plattformen angebotenen Diensten werden die Daten heute auf Internetservern und nicht mehr auf dem Gerät selbst gespeichert. Die Folge dieser beiden Entwicklungen ist, dass digitale Plattformen die volle Kontrolle über diese Daten haben, was einerseits gut ist, da z.B. Datensicherung und Zugriffskontrolle viel professioneller gehandhabt werden, als dies eine durchschnittliche Person, die sie nutzt, tun könnte. Andererseits müssen die Nutzenden diesen digitalen Plattformen volles Vertrauen entgegenbringen, sowohl was den rechtmäßigen Zugriff von Strafverfolgungsbehörden auf diese Daten angeht als auch die kontinuierliche Verfügbarkeit von Diensten mit fairen Preismodellen. Vor allem bei digitalen Plattformen, wo Nutzende durch die Bereitstellung von Daten bezahlen (wie Google und Facebook), befinden sich diese durch die Nutzungsbedingungen in einer sehr schwachen Position, da es kaum Beschränkungen dafür gibt, was die Plattformen mit ihren persönlichen Daten tun können. Dies schafft eine Situation, die als »feudale Sicherheit« (Schneier 2012) bezeichnet wurde, eine Situation, in der wir die Kontrolle über unsere Daten aufgeben, aber im Gegenzug darauf vertrauen, dass »unsere [Feudal-]Herren uns gut behandeln und vor Schaden bewahren werden«.

Glaube an technologieorientierte Lösungen (Privacy Enhancing Technologies)

Feudale Sicherheit ist ein eher einseitiges Vertrauensverhältnis, da digitale Plattformen kaum formale Garantien für ihre Dienste geben. Während viele Dienste behaupten, den Zugriff auf personenbezogene Daten durch Unbefugte zu verhindern, können die Diensteanbieter selbst die Daten in der Regel für jeden beliebigen Zweck nutzen (vgl. Stylianou/Venturini/Zingales 2015). Das übliche technische Modell des Datenschutzes basierte jedoch auf der Fähigkeit der oder des Einzelnen, die Offenlegung ihrer bzw. seiner persönlichen Daten zu steuern (vgl. Solove 2006). In digitalen Ökosystemen ist diese Fähigkeit (kodiert in den »Privatsphäre-Einstellungen« eines Kontos) jedoch darauf beschränkt, welche anderen *Nutzenden* (innerhalb desselben Ökosystems) auf die eigenen Daten zugreifen können. Der Zugriff durch die Plattform selbst kann nicht verhindert werden. Es gibt jedoch gut entwickelte technische Konzepte, mit denen Datenlecks und das Schutzniveau von persönlichen Daten zwischen beliebigen Parteien beschrieben werden können (vgl. Wagner/Eckhoff 2018). So besagt beispielsweise das Konzept der k-Anonymität (vgl. Sweeney 2002), dass innerhalb eines Datensatzes jede Person nicht von mindestens $k-1$ anderen Individuen des Datensatzes unterschieden werden kann. Es ist also möglich, präzise technische Garantien dafür zu formulieren, in welchem Umfang Informationen bei der Nutzung eines Dienstes geschützt sind.

Ähnlich wie bei der Entwicklung präziser Begriffe zur Beschreibung von Datenlecks hat die moderne Kryptografie ein Universum von Werkzeugen entwickelt, mit denen nahezu beliebige Berechnungen durchgeführt werden können, ohne dass Informationen an eine unbefugte Partei weitergegeben werden, die diese Informationen vorher nicht kannte. Zero-Knowledge-Protokolle ermöglichen es beispielsweise, eine Partei von einer Tatsache zu überzeugen, ohne diese Tatsache preiszugeben. Heutzutage ist es möglich, Algorithmen zu entwickeln, die es erlauben, die Weitergabe von persönlichen Daten auf ein einziges Informationsbit zu beschränken. Ein gutes Beispiel dafür ist das Protokoll für dezentrales, datenschutzgerechtes Proximity Tracing (DP3T), auf dem die meisten europäischen Warn-Apps gegen die Verbreitung von COVID-19 basieren (vgl. Troncoso et al. 2020). Das Protokoll stellt lediglich fest, ob eine möglicherweise gefährliche Begegnung innerhalb eines bestimmten Zeitraums stattgefunden hat oder nicht. Es ist nicht bekannt, mit wem der Kontakt stattfand oder wo er stattfand.

Dennoch ist die Entwicklung von Protokollen wie DP3T alles andere als trivial, insbesondere wenn die Anforderungen, wer was erfahren darf, nur vage definiert sind. In der Praxis wird daher argumentiert, dass der Schutz der Privatsphäre eine fortwährende »Identitätsmanagement«-Aufgabe ist, bei der Interessengruppen wie die Endnutzerinnen und -nutzer die Verwendung ihrer persönlichen Daten ständig überwachen und mithilfe technischer und rechtlicher Instrumente beeinflussen können. Im Zusammenhang mit der Nutzung von Gesundheitsdaten wurde dies als eine Verschiebung von einer reinen »Input-Orientierung« hin zu einer stärkeren »Output-Orientierung« charakterisiert (vgl. Deutscher Ethikrat 2017). Noch ist unklar, wie dieser Ansatz umgesetzt werden kann. Im technischen Bereich wurde die Idee untersucht, Entscheidungen an automatisierte Datenschutzassistenzprogramme unter der Kontrolle der oder des Einzelnen zu delegieren (vgl. Das et al. 2018), aber der Funktionsumfang, die Benutzungsfreundlichkeit und die Effektivität solcher Programme sind noch sehr rudimentär. Im Allgemeinen kann Identitätsmanagement auch an Organisationen delegiert werden, die die Datentreuhandschaft übernehmen, aber es ist unklar, wie für beide Seiten verständliche (und umsetzbare) Datenschutzerwartungen formuliert werden können (vgl. Rao et al. 2016). Es ist daher fraglich, ob digitale Plattformen als Datentreuhänder betrachtet werden können oder sollten.

Insgesamt scheint es immer noch eine ineffektive Kontrolle der Datenproduktion zu geben. Wie bereits erwähnt, ist der Preis für die Datenproduktion und die Datenspeicherung fast gleich null. Moderne Datenverarbeitungssysteme sind komplex, intransparent und schwer zu analysieren. Der Nachweis von Datenlecks ist zudem eine der komplexesten Aufgaben der Systemanalyse, da Informationen auf subtile Weise durch Seitenkanäle abfließen können, die sich willkürlich verschleiern lassen (vgl. Lampson 1973). Spezifische Vorschriften wie die Vorratsdatenspeicherung oder der gesetzlich vorgeschriebene Zugang zu Finanztransaktionen sind demnach Kompromisse, die technische mit rechtlichen Mechanismen koppeln, um diesem Dilemma zu begegnen.

Die Nichtexistenz von nicht personenbezogenen Daten

Obwohl auf den ersten Blick klar, ist die einfache Definition des Begriffs »personenbezogene Daten« in der Datenschutz-Grundverordnung (»alle Informationen über eine bestimmte oder bestimmbare natürliche Person«, vgl. European Parliament/Council of Europe 2016) auf den zweiten Blick erstaunlich komplex. Die Komplexität ergibt sich aus dem Attribut »bestimmbar«, d.h.

Daten können auch dann personenbezogene Daten sein, wenn sie sich nicht unmittelbar auf eine natürliche Person beziehen, aber mit einigem zusätzlichen »angemessenen« Aufwand zur Identifizierung einer Person verwendet werden können. Eine der zentralen Streitfragen im Bereich des Datenschutzes ist die Frage, was unter »angemessenem Aufwand« zu verstehen ist.

Es liegt auf der Hand, dass manche Daten viel weniger geeignet sind, einer Person zugeordnet zu werden, als andere. Der Temperaturwert des Meeresspiegels, der von einer automatischen Sensorstation in der Antarktis am Weihnachtsabend des vergangenen Jahres erfasst wurde, ist ein gängiges Beispiel für ein nicht personenbezogenes Datum. Aus konzeptioneller Sicht ist es jedoch immer möglich, eine Verbindung zu einem Menschen herzustellen und den Datenwert dieser natürlichen Person zuzuordnen. So könnte beispielsweise die Temperatur für die Personen relevant sein, die in der nahe gelegenen Forschungsstation arbeiten, oder die Daten könnten der Person zugeordnet werden, die die Datenabfrage durchgeführt hat. Für jeden Datensatz D gibt es also einen Datensatz D', sodass die Vereinigung von D und D' sich auf eine identifizierbare natürliche Person bezieht. Diese Argumentation zeigt, dass es zumindest theoretisch keine »nicht personenbezogenen Daten« gibt. Zwar ist diese Feststellung aus praktischer Sicht nicht sehr hilfreich, trägt aber zur Klärung bei, was mit dem »angemessenen Aufwand« erfasst werden soll (das Finden des Datensatzes D').

Die Literatur ist voll von Beispielen, in denen öffentlich verfügbare (und scheinbar anonyme) Datensätze durch Korrelation mit zusätzlichen Daten de-anonymisiert wurden. Eines der bekanntesten Beispiele aus der technischen Literatur ist die Re-Identifizierung von Personen, die sich bestimmte Filme angesehen haben – anhand eines von Netflix veröffentlichten anonymisierten Datensatzes (vgl. Narayanan/Shmatikov 2008). In jüngerer Zeit haben Forschende gezeigt, dass 99,98 Prozent der US-Amerikanerinnen und US-Amerikaner in jedem Datensatz anhand von nur 15 demografischen Attributen korrekt re-identifiziert werden können (vgl. Rocher/Hendrickx/de Montjoye 2019). In Anbetracht der Tatsache, dass es sehr genaue Datensammlungen über Verbrauchende gibt (insbesondere in den USA), ist es fraglich, ob ein ausreichendes Maß an Anonymität in einem öffentlich verfügbaren Datensatz erreicht werden kann: Während das Entfernen von Merkmalen aus einem Datensatz die Größe der Anonymitätsmenge schnell erhöht, hat das Hinzufügen von Merkmalen auch das Potenzial, die De-Anonymisierung exponentiell zu beschleunigen. Zumindest zeigen die obigen Beispiele deutlich, dass die Tatsache, ein personenbezogenes Datum zu sein, kein statisches,

sondern ein dynamisches Attribut ist, d.h. Daten, die zu einem Zeitpunkt keinen identifizierbaren Bezug zu einer natürlichen Person haben, können zu einem späteren Zeitpunkt tatsächlich einen solchen aufweisen (vgl. Hornung/Wagner 2019) – eine Erkenntnis, die die Materie nicht vereinfacht.

3. Herausforderungen der Selbstbestimmung

Mangelhafte mentale Modelle der Technologie

Mentale Modelle sind Vorstellungen der Menschen davon, wie Prozesse und Systeme funktionieren (vgl. Volkamer/Renaud 2013). Sie sind für die Entscheidungsfindung notwendig und müssen nicht korrekt sein. Stattdessen sollten mentale Modelle adäquat sein, d.h. sie sollten zu Entscheidungen führen, deren Ergebnisse für die Nutzenden von Vorteil sind (vgl. Camp 2009). So verstehen die Nutzenden womöglich nicht, wie ein Computervirus funktioniert, aber die Überzeugung, dass Viren ernsthaften Schaden anrichten können, kann zur Installation von Antivirensoftware führen (vgl. Wash 2010; Wash/Rader 2015).

Obwohl digitale Geräte und Dienste tagtäglich für eine Vielzahl wichtiger Aufgaben verwendet werden, fehlt vielen Nutzenden ein grundlegendes Verständnis dieser modernen Technologien, sodass ihre mentalen Modelle mangelhaft sind. Eine Untersuchung der mentalen Modelle von Smartphone-Apps ergab beispielsweise, dass die Hälfte der 24 befragten Nutzerinnen und Nutzer Apps nicht als Softwareprogramme wahrnahm, die Zugriff auf ihre Daten haben, sondern sie für Verknüpfungen zu Webseiten oder »Icons« hielt, mit denen sie nützliche Dinge tun konnten (vgl. King 2012). Selbst Nutzende mit hohem Allgemeinbildungsniveau weisen in diesem Bereich oft ein unzureichendes Bewusstsein und Wissen auf. So stellte sich heraus, dass mehrere aktive und gut ausgebildete E-Mail-Nutzende das Client-Server-Paradigma des Versendens und Empfangens von E-Mails nicht kannten und daher keinen Bedarf an einer Ende-zu-Ende-Verschlüsselung von E-Mails sahen (vgl. Renaud/Volkamer/Renkema-Padmos 2014). Kang et al. (2015) zitieren eine Person, die an ihrer Studie teilgenommen hat und die Funktionsweise des Internets folgendermaßen beschreibt: »Meine Daten gehen einfach überall hin.« Sie stellen fest, dass die mentalen Modelle der teilnehmenden Person ohne Informatikhintergrund »die Internet-Ebenen, -Organisationen und -Einheiten auslie-

ßen« (ebd.), was zu vielen falschen Vorstellungen über Sicherheit und Datenschutz führe.

Das Fehlen geeigneter mentaler Modelle der Technologie macht es den Nutzenden unmöglich, Gefahren für ihre Sicherheit und Privatsphäre in der digitalen Welt realistisch wahrzunehmen, Risiken richtig einzuschätzen und angemessene Schutzmechanismen anzuwenden. Dies macht auch die Entwicklung von benutzungsfreundlichen Schutzmechanismen zu einer besonderen Herausforderung: Forschung im Bereich Human-Computer-Interaction (HCI) zur Entwicklung von Maßnahmen zum Schutz von Sicherheit und Privatsphäre hat gezeigt, dass Laien meist nicht in der Lage sind, diese Maßnahmen zu verstehen und anzuwenden (vgl. Herley 2013).

Digitale Selbstbestimmung versus »Sicherheitsmüdigkeit«

Abgesehen davon, dass sie schwer zu verstehen und richtig anzuwenden sind, wurde festgestellt, dass Schutzmaßnahmen einen zu hohen kognitiven und zeitlichen Aufwand erfordern (bis hin zur Undurchführbarkeit) oder nicht den Sicherheitsnutzen bieten, der den Aufwand rechtfertigen würde (vgl. Herley 2009; Reeder/Ion/Consolvo 2017). Es gibt sogar den Begriff der Sicherheitsmüdigkeit (engl. *security fatigue*), der die Resignation der Benutzenden angesichts der vielen Sicherheitsanforderungen beschreibt, die oft mit ihren Lebens- und Arbeitsrealitäten und manchmal sogar miteinander in Konflikt stehen (vgl. Stanton et al. 2016).

Ähnlich unbefriedigend ist die Situation beim Datenschutz. Viele Nutzende scheinen gegenüber dem Schutz ihrer persönlichen Daten zu resignieren und akzeptieren einfach, dass sie nichts dafür tun können (vgl. Turow/Hennessy/Draper 2015). Darüber hinaus gibt es starke Hinweise aus der Verhaltensökonomie, dass Menschen, selbst wenn sie ihre Privatsphäre schützen wollten, dazu nicht in der Lage wären. So haben mehrere experimentelle Studien gezeigt, dass Entscheidungen zum Schutz der Privatsphäre durch einfache experimentelle Bedingungen, wie beispielsweise eine 15-sekündige Ablenkung oder leicht unterschiedliche Formulierungen der Datenschutzfragen, in Richtung einer größeren Offenlegung von Daten manipuliert werden können (vgl. Acquisti/Brandimarte/Loewenstein 2015).

Diese Erkenntnisse sind wichtig für die Datenschutz-Grundverordnung und ähnliche Vorschriften: Einerseits brauchen wir diese Regelungen, weil die Forschung wiederholt gezeigt hat, dass Menschen nicht in der Lage sind, ihre Entscheidungen in Bezug auf Sicherheit und Datenschutz so zu treffen, dass

sie für sie vorteilhaft sind und ihren Wünschen und Präferenzen entsprechen. Andererseits könnten die Forderungen dieser Verordnungen, den Nutzenden mehr Transparenz und Kontrolle über ihre persönlichen Daten zu geben, gerade deshalb vergeblich sein, weil es den Nutzenden an Wissen, Fähigkeiten und Ressourcen fehlen könnte, um ihre gesetzlichen Rechte sinnvoll zu nutzen.

Von technologiezentrierter zu menschenzentrierter Gestaltung von Schutzmaßnahmen

Sicherheitsfachleute geben häufig den Benutzenden die Schuld für ihr mangelndes Sicherheitsbewusstsein, für die Nichtbefolgung von Sicherheitsratschlägen und für die Verursachung von Sicherheitsproblemen. Diese Denkweise ist als »Fix the User«-Paradigma bekannt (vgl. Schneier 2016): Das Verhalten der Nutzenden sollte kontrolliert und eingeschränkt werden, und es müssen Verhaltensänderungen erzwungen werden, um Sicherheitsprobleme einzudämmen. Die bahnbrechende Veröffentlichung *Users are not the enemy* von Adams und Sasse (1999) hat es geschafft, zumindest einen Teil dieser auf Schuldzuweisungen basierenden IT-Sicherheitskultur grundlegend in Richtung konstruktiverer Ansätze zu verändern. In einer Reihe von Interviews zeigten die Autorinnen, dass die Nutzenden nicht aus Unachtsamkeit oder bösem Willen heraus unsicher handeln, sondern weil die Sicherheitsmechanismen nicht nutzerzentriert entwickelt wurden und dadurch eine schlechte Gebrauchstauglichkeit aufweisen.

Menschenzentrierte IT-Sicherheit und Datenschutz setzen auf das Verständnis der Fähigkeiten der Menschen für Sicherheits- und Datenschutzaufgaben. Außerdem sollen Sicherheit und Datenschutz stets im Kontext betrachtet werden: Was sind die Lebens- und Arbeitsrealitäten der Nutzenden, was sind ihre primären Ziele und Aufgaben? Wie interagieren die Ziele von Sicherheit und Datenschutz mit diesem Kontext? Dieses Verständnis ist notwendig für die Entwicklung angemessener Sicherheits- und Datenschutzmaßnahmen, die von der Mehrheit der Nutzenden verwendet werden können und dann auch verwendet werden. Das heißt, anstatt die Nutzenden an die Technologie anzupassen (was sowieso nicht gelingt), sollte man die Technologie an die Bedürfnisse und Fähigkeiten der Nutzenden anpassen (vgl. Sasse 2015). Mit dieser Anpassung beschäftigt sich seit mehr als 20 Jahren Forschung und Entwicklung im interdisziplinären Bereich »Usable Security« (Cranor/Garfinkel 2005; Garfinkel/Richter Lipford 2014). Zu Erfolgen dieser Forschung gehören u.a. die Änderungen der international anerkannten

Richtlinien zur Passwortverwaltung (vgl. NIST 2017). Diese Richtlinien vereinfachen die Passworthandhabung für die Nutzenden (z.B. werden keine komplizierten, langen Passwörter mehr verlangt) und zeigen gleichzeitig auf, welche organisatorischen und technischen Maßnahmen die Anbieter zur sicheren Verwaltung von vereinfachten Passwörtern treffen müssen.

Es gibt sowohl konzeptionelle als auch ökonomische Gründe dafür, dass benutzbare Sicherheitsmaßnahmen immer noch selten sind. Konzeptionell stellen »Benutzbarkeit« und »Sicherheit« zwei Systemeigenschaften dar, die in einem Aspekt zueinander sehr ähnlich charakterisiert werden: Sie sind für die Funktionsweise der Systeme nicht zwingend notwendig. Deswegen werden sie bei der Entwicklung oft später hinzugefügt, nachdem funktionale Eigenschaften implementiert und getestet wurden, was zu gegenseitiger Abschwächung der beiden Eigenschaften führt (vgl. Yee 2004): Wenn Sicherheit nicht von Anfang an in der Systemarchitektur vorhanden ist, werden beim späteren Hinzufügen der Sicherheitsmaßnahmen oft Hürden für die Benutzung eingebaut – die Benutzbarkeit sinkt. Bei dem Versuch, die Benutzbarkeit zu erhöhen, werden Sicherheitseigenschaften, beispielsweise vorgegebene Sicherheitseinstellungen, wieder abgeschwächt. Es ist deswegen unabdingbar, sowohl Sicherheit als auch Benutzbarkeit von Anfang an in den Entwicklungsprozess einzubinden (vgl. Sasse/Flechais 2005).

Jedoch würden solche Entwicklungsprozesse zusätzliche zeitliche, personelle und finanzielle Ressourcen beanspruchen, was aus der ökonomischen Perspektive problematisch ist. Die Verbraucherinnen und Verbraucher können die erhöhte IT-Sicherheit nicht selbst beurteilen und sind deswegen nicht bereit, für erhöhte Sicherheit mehr zu bezahlen (vgl. Anderson/Moore 2006). Folglich fehlen den Produzierenden ökonomische Anreize zum Erhöhen der IT-Sicherheit, die ihre begrenzten Ressourcen lieber in die schnelle Entwicklung von innovativen Lösungen mit attraktiver Funktionalität investieren. Dieses Marktversagen führt zur Notwendigkeit, IT-Sicherheit und Datenschutz zu regulieren, wie es beispielsweise im IT-Sicherheitsgesetz 2.0 (2021) und in der DSGVO (2016) vorgesehen ist. Zum Beispiel könnte das Anbringen von einem IT-Sicherheitskennzeichen die IT-Sicherheitsmaßnahmen der Hersteller sichtbar machen (vgl. Deutscher Bundestag 2021: Art. 1 § 9c) und zur Präferenz von Produkten mit besseren Sicherheitseigenschaften führen (vgl. Morgner et al. 2020). Regulierung führt in der Praxis jedoch nicht immer zur Verbesserung der Sicherheit und der Benutzbarkeit, wie die sehr negativen Nutzungserfahrungen mit kaum handhabbaren Cookie-Hinweisen auf den Webseiten gezeigt haben (vgl. Utz et al. 2019). Auch die praktische Umsetzbarkeit des IT-Sicher-

heitsgesetzes 2.0 wurde von mehreren Seiten kritisiert (s. Kipker/Scholz 2021 für einen Überblick).

4. Das Daten-Dilemma: Datenschutz vs. Datennutz

Der aktuelle Trend, alles aufzuheben

Zur »digitalen Souveränität« gehört auch zu wissen, welche Daten über einen selbst erfasst und welche Schlussfolgerungen aus ihnen gezogen wurden. Durch die Informationen des Whistleblowers Edward Snowden wurde 2014 bekannt, dass die NSA die Aufzeichnung des kompletten Internetverkehrs plante (vgl. Greenwald 2014). Was damals technisch zumindest als herausfordernd erachtet wurde, scheint heute ein plausibler Teil der Geschäftsstrategie vieler Unternehmen zu sein. Zwar ist nicht vollständig bekannt, was die oben bereits genannten digitalen Plattformen – die man summarisch gern als GAFAM (Google [Alphabet], Amazon, Facebook, Apple, Microsoft) bezeichnet – mit den Daten tun, die auf ihren Plattformen anfallen, aber mit großer Wahrscheinlichkeit heben auch sie erst einmal alles auf, was sie bekommen können, und nutzen es für ihre Zwecke. Wie oben schon diskutiert, sind das erst einmal Wege zur personalisierten Werbung und in gewissem Umfang auch zu personalisierten Services – ob die Benutzenden das nun wollen oder nicht. Es ist dann nur noch ein kleiner Schritt hin zu Manipulation und zur Modifikation des Verhaltens (vgl. Zuboff 2019). Hier wird definitiv feudale Sicherheit geboten, aber es ist nicht sicher, ob sich die Benutzenden das auch so wünschen.

Speicherplatz ist heute vergleichsweise preiswert und deshalb ausreichend verfügbar. Weil die Entscheidung, welche Daten wichtig sind und welche nicht, manchmal schwierig zu treffen ist, gibt es die deutliche Tendenz, erst einmal alles zu speichern und dann erst später zu entscheiden, was mit den Daten gemacht wird. Diese Herangehensweise wird im Volksmund gern als »Datengier« bezeichnet. Das bedeutet, es werden mehr Daten erfasst, als derzeit gebraucht werden, nur weil sie später vielleicht einmal nützlich sein könnten. Die nachträgliche Erweiterung eines Datenbestands ist immer mühsam und teuer, weshalb Entwickelnde dies zu vermeiden versuchen. Beides zusammen führt dann zu sogenannten »Daten-Friedhöfen«, also großen Datenmengen, die zwar gespeichert sind, aber nie gelesen werden.

Die Versprechen des Data Mining und der öffentlichen Daten, das Problem der Datenqualität und die Nachvollziehbarkeit von maschinellem Lernen

Viele Daten werden heute also ohne bestimmten Zweck und auf Vorrat gespeichert. Wenn sie nicht dauerhaft auf Daten-Friedhöfen landen, werden solche Datenbestände manchmal wiederentdeckt in der Hoffnung, sie doch noch nutzen zu können. Man vermutet darin etwa wichtige Informationen über das Verhalten von Kundinnen und Kunden oder über die Verbreitung von Infektionen. Was dann allerdings vorliegt, sind zumeist nur sogenannte »Rohdaten«, was bedeutet, dass sie meist nicht unmittelbar für eine Analyse geeignet sind. Es gibt eigentlich immer die wohlbekannten Probleme mit der Datenqualität: Unvollständigkeit (fehlende Werte), Ungenauigkeit, schlichte Fehler und Veralterung, um nur einige zu nennen.

Viele Algorithmen wurden schon vorgeschlagen, um diese Probleme zu beheben (vgl. Müller/Freytag 2003) (soweit das überhaupt möglich ist). Alle diese Verfahren der Datenbereinigung (*data cleaning* oder *data cleansing* genannt) nutzen aber Extrapolationen oder Schätzungen, führen also unter Umständen selbst auch wieder Ungenauigkeiten ein. Die tatsächlichen Werte zu finden, erfordert oft manuellen Eingriff und ist deshalb viel zu teuer. Es ist also notwendig, genau zu dokumentieren, welche Daten tatsächlich erhoben und welche nachträglich auf die eine oder andere Art rekonstruiert wurden. Bei dieser Dokumentation handelt es sich um Metadaten, die für die Nutzung der Daten von eminenter Bedeutung sind.

Eine zweite, ebenfalls sehr wichtige Aufgabe für die Sammelnden und Besitzenden ist in der Vorbereitung von Daten für die Analyse die Integration von Datenbeständen aus verschiedenen Quellen. Auch das verändert die Daten: Werte werden in ein gemeinsames Format umgewandelt, Dubletten identifiziert und gelöscht. Eventuell gelingt das nicht perfekt und verzerrt die Daten dadurch ein kleines bisschen, auch bei allem Bemühen, das zu vermeiden. Die Bestände sind aber oft einfach zu groß, um jedes Detail noch einmal zu prüfen. Es ist für die Betroffenen wichtig, diese Vorverarbeitung bei Bedarf nachvollziehen zu können und vor allem zu prüfen, ob damit (womöglich sogar ohne Absicht) falsche Eindrücke entstehen könnten.

Nach der Erfassung von Metadaten und der Datenintegration kann dann die Analyse beginnen. Es werden Schlüsse aus den Daten gezogen, die gravierende Konsequenzen haben können. Neue Daten werden aus den Beständen abgeleitet, zumeist in sehr viel kompakterer Form (»aggregiert«), sodass inter-

essante Dinge wie z.B. Trends, Gruppierungen und Verhaltensmuster besser erkannt werden können. Die abgeleiteten Daten beziehen sich auf eine große Gruppe von Elementen, die dann statistisch ausgewertet wird, oder auch auf ein einzelnes Individuum, das klassifiziert oder in einer größeren Menge lokalisiert wird.

Die Methoden der Datenanalyse können wie folgt gruppiert werden (vgl. Han/Kamber/Pei 2011):

a) Vorhersage: Was wird eine Person als nächstes tun (z.B. kaufen)? Wohin wird eine Person gehen?
b) Klassifikation: Was für eine Person ist es? Gut oder böse? Reich oder arm? Gesund oder krank?
c) Gruppierung (Clusterbildung): Welche Personen bilden eine Gruppe mit ähnlichen Eigenschaften?
d) Ausreißer: Was macht eine Person besonders, anders als die anderen?
e) Assoziationen: Was kommt oft zusammen vor?

All diese Methoden brauchen große Datenbestände (»Big Data«). Wichtiger noch, sie brauchen alle auch menschliche Eingriffe. So müssen Schwellenwerte festgelegt und Hyperparameter gesetzt werden. Auch was Ähnlichkeit ist, muss für die Methoden erst festgelegt werden. Dadurch kann das Ergebnis aber auch manipuliert werden. Es ist heute durchaus üblich, diese Systeme schrittweise zu verbessern, indem die Einstellungen einfach ausprobiert werden. Passen die Ergebnisse nicht, werden die Einstellungen etwas abgeändert und die Daten erneut analysiert. Auch hier sind Transparenz und Nachvollziehbarkeit für die Betroffenen ein Desiderat.

Verzerrungen zwischen gespeicherten Daten und der Realität: Kann ein perfekter Zwilling erzeugt werden?

Das Konzept des »digitalen Zwillings« ist heutzutage sehr populär. Mit diesem Begriff ist eine digitale Repräsentation eines echten Objekts gemeint, die in möglichst vielen Aspekten diesem Objekt ähnelt. Genutzt werden digitale Zwillinge bei Gegenständen und Produkten, die überwacht und gewartet werden müssen. Das Konzept kann aber auch auf Menschen angewendet werden. Der Glaube daran, einen perfekten digitalen Zwilling erschaffen zu können, führt dazu, dass der Zwilling in manchen Situationen das reale Objekt ersetzt und allein die digitale Repräsentation die Basis für Entscheidungen liefert, die

das reale Objekt betreffen. Handelt es sich dabei um eine Person, kann deren Souveränität dadurch beeinträchtigt werden, denn man »fragt« nicht mehr sie, sondern den Zwilling.

Allerdings ist bereits aus den frühen Tagen des Datenbankeinsatzes bekannt, dass Daten nur ein Modell der Realität sind und dieses niemals ein vollständiges Bild ergeben kann (vgl. Kent 2012 [1978]). Deshalb ist der digitale Zwilling eigentlich eine Fiktion: Er suggeriert Vollständigkeit, die nie erreicht werden kann. (Vielleicht ist deshalb inzwischen auch die Bezeichnung »digitaler Schatten« auf dem Vormarsch.) Daher ist es sehr viel sinnvoller, einen bestimmten Zweck für das Sammeln und die Analyse der Daten zu definieren. In der Entwicklung von Datenbanken wurde gern der Begriff »Miniwelt« verwendet, um deutlich zu machen, dass nur ein kleiner Ausschnitt der Realität durch die Daten beschrieben wird. Trotzdem waren und sind die mit dieser Einschränkung entwickelten Datenbanken durchaus nützlich, und zwar sowohl für die Betreibenden als auch für deren Kundschaft, weil Vorgänge effizienter abgewickelt werden können. Entscheidend ist, dass die Daten trotz ihrer Unvollständigkeit nicht einfach pragmatisch als digitaler Zwilling genutzt werden dürfen. Diese Gefahr besteht, wenn keine anderen Daten vorliegen. Hier sollte die »digitale Souveränität« der Betroffenen erlauben, den »Zwilling« zu überprüfen und auf dessen Unvollständigkeit hinzuweisen – ohne gleich gezwungen zu werden, dann doch bitte für Vollständigkeit zu sorgen.

Die Schwierigkeiten der Datensparsamkeit, Kontrolle über die Daten und das Prinzip der Zweckbindung

Leider kann der Zweck des Datensammelns oft nur unpräzise definiert werden, außerdem ändert er sich mit der Zeit. Trotzdem ist es sehr sinnvoll, zumindest den Versuch zu machen, diesen Zweck zu identifizieren. Er sollte dokumentiert und, wo möglich, auch publiziert werden. Das hilft, die oben beschriebenen Tendenzen zu vermeiden, und es kann in erheblichem Umfang Ressourcen einsparen – vom Energieverbrauch bis zur Arbeitszeit der Angestellten. Datensparsamkeit, also die Beschränkung auf die Daten, die tatsächlich gebraucht werden, ist also nicht nur eine Frage des Datenschutzes, sondern entfaltet auch ökonomische und ökologische Wirkungen. Eine Konsequenz daraus ist auch das Löschen der Daten, die nicht mehr gebraucht werden.

5. Herausforderungen für die Forschung

Wie steht es nun um die »digitale Souveränität« des Individuums in der digitalisierten Welt? Unter welchen Umständen kann man heute und in Zukunft ein selbstbestimmtes und autonomes Leben führen? Die Betrachtung aus technischer und soziotechnischer Perspektive wirft mehr Fragen auf, als sie Antworten gibt, denn die großen Diensteanbieter entziehen Individuen eher Kontrolle über ihre Geräte und Daten, was faktisch und gefühlt zu einem Weniger an Selbstbestimmung führt. Dies schafft Abhängigkeiten, die auch Auswirkungen auf die Möglichkeit von Nutzenden haben, etwaige Bedrohungen für die Sicherheit und den Schutz der Privatsphäre realistisch wahrzunehmen. Noch ist in höchstem Maße unklar, unter welchen Umständen Nutzende überhaupt selbstbestimmte Entscheidungen über ihre eigene Sicherheit und Privatsphäre in einer digitalisierten Welt treffen können. Sicherlich gehört die menschzentrierte Gestaltung von Schutzmechanismen dazu, welche aber auch immer durch den Grad des Verständnisses für die technischen und soziotechnischen Zusammenhänge bedingt sind. Was muss man wissen und verstehen, um in einer digitalisierten Welt selbstbestimmt zu leben? Um die Risiken für die eigene Sicherheit und Privatsphäre richtig einzuschätzen sowie angemessene Schutzmechanismen anzuwenden? Was ist überhaupt ein akzeptables individuelles Schutzniveau? Wo liegen die Grenzen von Schutzmaßnahmen, und unter welchen Umständen ist es demnach sinnvoller Risiken einzugehen? Wie können Schutzmaßnahmen aussehen, die Nutzende dabei unterstützen, die vorhandenen Zielkonflikte besser abzuwägen? Wie verändern sich diese Umstände über die Zeit?

Über allem steht der oft suggerierte »hohe Wert« von (personenbezogenen) Daten, der immer wieder durch einen Blick auf den spekulativen Börsenwert der Internetkonzerne eher anekdotisch bestätigt wird. Wie sich dieser Wert jenseits von der Hoffnung auf zukünftige Unternehmensgewinne ausdrückt und unter welchen Umständen er entsteht, ist weiterhin hochgradig unklar – denn zu wenig ist bekannt über die Genauigkeit und die Nützlichkeit dieser Daten sowie über die Wirksamkeit der Algorithmen, mit denen sie verarbeitet werden. Eine differenzierte Betrachtung dieses Phänomens tut Not, denn ein gefühlter Verlust an individueller Souveränität ist langfristig genauso wirksam wie ein real existierender.

Literaturverzeichnis

Acquisti, Alessandro/Brandimarte, Laura/Loewenstein, George (2015): »Privacy and human behavior in the age of information«, in: Science 347 (6221), S. 509–514.

Adams, Anne/Sasse, M. Angela (1999): »Users are not the enemy«, in: Communications of the ACM 42 (12), S. 41–46.

Anderson, Ross/Moore, Tyler (2006): »The economics of information security«, in: Science 314 (5799), S. 610–613.

Arp, Daniel/Quiring, Erwin/Wressnegger, Christian/Rieck, Konrad (2017): »Privacy threats through ultrasonic side channels on mobile devices«, in: 2017 *IEEE European Symposium on Security and Privacy (EuroS&P)*, S. 35–47, https://doi.org/10.1109/EuroSP.2017.33.

Aziz, Arslan/Telang, Rahul (2015): What is a cookie worth? 14th *Annual Workshop on the Economics of Information Security (WEIS)*. Preliminary Draft, Heinz College, Carnegie Mellon University, Pittsburgh. Online unter: www.econ-infosec.org/archive/weis2015/papers/WEIS_2015_aziz.pdf, abgerufen am 24.02.2022.

BVerfG – Bundesverfassungsgericht (1983): Urteil des Ersten Senats vom 15. Dezember 1983 – 1 BvR 209/83 –, Rn. 1–215.

Camp, L. Jean (2009): »Mental models of privacy and security«, in: IEEE Technology and Society Magazine 28 (3), S. 37–46.

Cranor, Lorrie Faith/Garfinkel, Simson (2005): Security and usability: Designing secure systems that people can use, Beijing u.a.: O'Reilly.

Das, Anupam/Degeling, Martin/Smullen, Daniel/Sadeh, Norman (2018): »Personalized privacy assistants for the internet of things: Providing users with notice and choice«, in: IEEE Pervasive Computing 17 (3), S. 35–46, https://doi.org/10.1109/MPRV.2018.03367733.

Deutscher Bundestag (2021): Zweites Gesetz zur Erhöhung der Sicherheit informationstechnischer Systeme (IT-Sicherheitsgesetz 2.0), Bundesgesetzblatt Teil I Nr. 25, Bonn: Bundesanzeiger Verlag.

Deutscher Ethikrat (2017): Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung. Stellungnahme, 30.11.2017. Online unter: https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-big-data-und-gesundheit.pdf, abgerufen am 20.08.2021.

Dhar, Vasant (2013): »Data science and prediction«, in: Communications of the ACM 56 (12), S. 64–73, https://doi.org/10.1145/2500499.

European Parliament/Council of the European Union (2016): Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Online unter: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679, abgerufen am 20.08.2021.

Garfinkel, Simson (2001): Database nation, Beijing u.a.: O'Reilly.

Garfinkel, Simson/Richter Lipford, Heather (2014): »Usable security: History, themes, and challenges«, in: Synthesis Lectures on Information Security, Privacy, and Trust 5 (2), S. 1–124.

Gawer, Annabelle (Hg.) (2010): Platforms, markets and innovation, Cheltenham/Northampton: Edward Elgar Publishing.

Gollmann, Dieter (2011): Computer security, Chichester: Wiley.

Grassi, Paul A./Fenton, James L./Newton, Elaine M./Perlner, Ray A./Regenscheid, Andrew R./Burr, William E./Richter, Justin P./Lefkovitz, Naomi B./Danker, Jamie M./Choong, Yee-Yin/Greene, Kristen K./Theofanos, Mary F. (2017): Digital identity guidelines. Authentication and lifecycle management (NIST Special Publication 800–63B), https://doi.org/10.6028/NIST.SP.800-63b.

Greenwald, Glenn (2014): No place to hide: Edward Snowden, the NSA, and the U.S. surveillance state, New York: Picador Metropolitan Books, Henry Holt.

Han, Jiawei/Kamber, Micheline/Pei, Jian (2011): Data mining: Concepts and techniques (= The Morgan Kaufmann Series in Data Management Systems), Burlington: Morgan Kaufmann.

Herley, Cormac (2009): »So long, and no thanks for the externalities: The rational rejection of security advice by users«, in: New security paradigms workshop (NSPW'09), Association for Computing Machinery (ACM), S. 133–144, https://doi.org/10.1145/1719030.1719050.

Herley, Cormac (2013): »More is not the answer«, in: IEEE Security & Privacy 12 (1), S. 14–19.

Hoofnagle, Chris Jay (2007): »Identity theft: Making the known unknowns known«, in: Harvard Journal of Law and Technology 21 (1).

Hopcroft, John E./Ullman, Jeffrey D./Motwani, Rajeev (2006): Introduction to automata theory, languages, and computation, Upper Saddle River: Prentice Hall.

Hornung, Gerrit/Wagner, Bernd (2019): »Der schleichende Personenbezug«, in: Computer und Recht 35 (9), S. 565–574.

Kang, Ruogu/Dabbish, Laura/Fruchter, Nathaniel/Kiesler, Sara (2015): »My data just goes everywhere: User mental models of the internet and implications for privacy and security«, in: Eleventh Symposium on Usable Privacy and Security (SOUPS), USENIX Association, S. 39–52.

Kent, William (2012 [1978]): Data and reality: A timeless perspective on perceiving and managing information in our imprecise world. Westfield: Technics Publications.

King, Jennifer (2012): How come I'm allowing strangers to go through my phone? Smartphones and privacy expectations, SSRN vom 15.03.2012, ht tp://dx.doi.org/10.2139/ssrn.2493412.

Kipker, Dennis-Kenji/Scholz, Dario E. (2021): »Das IT-Sicherheitsgesetz 2.0«, in: Datenschutz und Datensicherheit – DuD 45.1, S. 40–45.

Kranig, Thomas/Sachs, Andreas/Gierschmann, Markus (2019): Datenschutz-Compliance nach der DS-GVO. Köln: Reguvis Fachmedien.

Kurtz, Andreas/Gascon, Hugo/Becker, Tobias/Rieck, Konrad/Freiling, Felix C. (2016): »Fingerprinting mobile devices using personalized configurations«, in: Proceedings on Privacy Enhancing Technologies (1), S. 4–19.

Lampson, Butler W. (1973): »A note on the confinement problem«, in: Communications of the ACM 16 (10), S. 613–615.

Matt, Christian/Hess, Thomas/Benlian, Alexander (2015): »Digital transformation strategies«, in: Business & Information Systems Engineering 57 (5), S. 339–343, https://doi.org/10.1007/s12599-015-0401-5.

Morgner, Philipp/Mai, Christoph/Koschate-Fischer, Nicole/Freiling, Felix/Benenson, Zinaida (2020): »Security update labels: establishing economic incentives for security patching of IoT consumer products«, in: 2020 IEEE Symposium on Security and Privacy (S&P), S. 429–446, https://doi.org/10.1109/SP40000.2020.00021.

Müller, Heiko/Freytag, Johann-Christoph (2003): Problems, methods, and challenges in comprehensive data cleansing. Technical report, Humboldt-Universität zu Berlin. Online unter: http://dc-pubs.dbs.uni-leipzig.de/files/Mller2003ProblemsMethodsand.pdf, abgerufen am 20.06.2022.

Narayanan, Arvind/Shmatikov, Vitaly (2008): »Robust de-anonymization of large sparse datasets«, in: 2008 IEEE Symposium on Security and Privacy (S&P), S. 111–125, https://doi.org/10.1109/SP.2008.33.

Nikiforakis, Nick/Kapravelos, Alexandros/Joosen, Wouter/Kruegel, Christopher/Piessens, Frank/Vigna, Giovanni (2013): »Cookieless monster: Exploring the ecosystem of web-based device fingerprinting«, in: 2013 IEEE Symposium on Security and Privacy (S&P), S. 541–555.

Pugliese, Gaston/Riess, Christian/Gassmann, Freya/Benenson, Zinaida (2020): »Long-term observation on browser fingerprinting: Users' trackability and perspective«, in: Proceedings on Privacy Enhancing Technologies (2), S. 558–577, https://doi.org/10.2478/popets-2020-0041.

Rao, Ashwini/Schaub, Florian/Sadeh, Norman M./Acquisti, Alessandro/Kang, Ruogu (2016): »Expecting the unexpected: Understanding mismatched privacy expectations online«, in: Twelth Symposium on Usable Privacy and Security (SOUPS), S. 77–96.

Reeder, Robert W./Ion, Iulia/Consolvo, Sunny (2017): »152 simple steps to stay safe online: Security advice for non-tech-savvy users«, in: IEEE Security & Privacy 15 (5), S. 55–64.

Renaud, Karen/Volkamer, Melanie/Renkema-Padmos, Arne (2014): »Why doesn't Jane protect her privacy?«, in: Emiliano Cristofaro/Steven J. Murdoch (Hg.), Privacy Enhancing Technologies. 14th International Symposium PETS 2014, Amsterdam, The Netherlands, July 16–18, 2014, Proceedings, Cham: Springer, S. 244–262.

Rocher, Luc/Hendrickx, Julien M./Montjoye, Yves-Alexandre de (2019): »Estimating the success of re-identifications in incomplete datasets using generative models«, in: Nature Communications 10, Artikel Nr. 3069, https://doi.org/10.1038/s41467-019-10933-3.

Rochet, Jean-Charles/Tirole, Jean (2003): »Platform competition in two-sided markets«, in: Journal of the European Economic Association 1 (4), S. 990–1029, https://doi.org/10.1162/154247603322493212.

Sasse, M. Angela (2015): »Scaring and bullying people into security won't work«, in: IEEE Security & Privacy 13 (3), S. 80–83.

Sasse, M. Angela/Flechais, Ivan (2005): »Usable security: Why do we need it? How do we get it?«, in: Lorrie Faith Cranor/Simson Grafinkel (Hg.), Security and usability: Designing secure systems that people can use, Beijing u.a.: O'Reilly, S. 13–30.

Schneier, Bruce (2012): »When it comes to security, we're back to feudalism«, in: Wired 11. Online unter: https://www.wired.com/2012/11/feudal-security/, abgerufen am 01.06.2021.

Schneier, Bruce (2016): »Stop trying to fix the user«, in: IEEE Security & Privacy 14 (5), S. 96.

Shannon, Claude E. (1938): »A symbolic analysis of relay and switching circuits«, in: Transactions of the American Institute of Electrical Engineers 57 (12), S. 713–723, https://doi.org/10.1109/T-AIEE.1938.5057767.

Solove, Daniel J. (2006): »A taxonomy of privacy«, in: University of Pennsylvania Law Review 154, S. 477.

Spitz, Malte (2017): Daten – das Öl des 21. Jahrhunderts? Nachhaltigkeit im digitalen Zeitalter, Hamburg: Hoffmann und Campe.

Stanton, Brian/Theofanos, Mary F./Spickard Prettyman, Sandra/Furman, Susanne (2016): »Security fatigue«, in: IT Professional 18 (5), S. 26–32.

Stylianou, Konstantinos/Venturini, Jamila/Zingales, Nicolo (2015): »Protecting user privacy in the cloud: An analysis of terms of service«, in: European Journal of Law and Technology 6 (3). Online unter: https://ssrn.com/abstract=2707852, abgerufen am 01.07.2022.

Sweeney, Latanya (2002): »k-Anonymity: A model for protecting privacy«, in: International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10 (5), S. 557–570.

Syrmoudis, Emmanuel/Mager, Stefan/Kuebler-Wachendorff, Sophie/Pizzinini, Paul/Grossklags, Jens/Kranz, Johann (2021): »Data portability between online services: An empirical analysis on the effectiveness of GDPR Art. 20«, in: Proceedings on Privacy Enhancing Technologies 2021 (3), S. 351–372, https://doi.org/10.2478/popets-2021-0051.

Troncoso, Carmela/Payer, Matthias/Hubaux, Jean-Pierre/Salathe, Marcel/Larus, James/Bugnion, Edouard/Lueks, Wouter/Stadler, Theresa/Pyrgelis, Apostolos/Antonioli, Daniele/Barman, Ludovic/Chatel, Sylvain/Paterson, Kenneth/Capkun, Srdjan/Basin, David/Beutel, Jan/Jackson, Dennis/Roeschlin, Marc/Leu, Patrick/Preneel, Bart/Smart, Nigel/Abidin, Aysajan/Gürses, Seda/Veale, Michael/Cremers, Cas/Backes, Michael/Tippenhauer, Nils Ole/Binns, Reuben/Cattuto, Ciro/Barrat, Alain/Fiore, Dari/Barbosa, Manuel/Oliveira, Rui/Pereira, Jose (2020): Decentralized privacy-preserving proximity tracing. White Paper, Version vom 25.05.2020. Online unter: https://github.com/DP-3T/documents/blob/master/DP3T%20White%20Paper.pdf, abgerufen am 25.08.2021.

Turing, Alan M. (1937): »On computable numbers, with an application to the Entscheidungsproblem«, in: Proceedings of the London Mathematical Society s2-42 (1), S. 230–265, https://doi.org/10.1112/plms/s2-42.1.230.

Turow, Joseph/Hennessy, Michael/Draper, Nora (2015): The tradeoff fallacy: How marketers are misrepresenting American consumers and opening them up to exploitation, SSRN vom 10.08.2016, http://dx.doi.org/10.2139/ssrn.2820060.

Utz, Christine/Koloßa, Stephan/Holz, Thorsten/Thielbörger, Pierre (2019): »Die DSGVO als internationales Vorbild? Erste Forschungsergebnisse zu

Grundprinzipien der DSGVO und Gedanken zu ihrer Umsetzbarkeit«, in: Datenschutz und Datensicherheit – DuD 43, S. 700–705.

Volkamer, Melanie/Renaud, Karen (2013): »Mental models – general introduction and review of their application to human-centred security«, in: Marc Fischlin/Stefan Katzenbeisser (Hg.), Number theory and cryptography. Papers in honor of Johannes Buchmann on the occasion of his 60th birthday, Berlin/Heidelberg: Springer, S. 255–280.

Wagner, Isabel/Eckhoff, David (2018): »Technical privacy metrics: A systematic survey«, in: ACM Computing Surveys 51 (3), Artikel Nr. 57, S. 1–38.

Wash, Rick (2010): »Folk models of home computer security«, in: Sixth Symposium on Usable Privacy and Security (SOUPS), Artikel Nr. 11, S. 1–16, https://doi.org/10.1145/1837110.1837125.

Wash, Rick/Rader, Emilee (2015): »Too much knowledge? Security beliefs and protective behaviors among United States internet users«, in: Eleventh Symposium on Usable Privacy and Security (SOUPS), USENIX Association, S. 309–325.

Westin, Alan F. (1970): Privacy and freedom, New York: Atheneum.

Wylie, Christopher (2019): Mindf*ck. Cambridge analytica and the plot to break America, New York: Random House.

Yee, Ka-Ping (2004): »Aligning security and usability«, in: IEEE Security & Privacy 2 (5), S. 48–55.

Zuboff, Shoshana (2019): The age of surveillance capitalism: The fight for a human future at the new frontier of power, New York: PublicAffairs.