

Abschlussbericht

Verbraucherspuren in der Digitalen Welt – Wie groß und nutzbar ist der Datenschatten?

Autoren: Robert Follmer, Alexander Markowetz, Janina Belz, Jana Hölscher

Auftraggeber:

Bundesanstalt für Landwirtschaft und Ernährung (ptble) für
das Bundesministerium der Justiz und für Verbraucherschutz (BMJV)

Projektnummer: 314-06.01-2816EHV003

Projektlaufzeit: 29.12.2016 – 30.04.2018

infas Institut für angewandte
Sozialwissenschaft GmbH

Friedrich-Wilhelm-Straße 18
D-53113 Bonn
Tel. +49 (0)228/38 22-0
Fax +49 (0)228/31 00 71
info@infas.de
www.infas.de

Bericht an

Projektträger Bundesanstalt für Landwirtschaft und Ernährung (ptble)

Referat 314

z. Hd. Herrn Berg

Deichmanns Aue 29

53179 Bonn

Vorgelegt von

infas Institut für angewandte

Sozialwissenschaft GmbH

Friedrich-Wilhelm-Straße 18

53113 Bonn

Kontakt

Robert Follmer

Bereichsleitung Verkehrs- und Regionalforschung

Tel. +49 (0)228/38 22-419

Fax +49 (0)228/310071

E-Mail r.follmer@infas.de

Autoren

Robert Follmer, Alexander Markowetz, Janina Belz, Jana Hölscher

Projekt

5783

Bonn, April 2018

Fr, Bj, Ha

© infas Institut für angewandte Sozialwissenschaft GmbH
Der Inhalt dieses Berichts darf ganz oder teilweise
nur mit unserer schriftlichen Genehmigung veröffentlicht,
vervielfältigt, gedruckt oder in Informations- und
Dokumentationssystemen (information storage and
retrieval systems) gespeichert, verarbeitet oder ausgegeben
werden.

infas ist zertifiziert
nach ISO 20252 für die Markt-,
Meinungs- und Sozialforschung

ISO 20252



infas ist Mitglied im
Arbeitskreis Deutscher Markt- und
Sozialforschungsinstitute e.V. (ADM)
und ESOMAR



ESOMAR
member

Im nachfolgenden Bericht wird eine geschlechtsneutrale Sprache verwendet. Sollte aus Gründen der besseren Lesbarkeit nur die männliche Form, zum Beispiel „Nutzer“ oder „Verbraucher“ gewählt sein, sind damit auch immer die weiblichen Personen – die Nutzerinnen und Verbraucherinnen – gemeint.

Der Zugriff auf alle in diesem Bericht angegebenen Websites erfolgte zwischen Januar bis April 2018.

Inhaltsverzeichnis

1	Ziele und Aufgabenstellung des Projekts	11
1.1	Ziele	14
1.2	Aufgabenstellung	15
2	Planung und Ablauf des Projekts	16
2.1	Projektplanung	16
2.2	Projektablauf	17
3	Wissenschaftlicher und technischer Stand	19
4	Material und Methoden	23
4.1	Rekrutierung und Kommunikation mit den Probanden	24
4.1.1	Datenschutzerklärung	27
4.2	Online-Befragungen	28
4.3	App Tracy und Recall-Befragungen	29
4.4	Selbstauskunft	31
5	Darstellung der Ergebnisse	37
5.1	Exkurs: Modell des Datenschattens	37
5.1.1	Vorüberlegungen	37
5.1.2	Methoden der Datenerfassung	40
5.1.3	Weitergabe und Austausch von Daten mittels Identifikatoren	47
5.1.4	Analyse und Anreicherung von Daten	49
5.1.5	Automatisierte Entscheidung und die Rückwirkung des Datenschattens auf die Verbraucher	51
5.1.6	Größe und Ausmaß des Datenschattens	56
5.1.7	Anonymisierung und Pseudonymisierung – Datenverarbeitung ohne Datenschatten?	57
5.2	Online-Verhalten der Projektteilnehmer	59
5.2.1	Online-Aktivitäten	60
5.2.2	Datenschutzaktivitäten	67

5.3	Datensammlung und –anreicherung: Ergebnisse der Selbstauskünfte	80
5.3.1	Grundsätzliche Hinweise	80
5.3.2	Beschreibung der gelieferten Auskünfte	82
5.4	Umfang des Datenschattens und Einflussgrößen	94
5.5	Ausblick	102
6	Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse	106
7	Zusammenfassung	109
8	Gegenüberstellung der ursprünglich geplanten zu den tatsächlich erreichten Zielen	110
9	Verwendete Literatur	112
	Anhang	113

Abbildungsverzeichnis

Abbildung 1	Projektschritte und zeitlicher Ablauf	18
Abbildung 2	Erhebungsmodule	23
Abbildung 3	Ansicht des Projektflyers	24
Abbildung 4	Auszüge aus der Projektwebsite „ <i>datenspuren.infas.de</i> “	25
Abbildung 5	Teilnehmer an Projektmodulen nach Geschlecht	26
Abbildung 6	Teilnehmer an Projektmodulen nach Alter	27
Abbildung 7	Modell der Datenschattenentstehung	40
Abbildung 8	Selbstzuschreibung Online-Typ	60
Abbildung 9	Private Nutzung technischer Geräte	61
Abbildung 10	Private Online-Aktivitäten zu Unterhaltung und Kommunikation	62
Abbildung 11	Private Online-Aktivitäten zu Einkauf, Banking und Mobilität	63
Abbildung 12	Genutzte Internetseiten (Basis: Recall-Befragung)	64
Abbildung 13	Tägliche Dauer der App-Prozesse nach Inhalten	65
Abbildung 14	Profilbildung zu den Online-Aktivitäten	66
Abbildung 15	Selbst- und Bevölkerungswahrnehmung zur Datenschutzsensibilität	67
Abbildung 16	Meinungen zu den eigenen Daten im Netz	68
Abbildung 17	Browser- und Sicherheitseinstellungen bei privat genutzten Geräten	69
Abbildung 18	Art und Häufigkeit der Passwort-Vergabe	70
Abbildung 19	Durchschnittliche Anzahl der GPS-Locations am Tag	71
Abbildung 20	Profilbildung zu Datenschutzaktivitäten	72
Abbildung 21	Nachdenken und Informiertheit über den eigenen Datenschatten	74
Abbildung 22	Vorstellungen von Größe und Inhalt des Datenschattens	74
Abbildung 23	Wissen über die Verwendung der eigenen Daten	75
Abbildung 24	Einflussmöglichkeiten auf den eigenen Datenschatten	76
Abbildung 25	Kontakt zu ausgewählten Unternehmen	78

Abbildung 26	Profilbildung zur Kundenbeziehung	79
Abbildung 27	Gründe der Nicht-Teilnahme bei 17 Unternehmen	85
Abbildung 28	Schätzwert des Datenschattenindex nach Online-Aktivität	98
Abbildung 29	Schätzwert des Datenschattens nach Datenschutzaktivität	99
Abbildung 30	Schätzung des Datenschattenindex nach Zahl der Social Media-Profile	100
Abbildung 31	Ableitungen und Empfehlungen zu möglichen Regelungsansätzen	104
Abbildung 32	Präferierte Informationsquellen zum Thema Datenschatten	107
Abbildung 33	Zugeschriebene Verantwortung für den Schutz von Verbraucherdaten	108

Tabellenverzeichnis

Tabelle 1	Datenschutzaktivitäten nach Online-Aktivitäten	72
Tabelle 2	Gestellte Auskunftersuchen und Resonanz	84
Tabelle 3	Bearbeitete Auskunftersuchen und vorliegende Daten	87
Tabelle 4	Bearbeitete Ersuchen: Informationen zu Weitergabe und Speicherung	89
Tabelle 5	Bearbeitete Ersuchen: Häufig genannte Inhalte	90
Tabelle 6	Bearbeitete Ersuchen: Selten genannte Inhalte	91
Tabelle 7	Codeplan und Inhalte der inhaltsanalytischen Auswertung	92
Tabelle 8	Bildung Datenschattenindex	96

1 Ziele und Aufgabenstellung des Projekts

Im Rahmen der Digitalisierungsprozesse werden zunehmend nicht nur technische, sondern auch personenbezogene Daten mittels automatisierter Datenverarbeitungsprozesse erhoben, verarbeitet und genutzt. Durch die stärkere Verbreitung von Smartphones sowie die zunehmende Nutzung von Apps und internetfähigen Produkten und Dienstleistungen, wie beispielsweise Smart Home-Anwendungen oder Wearables, werden digitale Verbraucherspuren in Zukunft weiter wachsen. Diese Verbraucherspuren bestehen zu großen Teilen aus „sozialen Daten“, das heißt Daten über Bewegungsmuster, Verhalten, Interessen, Beziehungen zu anderen Menschen oder Einstellungen zu Produkten sowie Dienstleistungen.

„Insgesamt wächst die Menge der sozialen Daten – ob passiv oder aktiv, erforderlich oder freiwillig, präzise oder unscharf – exponentiell. (...) Die Menge der Daten, die wir im Laufe des gesamten Jahres 2000 produzieren, wird heute im Lauf eines Tages erzeugt. Bei unserer gegenwärtigen Wachstumsrate werden wir 2020 diese Menge in weniger als einer Stunde hervorbringen.“¹

Im Auftrag der Bundesanstalt für Landwirtschaft und Ernährung (BLE) als Projektträger und des Bundesministeriums der Justiz und für Verbraucherschutz (BMJV) als inhaltlicher Auftraggeber führte infas zu dieser Thematik von Januar 2017 bis April 2018 die Studie „Verbraucherspuren in der digitalen Welt – Wie groß und wie nutzbar ist der Datenschatten?“ durch.

„Big Data“ ist ein vereinfachender Begriff, der dieses rasante Wachstum widerspiegelt und seit einigen Jahren in unterschiedlichen Zusammenhängen diskutiert wird. Einerseits geht es dabei um die Potenziale bei der Auswertung und Analyse dieser Daten, andererseits um die Herausforde-

¹ Weigend 2017, S. 17.

rungen, die diese Entwicklung mit sich bringt. Oftmals stehen der Datensowie der Verbraucherschutz im Mittelpunkt der Diskussion, wenn es um die Erhebung, Anwendung und Weitergabe von Daten mit Personenbezug geht. Neben verschiedenen anderen Autoren betont der Politiker und Datenschützer Malte Spitz, dass die Rettung der informationellen Selbstbestimmung und des Datenschutzes neben dem Klimawandel und dem Umgang mit dem demografischen Wandel eine der zentralen Aufgaben und Machtfragen des 21. Jahrhunderts sein werde.² Die Frage, auf welche Art und Weise ein für die Verbraucher wirksamer Datenschutz umgesetzt werden kann und welche Rolle Politik, Unternehmen und Verbraucher dabei spielen, wird in den kommenden Jahren zu klären sein. Sicher ist jedoch, dass sich Lösungsansätze vor dem Hintergrund der rasanten technischen Entwicklung nicht am Status Quo orientieren sollten.

Es ist unbestritten, dass personenbezogene Daten für viele Unternehmen wertvoll sind. In den Medien ist in diesem Kontext häufig vom „Gold des digitalen Zeitalters“ die Rede. Bei einigen ‚Global Playern‘ aus dem Silicon Valley ist das Sammeln von Daten sogar das maßgebliche Geschäftsmodell. So kann beispielsweise Google seinen Kunden attraktive Werbeflächen passgenau für die jeweilige Zielgruppe verkaufen. Dafür „bezahlt“ jeder Nutzer, der die Google-Suchmaschine oder andere Google-Dienstleistungen nutzt, diesen Service mit seinen Daten. Laut einer Studie des Sachverständigenrates für Verbraucherfragen aus dem Jahr 2017 lag der Umsatz von Google zwischen 2010 und 2015 zwischen 38 und 45 US-Dollar pro aktivem Nutzer.³ Aber auch viele andere Unternehmen haben ein Interesse daran, möglichst umfassende Verbraucherinformationen zu sammeln, um diese für wirtschaftliche Zwecke einsetzen zu können. Dabei werden häufig unterschiedliche Verbraucherdaten zusammengeführt und zu digitalen Verbraucherprofilen aggregiert. Unter Verwendung dieser aggregierten Daten

² Vgl. Spitz 2014, S. 23.

³ Vgl. SVRV 2017, „Der Wert persönlicher Daten. Ist Datenhandel der bessere Datenschutz?“, S. 17.

können die Unternehmen mittels Algorithmen automatisierte Entscheidungen über Verbraucher treffen. Deren Einfluss wird seit dem jüngsten Datenskandal um die Nutzung von Facebook-Profilen durch die britische Big-Data-Firma Cambridge Analytica zunehmend auch in der breiten Öffentlichkeit diskutiert.

Die Genese eines Datenschattens ist durch vier Faktoren bedingt. Erstens erfordert sie eine gewisse Digitalisierung des Alltagslebens. Das heißt, Bürger interagieren mit digitalen Geräten und Medien und erzeugen somit den notwendigen Input für einen Datenschatten. Zweitens bedarf es technischer Möglichkeiten, um Daten zu sammeln, zu speichern, weiter zu verteilen und auszuwerten. Drittens müssen Unternehmen diese Möglichkeiten auch einsetzen, damit ein tatsächlicher Datenschatten entsteht. Schließlich können Verbraucher versuchen, ihren Datenschatten durch bestimmte Verhaltensweisen klein zu halten oder diesen aktiv zu managen.

Wie die einzelnen Faktoren miteinander zusammenhängen und welche Rolle sie für das Ausmaß des Datenschattens spielen, ist Untersuchungsgegenstand dieser Studie. Das genaue Ausmaß dieses sogenannten „Datenschattens“ ist bisher weder hinsichtlich seiner Quantität noch seiner Qualität bekannt. Daten- und Verbraucherschützer befürchten durch das wachsende Ausmaß dieser Datenverarbeitungsprozesse negative Konsequenzen für die Verbraucher. Auch viele Bürger sorgen sich um den Schutz ihrer Privatsphäre. Dennoch zeigt sich, dass viele Menschen, zum Beispiel für einen Preisnachlass oder im Tausch für den Service einer App, ihre persönlichen Daten relativ freigiebig preisgeben. Dieser Widerspruch zwischen der Einstellung und dem tatsächlichen Verhalten in Bezug auf den Datenschutz wurde in verschiedenen Studien belegt. Er wird auch als „Privacy Paradox“ bezeichnet.⁴

⁴ Institut der deutschen Wirtschaft, IW-Kurzberichte 57.2017

Alle Entwicklungen in diesem Bereich weisen eine hohe Dynamik auf. Die von infas durchgeführte Studie muss in diesen Kontext eines rasanten technischen Fortschritts und vor dem Hintergrund der unterschiedlichen Interessen von Verbrauchern und Unternehmen eingeordnet werden. Hinzu kommen zentrale rechtliche Veränderungen. Die neue EU-Datenschutz-Grundverordnung (DSGVO), welche im Mai 2018 endgültig in Kraft getreten ist, wird die Rahmenbedingungen für den Datenschutz in den kommenden Jahren vorgeben.

1.1 Ziele

Das Hauptziel der Studie besteht in einer Analyse, welche persönlichen Daten Verbraucher bei alltäglichen Handlungen digital hinterlassen und wie diese Daten von Unternehmen genutzt werden.

Es soll daher einerseits aufgezeigt werden, ob die Verbraucher selbst durch ihr alltägliches Verhalten zu ihrem Datenschatten beitragen und wie groß ihr diesbezüglicher Einfluss ist. In diesem Zusammenhang wird untersucht, in welchen Bereichen des Alltags und mittels welcher Technik die Daten anfallen. Die Smartphone-Nutzung bildet dabei aufgrund ihrer großen und in Zukunft weiter zunehmenden gesellschaftlichen Bedeutung⁵ einen Schwerpunkt der Untersuchung.

Auf der anderen Seite soll betrachtet werden, welche Datentypen von den Unternehmen erfasst und verarbeitet werden. Bei der Auswertung werden verschiedene Unternehmenstypen, wie beispielsweise Online-Shops, soziale Netzwerke oder Adress- und Datenhändler, berücksichtigt. Die Ergebnisse der Studie sollen das BMJV bei einer evidenzbasierten Verbraucherpolitik, etwa im Hinblick auf den Dialog mit der Wirtschaft und der Stärkung von Verbraucherrechten, unterstützen. Darüber hinaus soll die

⁵ Laut einer bitkom-Studie aus dem Jahr 2017 nutzen bereits knapp 80 Prozent der Deutschen ein Smartphone. Damit hat sich die Quote seit 2013 verdoppelt. Nähere Informationen siehe: <https://www.bitkom.org/Presse/Anhaenge-an-PIs/2017/02-Februar/Bitkom-Presskonferenz-Smartphone-Markt-Konjunktur-und-Trends-22-02-2017-Praesentation.pdf>

Studie zur Transparenz und Debatte über den „gläsernen“ Verbraucher beitragen sowie Empfehlungen für die zukünftige Datenschutzpolitik geben.

1.2 Aufgabenstellung

Exemplarisch sollte im Rahmen einer dreimonatigen Feldphase das digitale Verhalten von mindestens 100 Studienteilnehmerinnen und -teilnehmern (50 Frauen und 50 Männer), verteilt über unterschiedliche Altersgruppen ab 18 Jahren, untersucht werden. Die geplante Teilnehmerzahl konnte durch die erfolgreiche Rekrutierung sogar verdoppelt werden (siehe Kapitel 4.1). Laut Ausschreibung war zu ermitteln, welche persönlichen Daten die Studienteilnehmer bei ihrer individuellen Nutzung digitaler Angebote Dritten zugänglich machen. Die Untersuchung beschränkte sich auf die privaten Aktivitäten der Verbraucher im Alltag, wie etwa die Smartphone-Nutzung, das Surfen im Internet oder Kaufvorgänge. Repräsentative Aussagen sind mit der vorgesehenen Fallzahl nicht zuverlässig möglich. Für die Projektziele war dies aber auch nicht erforderlich. Im Vordergrund stand stattdessen die Erkennung von Mustern.

Neben der Analyse des Verbraucherverhaltens sollten 30 auf dem deutschen Markt tätige und besonders relevante Datenverarbeiter genauer untersucht werden. In diesem Zusammenhang war zu prüfen, welche Daten diesen Unternehmen zu den Studienteilnehmern vorliegen und zu welchem Zweck diese verwendet werden. Aus der Analyse des Verbraucherverhaltens einerseits und der Unternehmensdaten andererseits sollten Aussagen hinsichtlich des Verarbeitungsausmaßes und der Möglichkeiten zur Profilbildung abgeleitet werden. Schließlich sollte die aktuelle und zukünftige Relevanz dieser Persönlichkeits- und Nutzungsprofile bewertet werden.

2 Planung und Ablauf des Projekts

Im Mittelpunkt des in Kapitel 4 beschriebenen methodischen Ansatzes stehen die 200 Studienteilnehmer und deren individuelle Datenschatten. Um sich ein möglichst vollständiges Bild von den Verhaltensweisen und Einstellungen der Teilnehmer sowie deren digitalen Fußabdrücken bei den Unternehmen machen zu können, umfasst das Projekt verschiedene Bausteine. Dazu gehören zwei Online-Befragungen zum Nutzerverhalten, der Download sowie die Nutzung der für das Projekt erstellten infas-App „Tracy“, die verschiedene datenschutzrelevante Parameter auf dem Smartphone misst, und das Einholen einer Selbstauskunft im Namen der Teilnehmer bei 30 Unternehmen (siehe Kapitel 4).

2.1 Projektplanung

Nach dem Projektauftritt mit dem Auftraggeber entwickelte infas ein Gesamtkonzept, um die in der Projektausschreibung definierten Ziele zu erreichen. Die im Angebot beschriebenen Module und Arbeitsschritte wurden im Detail geplant und aufeinander abgestimmt. Dabei steht keines der Module für sich allein, sondern alle Erhebungsmethoden in Verbindung zueinander. Sie werden gemeinsam ausgewertet. Die Konkretisierung der einzelnen Projektmodule war auch im Hinblick auf die Rekrutierung der Teilnehmer wichtig, die vorab vollständig über den Erhebungsablauf informiert wurden. Die Überlegungen zum Gesamtkonzept beinhalteten neben den inhaltlichen und methodischen Fragen auch die Dokumentation einer studienspezifischen Datenschutzinformation. Da die Studienteilnehmer im Rahmen des Moduls „Selbstauskunft“ persönliche Daten an infas übergeben, die in hohem Maße schutzwürdig sind, war auch die Erarbeitung einer projektspezifischen Datenschutzerklärung Bestandteil der Erhebungsvorbereitungen. Zur Rekrutierung der Teilnehmer aus einem infas-eigenen Panel wurden zudem verschiedene Erhebungsunterlagen mit umfassenden Studieninformationen erstellt (siehe Anhang).

2.2 Projektablauf

Für die Bearbeitung des Projekts war ein Zeitbedarf von 12 Monaten ab Auftragsvergabe am 29. Dezember 2016 vorgesehen. Um die Teilnehmer vorab umfassend und zuverlässig über den Projektverlauf informieren zu können, war es wichtig, die Projektmodule und deren Ablauf im Vorfeld der für das Frühjahr 2017 geplanten Rekrutierung detailliert abzustimmen. Insbesondere die inhaltliche Komplexität der einzelnen Projektschritte und die enge Verzahnung der Projektmodule sowie der darin enthaltenen Daten und Informationen erforderten eine intensive Vorbereitungsphase. Sie wurde mit der Erarbeitung der Erhebungsinstrumente und -unterlagen Ende April abgeschlossen. Danach begann die Rekrutierung der Teilnehmer und im Anschluss die Erhebungsphase der Hauptmodule der Studie.

Während die beiden Online-Befragungen der Teilnehmer sowie der Einsatz der App „Tracy“ planmäßig durchgeführt und im August 2017 abgeschlossen werden konnten, kam es bei der Durchführung des Moduls „Selbstauskunft“ zu unvorhersehbaren Verzögerungen. Grund dafür war vor allem die geringe Resonanz der Unternehmen auf das Ersuchen um Selbstauskunft unter Anwendung des § 34 Bundesdatenschutzgesetz (BDSG) im Namen der Projektteilnehmer. Des Weiteren warfen einige angeschriebene Unternehmen die Frage auf, ob der § 34 BDSG in dieser Form im Rahmen einer Studie überhaupt angewendet werden kann. Diese Frage wurde daraufhin von verschiedenen kontaktierten Landesdatenschutzbehörden diskutiert. infas stand diesbezüglich mit der für das Institut zuständigen Landesdatenschutzbehörde NRW in intensivem fachlichen Austausch. In Kapitel 4.4 werden die Planung, Umsetzung und Diskussionen zur Selbstauskunft und die daraus resultierenden Änderungen im Projektablauf näher erläutert. Die im Herbst 2017 erfolgten Absprachen mit Vertretern des Landesdatenschutzbeauftragten für Datenschutz und Informationsfreiheit Nordrhein-Westfalen (LDI) sowie mit dem Auftraggeber führten zu maßgeblichen Verzögerungen des Projekts. Da in Absprache mit den Beteiligten beschlossen wurde, bis dato nicht teilnahmebereiten Unternehmen erneut um

Kooperation zu bitten, konnte der ursprünglich geplante zeitliche Projekt-
rahmen von 12 Monaten nicht eingehalten werden. infas stellte daher im
November 2017 einen Antrag auf viermonatige Projektverlängerung, der im
Dezember 2017 durch den Auftraggeber genehmigt wurde.

Die einzelnen Projektschritte und zeitlichen Abläufe werden in Abbildung 1
illustriert. In Kapitel 4 werden die Rekrutierung der Teilnehmer, die ver-
schiedenen Online-Instrumente, die App Tracy und das Modul „Selbstauss-
kunft“ detailliert beschrieben.

Abbildung 1 Projektschritte und zeitlicher Ablauf

Meilensteine	Jan	Feb	März	April	Mai	Juni	Juli	Aug	Sep	Okt	Nov	Dez	Jan	Feb	März	April
Projektaufstart																
Vorbereitung Instrumente																
Datenschutzkonzept																
Rekrutierung Teilnehmer																
Abgabe Zwischenbericht																
Erhebungsphase Online + App																
Selbstauskunft 1. Versuch																
Datenaufbereitung																
Selbstauskunft 2. Versuch																
Auswertungsphase																
Abschlussgespräch																
Berichterstattung																

Studie Verbraucherspuren in der Digitalen Welt – April 2018

infas

3 Wissenschaftlicher und technischer Stand

Eine umfassende Übersicht über verfügbare Technologien und Arbeiten sowie Initiativen zum Themenbereich Datenspurten und -schutz würde den Rahmen dieser Studie sprengen. Daher beschreibt dieses Kapitel grob die eingesetzte Technologie und den aktuellen Stand der Forschung.

Die Beobachtung der App-Nutzung durch Teilnehmer der vorliegenden Studie geschah auf Grundlage der entwickelten Technik im Rahmen des Mental Projekts.⁶ Die dort entwickelte App misst beispielsweise, wann ein Nutzer sein Telefon ein- oder ausschaltet, wann er eine App aufruft, oder zu einer anderen wechselt. Die App erhebt ebenfalls die Zahl der auf einem Telefon installierten Apps bzw. der verschiedenen Programmmodule. In diesem Kontext wird beispielsweise die Tastatur als eigenständige App betrachtet. Der Blickwinkel unterscheidet sich also von dem eines gewöhnlichen Nutzers, der bei der Verwendung seines Gerätes nur die Apps mit Nutzerschnittstellen wahrnimmt. Des Weiteren versucht die App, alle 30 Minuten eine Lokalisierung des Nutzers mittels GPS zu erheben. Die konkrete Position des Nutzers ist dabei im Rahmen dieser Studie irrelevant. Von Interesse ist lediglich, ob es prinzipiell möglich ist, den Nutzer entsprechend feingranular zu tracken; ob dies also anderen Apps mit entsprechenden Berechtigungen ebenfalls möglich wäre. Die Informationen über entsprechend identifizierten Phone- und App-Sessions sendet die App dann an einen Analyseserver.

Das Mental-Framework funktionierte ursprünglich lediglich auf Android-Telefonen. Im Rahmen der vorliegenden Studie wurde es auf Apples iOS portiert. Jedoch erlaubten die strengereren Sicherheitsregeln von iOS lediglich, eine Teilmenge der Funktionalität abzubilden. Insbesondere ist es auf iOS nicht möglich zu erheben, welche anderen Apps durch den Nutzer

⁶ mental.org

aktiv aufgerufen werden. Neben Menthal gibt es einige weitere Frameworks mit ähnlicher Funktionalität, beispielsweise FUNF⁷ und Aware⁸.

Die App-Nutzung wird auch für kommerzielle Zwecke gemessen, beispielsweise durch eingebaute SDKs⁹. Die entsprechenden Dienste erlauben es Anbietern, die Popularität ihrer Apps zu tracken, insbesondere mit Blick auf Konkurrenten.¹⁰ Die Sichtweise dieser Tools ist jedoch sehr speziell und im Kontext der vorliegenden Studie wenig hilfreich.

Eine der ersten Studien zu dem wahrgenommenen Datenschatten von Nutzern stammt aus dem Jahr 2007.¹¹ Ein Jahr später findet der erste Versuch statt, den tatsächlichen Datenschatten von Nutzern zu quantifizieren – wenn auch aus einer eher technischen Sicht, die das Gesamtvolumen der weltweit verfügbaren Daten betrachtet.¹² In Deutschland versuchte der Grünen-Politiker Malte Spitz im Jahr 2014 den eigenen Datenschatten zu ermitteln, indem er breit angelegt Anträge auf Datenselbstauskunft stellte. Obwohl nicht alle Firmen Daten sendeten und nicht vollständig Auskunft gaben, gelang es ihm, seinen Datenschatten zumindest grob zu umreißen.¹³

Unter *selbstauskunft.net* finden Nutzer Informationen rund um das Erstellen von Anträgen auf Datenselbstauskunft. Das kommerzielle Projekt liefert Vordrucke, des Weiteren können Nutzer ihre Erfahrungen mit den verschiedenen Diensten schildern. Allerdings liegen die Bewertungen hauptsächlich für Wirtschaftsauskunfteien und selten für Internet-Giganten vor. Die Nutzung der Auskunftsmöglichkeit nach § 34 BDSG im Rahmen eines vergleichbaren wissenschaftlichen Projekts für eine größere Personengruppe ist nach unserer Kenntnis bislang nicht erfolgt. In diesem Zusammen-

⁷ <http://funf.codysumter.com/>

⁸ <http://www.awareframework.com/>

⁹ Ein Software Development Kit (SDK) ist eine Sammlung von Programmierwerkzeugen und Programmbibliotheken, die zur Entwicklung von Software dient. Es unterstützt Softwareentwickler, darauf basierende Anwendungen zu erstellen. In der Regel gibt es zu jeder Programmiersprache ein Software Development Kit. Bei interpretierten Sprachen kann dieses jedoch identisch mit der Laufzeitumgebung sein.

¹⁰ www.appannie.com & [flurry.com](http://www.flurry.com)

¹¹ <http://www.pewinternet.org/2007/12/16/digital-footprints/>

¹² https://readwrite.com/2008/03/24/new_tool_calculates_your_digital_footprint/

¹³ Vgl. Spitz 2014, S. 23.

hang stellt der in diesem Projekt vorgenommene Versuch in seiner Form nach unserer Kenntnis Neuland dar.

Die Literatur zu § 34 BDSG ist so vielfältig, dass sie hier nicht aufbereitet werden kann. Beispielhaft wird auf Hermstrüwer¹⁴ 2016 verwiesen, wo auch der Diskussionsstand aufbereitet wird.

Parallel zur vorliegenden Studie versuchen derzeit mehrere Ansätze durch breit gestreute Selbstauskünfte, den Datenschatten von Nutzern zu bestimmen, bzw. durch dessen Analyse die Scoring-Methoden der Firmen zu entschlüsseln.¹⁵ Neu dabei ist das Konzept der sogenannten Datenspende, bei der Nutzer ihre Daten für das Gemeinwohl zur Verfügung stellen, insbesondere auch Daten aus Selbstauskünften. Weiterhin gibt es einige journalistische Versuche. Ein Beispiel ist ein Beitrag zu Amazons Sprachassistenten Alexa¹⁶, der auch Messungen der Datenströme vorstellt. Ähnliches unternimmt ein Beitrag zu von Amazon erhaltenen Auskünften¹⁷. Ebenso zu nennen sind verschiedene Versuche im Verbraucherschutzbereich, etwa im Automobilsektor, der in dieser Studie nicht betrachtet wird.¹⁸

Das SensorenResidenz-Programm der OpenDataCity untersucht den Datenschatten-Beitrag von Smart-Home-Technologien. Dazu misst es in einer Modellwohnung von über 100 smarten Geräten ausgehende Datenströme.¹⁹

Auf amiunique.org können Nutzer überprüfen, ob ihre Browsereinstellungen in der Tat so einzigartig sind, dass man sie damit im Netz tracken kann. Die Webseite erklärt des Weiteren die verschiedenen genutzten Parameter, und wie stark sie zu einer eindeutigen Identifizierung des Nutzers beitragen.

Auch, wenn die vorliegende Studie – wie noch gezeigt wird – zu der Einschätzung gelangt, dass das Vermeiden der Entstehung eines Datenschat-

¹⁴ Vgl. Hermstrüwer, 2016

¹⁵ openschufa.de

¹⁶ <https://www.zeit.de/2018/14/amazon-alexa-sprachassistentin-persoennliche-daten-test>

¹⁷ <http://www.spiegel.de/netzwelt/web/amazon-experiment-was-der-konzern-mit-jedem-klick-erfaehrt-a-1205079.html>

¹⁸ <https://www.test.de/Connected-Cars-Die-Apps-der-Autohersteller-sind-Datenschnueffler-5231839-0/>

¹⁹ <https://opendatacity.de/project/wie-das-smarthome-unser-leben-aufzeichnet/>

tens ebenso wie dessen aktives Management nur bedingt möglich sind, bedeutet dies nicht, dass Nutzer es nicht zumindest versuchen sollten. Digitalcourage gibt hierzu Anleitung zur sogenannten „digitalen Selbstverteidigung“.²⁰

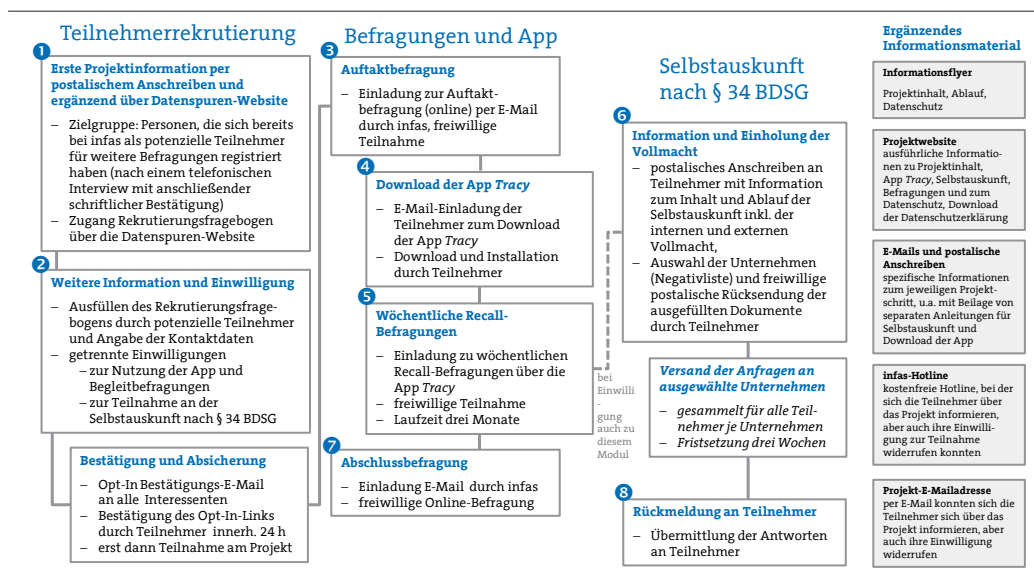
Eine umfassende Beschreibung der Entstehung eines Datenschattens finden Nutzer auf myshadow.org. Dieses Online-Projekt ermöglicht es, den eigenen Datenschatten zu „berechnen“, wobei jedoch nur ein grobes Gefährdungspotential anhand der Anzahl der verschiedenen genutzten Online-Medien erstellt wird.

²⁰ <https://digitalcourage.de/support/digitale-selbstverteidigung>

4 Material und Methoden

Die Anzahl der Datenquellen, der Daten sammelnden Unternehmen und der Verbraucher, die Datenspuren hinterlassen, ist schon heute so umfangreich und komplex, dass nur ein Ausschnitt des Status quo erhoben werden kann. In diesem Projekt soll ein zuverlässiger Eindruck über die Entstehung, Dichte und Qualität des entstehenden Datenschattens vermittelt werden – ein Anspruch auf Vollständigkeit kann dabei nicht erhoben werden. Das zu Grunde liegende Modell des Datenschattens wird in Kapitel 5.1 in Form eines Exkurses erläutert. Der methodische Ansatz konzentriert sich auf das Verhalten und die Einstellungen der rekrutierten Studienteilnehmer einerseits und auf ihren Datenschatten bei 30 ausgewählten Unternehmen andererseits. Die Studie zielt dabei nicht auf repräsentative Aussagen, sondern auf die Erklärung von Zusammenhängen, die Ableitung von Schlussfolgerungen zu zukünftigen Entwicklungen und die sich daraus ergebenden Empfehlungen. Die verschiedenen Module (siehe Abbildung 2) werden ab Kapitel 4.2 erläutert. Zunächst wird im Folgenden die Rekrutierung der Teilnehmer beschrieben.

Abbildung 2 Erhebungsmodule



Studie Verbraucherspuren in der Digitalen Welt – April 2018

infas

4.1 Rekrutierung und Kommunikation mit den Probanden

Vor Beginn der eigentlichen Feldphase wurden im Rahmen der Rekrutierung verschiedene Informationsangebote für die potenziellen Teilnehmer entwickelt. Dazu gehörten eine Projektwebsite²¹, ein Flyer sowie ein Anschreiben (siehe Anhang). Für die Kommunikation mit den Teilnehmern erhielt das Projekt den neutralen Studententitel „Datenspuren“. Für den Webaufttritt und die Erhebungsunterlagen wurde ein einheitliches Design entworfen. Im Mai 2017 wurden Teilnehmer eines infas-Panels, die sich projektübergreifend zu weiteren Kontaktaufnahmen von infas bereit erklärt hatten, schriftlich kontaktiert, über das Projekt „Datenspuren“ informiert und um freiwillige Teilnahme gebeten. Sie erhielten das Anschreiben, den Flyer mit den wichtigsten Informationen sowie eine Datenschutzerklärung.

Abbildung 3 Ansicht des Projektflyers



Zudem wurde das Projekt auf der Facebook-Seite des Mental Projekts²² der Universität Bonn vorgestellt. So konnten interessierte Leser ebenfalls

²¹ www.datenspuren.infas.de

²² <https://menthal.org/>

auf die Datenspurten-Studie aufmerksam werden und sich zum Projekt anmelden.

Die Anmeldung erfolgte in beiden Rekrutierungsvarianten über die Website der Studie. Diese enthielt auf der Startseite einen Überblick zum Hintergrund des Projekts sowie zum Ablauf und zu den Teilnahmevoraussetzungen. Ebenso wurden die Teilnehmer über die Incentivierung (gestaffelt, bis zu 200 € für die Teilnahme an allen Projektmodulen) informiert. In weiteren Menüpunkten wurden detaillierte Informationen zu den einzelnen Projektmodulen (App Tracy, Selbstauskunft, Online-Befragungen) und zum Datenschutz hinterlegt. Außerdem konnten auf der Projektwebsite der Info-Flyer und die Datenschutzerklärung heruntergeladen werden.

Abbildung 4 Auszüge aus der Projektwebsite „*datenspurten.infas.de*“



Der Anmeldeprozess umfasste zunächst einen kurzen Rekrutierungsfragebogen mit Fragen zur Smartphone-Nutzung und dem Online-Verhalten. Am Ende des Rekrutierungsfragebogens wurden die Kontaktdaten der Teilnehmer erfasst. Das verwendete Double-Opt-In-Verfahren sah eine zusätzliche E-Mail-Bestätigung zur Erklärung der Teilnahmebereitschaft vor. Ansonsten wurden die Daten binnen 24 Stunden wieder gelöscht. Für Rückfragen

waren im gesamten Projektverlauf geschulte infas-Mitarbeiter unter einer kostenlosen Hotline oder einer studienspezifischen E-Mail-Adresse erreichbar.

Um die angestrebte Fallzahl von 100 Personen – darunter 50 Männer und 50 Frauen sowie unterschiedliche Altersklassen – sicher zu erreichen, wurde zunächst eine größere Teilnehmeranzahl rekrutiert. So bestand die Möglichkeit, Teil- oder Komplettausfälle während der Projektlaufzeit zu kompensieren. Zudem wurde eine Übersteuerung bei der Teilnehmeranzahl geplant, um eine ausreichend große Auswertungsbasis beim Modul „Selbstauskunft“ zu haben. Hier war zu Projektbeginn von einer geringeren Teilnahmebereitschaft auszugehen, da die Teilnehmer dieses Modul optional wählen konnten.

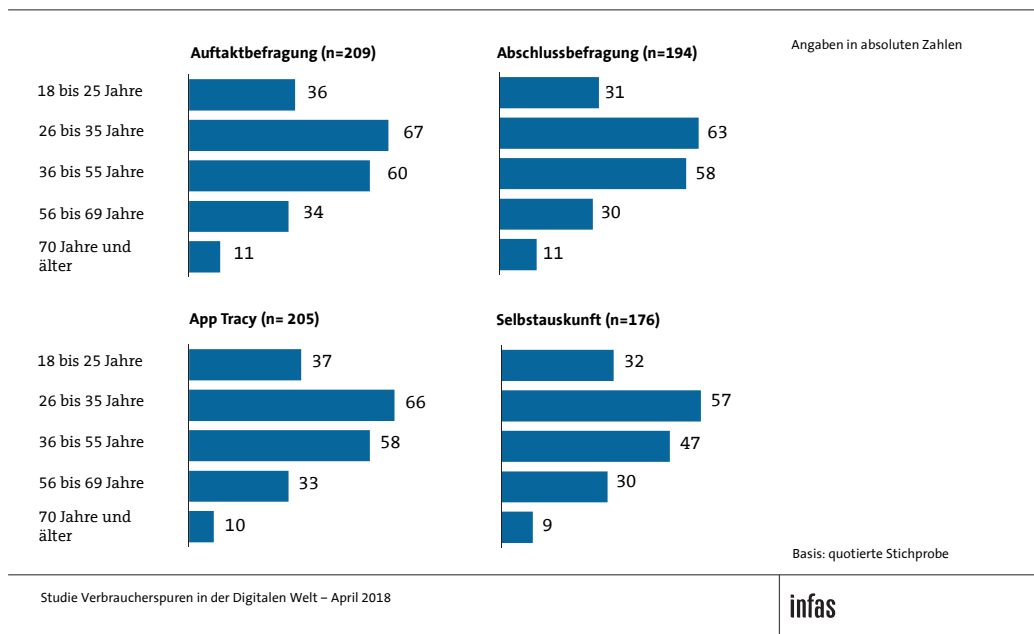
Abbildung 5 Teilnehmer an Projektmodulen nach Geschlecht



Abbildung 5 zeigt, dass die Anforderung von 100 Personen in allen Modulen deutlich übertroffen werden konnte. Das Geschlechterverhältnis in der Studie ist sehr ausgeglichen. Bei den Altersgruppen (Abbildung 6) liegen größere Schwankungen vor. So sind die beiden mittleren Altersgruppen von 26 bis 55 Jahre mit einem Anteil von etwa 60 Prozent

der Teilnehmer in allen Modulen am stärksten besetzt. Die Altersgruppen von 18 bis 25 Jahre sowie von 56 bis 69 Jahre sind mit jeweils etwa 30 Teilnehmern etwa gleichstark besetzt. Die Altersgruppe ab 70 Jahre ist erwartungsgemäß am schwächsten besetzt.

Abbildung 6 Teilnehmer an Projektmodulen nach Alter



Positiv hervorzuheben ist, dass über den gesamten Projektzeitraum von über drei Monaten kaum Teilnehmersausfälle zu beklagen waren. Fast alle Teilnehmer der Auftaktbefragung luden sich auch die App Tracy herunter. Selbst am optionalen Projektmodul der Selbstauskunft nahmen über 80 Prozent der Stichprobe teil. An der Abschlussbefragung beteiligten sich lediglich sieben Prozent weniger als an der Auftaktbefragung. Die sehr gute Teilnahmequote ist höchstwahrscheinlich zum einen auf die relativ hohe Incentivierung und zum anderen auf die regelmäßige E-Mail-Kommunikation mit den Teilnehmern zurückzuführen.

4.1.1 Datenschutzerklärung

Im Rahmen der einzelnen Projektschritte wurden umfassende persönliche Daten der Teilnehmer erhoben.

Den Teilnehmern wurde im Rahmen einer Datenschutzerklärung (siehe Anhang) zugesichert, dass

- die Teilnahme an dem Projekt freiwillig erfolgt und der Nutzung der Daten jederzeit widersprochen werden kann,
- alle ihre persönlichen Daten zum Projektende bei infas gelöscht werden,
- ihre Befragungsangaben nur in anonymisierter Form ausgewertet und berichtet werden und
- die Daten zu keinem Zeitpunkt an Dritte weitergegeben werden.

Die Datenschutzerklärung wurde den Teilnehmern schriftlich zugesendet und war zudem während der gesamten Projektlaufzeit über die Website abrufbar.

4.2 Online-Befragungen

Die Projektteilnahme beinhaltete zwei etwa 15-minütige Online-Befragungen jeweils zu Beginn und zum Abschluss der Erhebungsphase. Nach der Rekrutierung der Teilnehmer wurden diese Anfang Juni 2017 zunächst zur Auftaktbefragung (siehe Anhang) eingeladen. Die Auftaktbefragung hatte zum Ziel, das allgemeine Nutzerverhalten abzufragen sowie die Einstellungen der Teilnehmer zum Thema Datenschutz zu erfassen.

Weitere Themenbereiche aus der Auftaktbefragung waren unter anderem:

- Häufigkeit der Online-Nutzung bei verschiedenen privaten Aktivitäten,
- Nutzung verschlüsselter und nicht-verschlüsselter W-LAN Netzwerke,
- Privatsphäre-Einstellungen an unterschiedlichen Geräten,
- Passwort-Vergabe,
- Anzahl von Benutzerkonten und Kundenkarten,
- Kundenbeziehung zu verschiedenen Unternehmen und
- Einstellungen zum Thema Datenschutz.

Im Oktober wurde zum Abschluss des Erhebungszeitraums eine weitere etwa 15-minütige Online-Befragung (siehe Anhang) durchgeführt. Hierbei

sollten die Teilnehmer sowohl ihre unterschiedlichen Erfahrungen aus den drei Erhebungsmonaten schildern als auch die Größe und Qualität des eigenen Datenschattens einschätzen.

Weitere Themenbereiche aus der Abschlussbefragung waren unter anderem:

- Beschäftigung mit dem Thema Datenschatten,
- Einschätzung über Einflussmöglichkeit auf den Datenschatten,
- Maßnahmen zur Reduktion des Datenschattens,
- Informationsstand und genutzte Informationsquellen zum Datenschatten sowie
- Vertrauen und Erwartungen in Institutionen.

4.3 App Tracy und Recall-Befragungen

Während die Ergebnisse der Online-Befragungen die subjektiven Einschätzungen und Bewertungen der Teilnehmer widerspiegeln, dient die App Tracy der Messung des tatsächlichen Onlineverhaltens auf dem Smartphone. Dieses belegte erstmalig 2016 den Spitzenplatz bei den meistgenutzten Geräten für den Internetzugang.²³ Die Messung der Smartphone-Aktivität ist im Hinblick auf das Thema Datenspurten von großer Relevanz, da die Bedeutung des Smartphones inzwischen weit über die reine Kommunikation hinausgeht. Vor allem Apps dienen inzwischen in den unterschiedlichsten Bereichen als „Alltagshelfer“. Die beiden Marktführer iOS und Google Play erreichten im Jahr 2017 Downloadzahlen von fast 26 Milliarden weltweit – ein Anstieg von acht Prozent im Vergleich zum Vorjahr.²⁴ Bei der Installation und der Anwendung von Apps können sowohl durch den App-Betreiber als auch den App-Entwickler persönliche Daten erhoben und verarbeitet werden. Dies sind zum einen Daten, die der Nutzer bewusst eingibt,

²³ <http://www.zeit.de/digital/internet/2016-10/ard-zdf-onlinestudie-internetnutzer-deutschland-zuwachs>

²⁴ <https://www.appannie.com/de/insights/market-data/global-app-downloads-consumer-spend-hit-q3-2017-recap/>

andere werden wiederum automatisch erhoben. Im Kapitel 5.2 werden die Ergebnisse zu den spezifischen Verbraucherspurten bei der App-Nutzung berichtet. Auch durch andere Smartphone-Aktivitäten (z. B. die Nutzung nicht gesicherter W-LAN-Netzwerke, die Nutzung des Internetbrowsers) entstehen Datenspurten, die Rückschlüsse auf den Nutzer zulassen.

Die App Tracy wurde den Studienteilnehmern Anfang Juni 2017 als Android- oder iOS-Version zum Herunterladen zur Verfügung gestellt. Von den insgesamt 205 Teilnehmern dieses Moduls nutzten 141 Personen ein Smartphone mit Android-Betriebssystem und 64 ein iPhone von Apple.

Die App Tracy zeichnete verschiedene Parameter der Smartphone-Nutzung auf. Dazu gehörten:

- Anzahl der installierten und genutzten Apps,
- Anzahl der App-Sessions am Tag,
- Dauer der Telefonnutzung,
- Nutzung von verschlüsselten und nicht-verschlüsselten W-LAN-Netzwerken und
- GPS-Signale (alle 30 Minuten).

Des Weiteren konnte Tracy – abhängig vom Betriebssystem und Hersteller – verschiedene IDs des Smartphones (IMEI, Mac-Adresse, Ad-ID) auslesen. Anhand der getrackten Merkmale kann die Aktivität der Nutzer abgeleitet werden. Daraus lassen sich Rückschlüsse auf die Datenspurten der Teilnehmer ziehen.

Da neben der Smartphone-Nutzung auch die Onlineaktivitäten auf anderen Geräten der Teilnehmer eine Bedeutung für den Datenschatten haben, diese aber im Rahmen des Projekts nicht objektiv gemessen werden konnten, wurde das Konzept von wöchentlichen **Recall-Befragungen** angewendet. Alle acht Tage wurde den Teilnehmern automatisch über die App ein kurzer Fragebogen angezeigt. Dieser konnte direkt auf dem Smartphone bearbeitet werden. Damit blieb der Aufwand für die Teilnehmer überschaubar und das Risiko vorzeitiger Abbrüche wurde minimiert. Die regelmäßige

Streuung über zwölf Messzeitpunkte an unterschiedlichen Wochentagen hatte außerdem den Vorteil, dass alle Wochentage abgedeckt wurden und besondere Anlässe, zum Beispiel Urlaub, keinen entscheidenden Einfluss auf die Erhebung hatten. Die Recall-Fragen dienten vor allem dazu, Nutzungsmuster auf dem PC oder Tablet abschätzen zu können.

4.4 Selbstauskunft

Während die Entstehung des Datenschattens durch die Befragung der Teilnehmer und die App Tracy abgebildet werden sollte, war das Modul „Selbstauskunft“ vorgesehen, um die Quantität und Qualität des Datenschattens auf Unternehmensseite abzuschätzen. Daher sah der Projektplan vor, für mindestens 100 Probanden bei 30 ausgewählten Unternehmen das Recht auf informationelle Selbstauskunft nach § 34 BDSG in Anspruch zu nehmen.

Das Modul „Selbstauskunft“ verfolgte nicht das Ziel, einzelne Unternehmen im Hinblick auf die Beantwortung der Auskünfte oder die Reaktionsgeschwindigkeit zu bewerten, sondern eine möglichst gute Abschätzung der Art und Größe des Datenschattens bei den Unternehmen vornehmen zu können, die stellvertretend für bestimmte Branchen stehen. Gemäß Beschreibung sollten insbesondere die Bereiche Online-Shops, Smart-TV-Anbieter, soziale Netzwerke, Auskunftsteien, Zahlungsdienstleister, Versicherungen, Logistikunternehmen und Adress- sowie Datenhändler beleuchtet werden. Vorgegeben waren die Unternehmen Acxiom, Alphabet (Google), Amazon, Apple, Facebook, Microsoft und SCHUFA.

Versicherungen wurden in Absprache mit dem Auftraggeber in diesem Projektmodul ausgeklammert, da die Bandbreite der möglichen Versicherungen in Relation zur Teilnehmerzahl zu groß ist und sich die Personendaten bei Versicherungen in der Regel nur auf Kundenbeziehungen konzentrieren.

Die Auswahl der weiteren Unternehmen wurde mit dem Auftraggeber abgestimmt und umfasste die nachfolgenden Unternehmen:

1. Acxiom Deutschland GmbH
2. AdAudience GmbH
3. Amazon EU SARL (Niederlassung Deutschland)
4. Apple Retail Germany B.V. & Co. KG
5. AZ Direct GmbH
6. BILD GmbH & Co. KG
7. comScore, Inc.
8. Creditreform Boniversum GmbH
9. Deutsche Bahn AG
10. Deutsche Post DHL Group
11. Deutsche Post Direkt GmbH
12. DriveNow GmbH & Co. KG
13. eBay International AG
14. Facebook Germany GmbH
15. Google Germany GmbH
16. IKEA Deutschland GmbH & Co. KG
17. informa HIS GmbH
18. Microsoft Deutschland GmbH
19. Miles & More GmbH
20. Payback GmbH
21. PE Digital GmbH (Parship)
22. Samsung Electronics GmbH
23. Schober Information Group Deutschland GmbH
24. Schufa Holding AG
25. Semasio GmbH
26. Telekom Deutschland GmbH
27. Twitter Germany GmbH
28. United Internet AG (1&1)
29. WhatsApp Inc.
30. Zalando SE

Alle Unternehmen bis auf WhatsApp Inc. sind in Deutschland ansässig oder mit einer entsprechenden Niederlassung vertreten. Bei WhatsApp wurde die Unternehmensadresse in den USA kontaktiert.

Unternehmen können im Rahmen der Speicherung von Personendaten unterschiedliche Daten- bzw. Personenschlüssel verwenden. Bei einem Antrag auf Datenselbstauskunft müssen vom Antragsteller mehrere dieser eindeutigen Personenschlüssel angegeben werden, um die vom Unternehmen gespeicherten Daten zur eigenen Person zu erhalten. Traditionell besteht ein solcher Schlüssel aus Name, Anschrift und Geburtsdatum des jeweiligen Bürgers. In der digitalen Welt haben diese analogen Identifikatoren jedoch beinahe keine Verbreitung. Stattdessen verwenden Unternehmen digitale Schlüssel wie z. B. Handynummern, E-Mail-Adressen, Facebook-IDs oder eigens geschaffene Ad-IDs. Insbesondere in der Online-Werbeindustrie werden fast ausschließlich digitale Schlüssel oder davon abgeleitete Pseudonyme verwendet. Daher wurden im Rahmen des Projekts die üblichen Angaben einer Selbstauskunft durch zusätzliche potentielle Identifikationsmerkmale wie etwa Mobilfunknummern, E-Mail-Adressen und Accountkennungen²⁵ der Teilnehmer erweitert. Des Weiteren wurden explizit auch solche persönlichen Daten erfragt, die bei den Unternehmen unter abgeleiteten Pseudonymen (Hash-Schlüsseln) gespeichert wurden. Während die „klassischen“ Angaben für die Teilnahme an diesem Projektmodul notwendig waren, konnten die Probanden selbst über die Angabe der zusätzlich von infas ausgewählten Kennziffern entscheiden. Die Teilnehmer konnten darüber hinaus festlegen, für welche Unternehmen sie infas nicht zur Einholung der Selbstauskünfte bevollmächtigen wollten.

In Vorbereitung zum Projektmodul wurden unter juristischer Beratung verschiedene Dokumente zur postalischen Kommunikation mit den Teilnehmern und den Unternehmen erarbeitet (siehe Anhang):

²⁵ Dabei handelte es sich um die Facebook-ID, Apple-ID, Google-ID, IMEI des Smartphones, MAC-Adresse des Smartphones und AD-ID des Google-Accounts.

- eine ausführliche interne Vollmacht der Befragten gegenüber infas, innerhalb derer die Unternehmen individuell ausgewählt und diejenigen personenbezogenen Daten angegeben werden, die in das Auskunftersuchen aufgenommen werden sollen,
- eine externe Vollmacht zur Vorlage bei den jeweiligen Unternehmen und
- ein personalisiertes Anschreiben an die Unternehmen, das infas für die Befragten gesammelt an das jeweilige Unternehmen versendet hat.

Alle Studienteilnehmer wurden schriftlich ausführlich über das Vorgehen aufgeklärt und konnten der Teilnahme an diesem Modul durch den postalischen Rückversand der beiden Vollmachten zustimmen.

Durch die erforderliche schriftliche Kommunikation mit den Teilnehmern sowie die intensive Vorbereitung der individualisierten Anschreiben und Vollmachten konnte der postalische Versand der Anfragen an die Unternehmen erst Mitte August 2017 erfolgen. infas erbat Auskünfte im Namen derjenigen Studienteilnehmer, die sich zuvor zur Teilnahme an diesem Modul bereit erklärt hatten. Insgesamt beteiligten sich 176 Personen an diesem Projektmodul. Dies überschritt die geschätzte Teilnahmebereitschaft deutlich. Es wurden 4.959 Anfragen versendet (Anzahl der Personen x Anzahl der jeweils einbezogenen Unternehmen). Die Anfragen der Teilnehmer mit den entsprechenden Vollmachten wurden zwar separat erstellt und kuvertiert, aber jeweils gemeinsam an das jeweilige Unternehmen geschickt. Damit wurden der Umfang der Anfragen und auch der Projekthintergrund deutlich. Im Anschreiben wurde ebenfalls offengelegt, dass die Anfragen im Rahmen eines Forschungsprojekts eingeholt werden sollten. Den Unternehmen wurde eine dreiwöchige Frist zur Beantwortung der Anfragen eingeräumt.

Lediglich 13 der insgesamt 30 Unternehmen kamen der Aufforderung zur Bereitstellung der Selbstauskünfte nach, darunter eines erst nach erfolgter schriftlicher Nachfassaktion. Diese Rückmeldung erfolgte bei manchen Unternehmen in Form einer postalischen Datenlieferung an infas oder an

die Teilnehmer selbst. In anderen Fällen informierten die Unternehmen infas oder die Teilnehmer über das Nichtvorhandensein der angefragten Daten in ihrem Unternehmen. Den Unternehmen wurde es explizit überlassen, die Auskünfte nach eigener Bewertung auch verschlüsselt zu versenden.

Die geringe positive Resonanz wurde nach Abschluss der gesetzten Frist mit dem Auftraggeber im Hinblick auf das weitere Vorgehen diskutiert. Parallel befand sich infas im Austausch mit Vertretern des Landesbeauftragten für Datenschutz und Informationsfreiheit Nordrhein-Westfalen (LDI NRW). Nachdem infas dort Mitte September 2017 eine Stellungnahme zu unterschiedlichen, das Projekt betreffenden, datenschutzrechtlichen Fragen eingereicht hatte, konnte ein persönliches Treffen mit Vertretern des LDI am 20.12.2017 stattfinden. Aufgrund dieser Abläufe und der zu diesem Zeitpunkt ausstehenden Entscheidungen zur weiteren Vorgehensweise wurde das Projekt, wie in Kapitel 2.2 bereits thematisiert, bis Ende April 2018 verlängert.

In Absprache mit dem Auftraggeber und dem LDI setzte infas im Januar 2018 ein weiteres Schreiben an die Unternehmen auf (siehe Anhang). Darin wurde noch einmal näher auf den Projekthintergrund, die Einwilligung der Teilnehmer, die Bevollmächtigung von infas und die Datenübermittlung eingegangen.²⁶ Auf eine juristische Argumentation sowie die Androhung von Verwarngeldern bei Nicht-Kooperation wurde bewusst verzichtet, auf die grundsätzliche Abstimmung des Verfahrens mit dem LDI NRW jedoch verwiesen. Der Versand des zweiten Briefes erfolgte am 19.01.2017 erneut mit einer dreiwöchigen Frist an 13 Unternehmen, bei denen eine Kooperation noch möglich erschien. Nur eines der erneut angeschriebenen Unternehmen versendete daraufhin Selbstauskünfte an die Teilnehmer, sodass

²⁶ Ebenfalls Bestandteil des Schreibens war die Bezugnahme auf § 34 BDSG. Obwohl der Artikel nicht an einen bestimmten Anlass oder eine bestimmte Verwendungssituation geknüpft ist, ist nicht explizit geklärt, ob das Auskunftsrecht nach § 34 BDSG in der Art im Rahmen eines Forschungsprojekts eingesetzt werden kann. Zumal handelt es sich bei der Nutzung in diesem Projekt vermutlich um einen Präzedenzfall. In Kapitel 5.4 wird noch einmal auf die Verwendung des Auskunftsrechts und die unterschiedlichen Argumentationen eingegangen.

insgesamt 13 der 30 Unternehmen Informationen zu vorhandenen oder nicht vorhandenen Daten lieferten. Auf die Inhalte der gelieferten Selbstauskünfte wird in Kapitel 5.4 ausführlich eingegangen.

5 Darstellung der Ergebnisse

Der folgende Ergebnisteil hat vier Funktionen. Diese werden auch in der Gliederung abgebildet. Zunächst wird in Kapitel 5.1 in der Form eines Exkurses ein theoretisches Modell des Datenschattens entwickelt und anhand von Beispielen verdeutlicht. Dabei werden sowohl die Methoden der Datenerfassung, die Möglichkeiten der Datenweitergabe und -anreicherung und mögliche Rückwirkungen auf die Verbraucher beleuchtet. Daran schließt sich in Kapitel 5.2 die Vorstellung der Befragungsergebnisse sowie der technischen Messung über die App Tracy an. Diese Resultate zeigen das Ausmaß von Aktivitäten der Probanden auf, die datenschattenwirksam sind. Kapitel 5.3 stellt die Ergebnisse der gemäß § 34 BDSG für die Probanden eingeholten Selbstauskünfte bei 30 Unternehmen vor. Dies umfasst auch eine inhaltliche Bewertung der erzielten Qualität. Da diese Resultate aus Gründen der überwiegenden Nicht-Kooperation der Unternehmen bzw. aufgrund vermutlich oft nicht vollständiger Auskünfte keine valide Grundlage für eine Ableitung des tatsächlichen Datenschattens darstellen, wird im anschließenden Kapitel 5.4 ersatzweise der Versuch unternommen, den tatsächlichen Datenschatten anhand der Befragungsergebnisse sowie der technischen Messung zu schätzen. Hierzu wird ein **Datenschattenindex** entwickelt und vorgestellt.

5.1 Exkurs: Modell des Datenschattens

5.1.1 Vorüberlegungen

Der Datenschatten eines Verbrauchers definiert sich als Gesamtheit aller über einen Nutzer erhobenen Daten, verteilt über die speichernden Institutionen. Darüber hinaus streut er über Institutionen, die selber keine Daten speichern, jedoch über gespeicherte Daten bei anderen Institutionen jederzeit verfügen können. Im Rahmen dieser Studie ist es notwendig, vier verschiedene Formen des Datenschattens zu differenzieren: (a) den heute mög-

lichen, (b) den tatsächlichen, (c) den zukünftig möglichen und (d) den wahrgenommenen Datenschatten.

Der heute mögliche Datenschatten eines Menschen ist bedingt durch die technischen Möglichkeiten, Daten zu sammeln, zu übermitteln und zu speichern. Diese waren bis vor kurzer Zeit noch sehr begrenzt. Erst durch mobile Endgeräte und Mobilfunk werden menschliche Handlungen zunehmend digital erfassbar. Über leistungsfähige Glasfasernetze sind diese immensen Datenmengen ohne Grenzkosten weltweit verschiebbar beziehungsweise mit dritten Unternehmen teilbar. Gleichzeitig sind die Kosten für die Datenspeicherung rapide gefallen. Dies macht es möglich, menschliches Handeln umfassend zu dokumentieren, die Informationen mit anderen Datensätzen zusammenzuführen und gemeinsam zu analysieren.

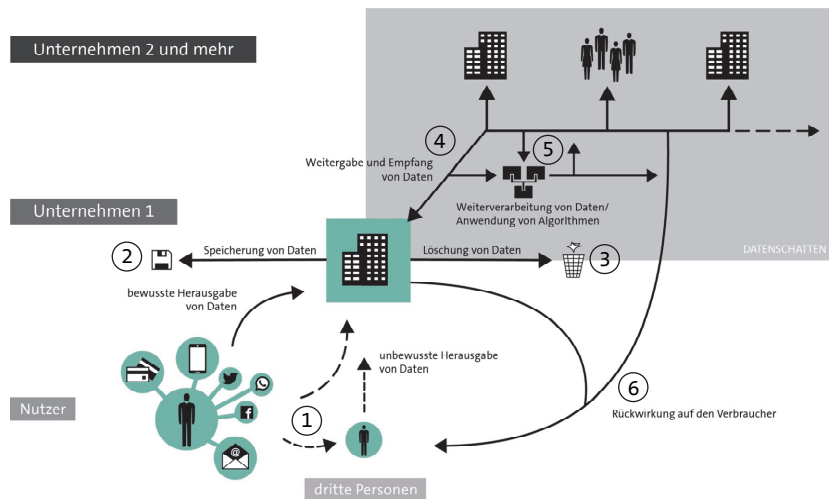
Der tatsächliche individuelle Datenschatten ergibt sich aus dem Grad, zu dem Unternehmen die derzeitigen technischen Möglichkeiten umsetzen. Diese faktische Sammlung und Auswertung von personenbezogenen Daten hinkt den technischen Möglichkeiten mitunter deutlich hinterher. Oft fehlt es Firmen an Kreativität und Know-how, um diese Technologien umfassend einzusetzen. Auch existierende IT-Landschaften sind aus ihrer Historie heraus oft so komplex, dass sich neue Module nicht einfach integrieren lassen. Es wäre jedoch irreführend, sich langfristig auf dem mangelnden Einsatz neuer Technologien zu verlassen. In Zukunft wird mit hoher Wahrscheinlichkeit jede neue Technologie umfassend eingesetzt werden. Unternehmen wie Google oder Facebook geben hier bereits die Richtung vor. Dadurch wird sich der tatsächliche Datenschatten über kurz oder lang dem möglichen Datenschatten angleichen. Insbesondere in Zeiträumen von zehn Jahren bis zwanzig Jahren, in denen Gesetzgeber (voraus-)denken müssen, ist daher der technisch mögliche Datenschatten maßgebend.

Darüber hinaus gibt es noch einen zukünftig möglichen Datenschatten, der durch Technologien erhoben und übermittelt wird, die aktuell noch nicht zur Verfügung stehen. Bei der aktuellen Innovationsgeschwindigkeit fallen viele dieser technischen Neuerungen jedoch in einen für den Gesetzgeber

maßgeblichen Zeitrahmen. Es ist daher notwendig, auch zukünftig technische Möglichkeiten in gesetzgeberisches Handeln einzubeziehen. Vor allem das Zielbild eines „Internet der Dinge“ (Engl.: Internet of Things, IoT)²⁷ kann als Orientierung für eine zukunftsfähige Gesetzgebung dienen. In dieser Vision sind sämtliche menschlichen Handlungen digital dokumentierbar. Umgekehrt werden sie in ihr nur als real angenommen, wenn sie auch tatsächlich digital belegt sind. Realität und Datenschatten sind demzufolge deckungsgleich. Jede technologische Erneuerung der nächsten Jahrzehnte wird höchstens eine Teilmenge des Internet der Dinge abbilden. Eine auf einem umfassenden IoT basierende Gesetzgebung kann daher als zukunftsfähig gelten.

Der wahrgenommene Datenschatten schließlich ist jener, den Verbraucher bei sich selbst vermuten. Er divergiert im Allgemeinen sowohl von dem tatsächlichen als auch dem möglichen Datenschatten. Dennoch kann ein wahrgenommener Datenschatten – egal wie realitätsnah dieser sein mag – Bürger in ihrem Verhalten beeinflussen. In der Regel versuchen diese, ihren Datenschatten durch verschiedene Maßnahmen zu verringern. Wie jedoch später in diesem Kapitel erläutert wird, sind die unmittelbaren Einflussmöglichkeiten der Nutzer sehr begrenzt. Der wahrgenommene Datenschatten und die daraus abgeleiteten Maßnahmen können damit auch nur einen begrenzten Effekt auf das Ausmaß des tatsächlichen Datenschattens haben.

²⁷ Das Internet der Dinge (IoT) bezeichnet eine technologische Vision. In ihr ist jedes Gerät mit einem Prozessor und einem Internetzugang ausgestattet. Damit erhebt, verarbeitet und versendet es Daten über seinen Betrieb. Die Daten des Gerätes stoßen weit verzweigte Prozessbäume zu verschiedensten Unternehmen und Geräten an, die wiederum weiter kommunizieren und senden. Daher wird die Vision auch als *Internet of Everything* bezeichnet (IoE). In ihr hat jedes Objekt eine digitale Kennung (Engl.: ID) und jeder physische Prozess wird digital gespiegelt. Physische und digitale Welt stimmen in dieser Vision vollends überein. Die Kommunikation im IoT bzw. IoE findet direkt zwischen den beteiligten Maschinen statt, menschliche Sachbearbeiter sind nicht vorgesehen.

Abbildung 7 Modell der Datenschattenentstehung


Studie Verbraucherspur in der Digitalen Welt – April 2018

infas

Abbildung 7 zeigt eine schematische Entstehungsgeschichte eines Datums. Sie beginnt mit der Erfassung menschlichen Handelns durch Maschinen, insbesondere durch die Nutzung digitaler Medien, aber auch der Datenerfassung durch Dritte (1). Die erfassten Daten werden von Unternehmen gespeichert (2) und können – meist auf Grundlage gesetzlicher Bestimmungen oder nach Aufforderung der Nutzer – wieder gelöscht werden (3). Viele Daten werden jedoch unbegrenzt gespeichert und mit anderen Unternehmen geteilt (4). Die gespeicherten Daten werden möglicherweise mit anderen Daten des Nutzers verknüpft oder unter Anwendung von Algorithmen ergänzt (5). Erst durch die Anwendung dieser Daten durch Unternehmen kann der Verbraucher Folgen des entstandenen Datenschattens zu spüren bekommen (6) – in der Regel in deutlich zeitlichem Abstand zur eigentlichen Datengenese. In den folgenden Kapiteln werden alle Einzelschritte bei der Entstehung des Datenschattens ausführlich erläutert.

5.1.2 Methoden der Datenerfassung

Die Entstehung eines persönlichen Datenschattens wurde erst mit der Einführung moderner IT-Technologien und Kommunikationsmethoden mög-

lich. Hierzu zählen zunächst die immer stärkere Miniaturisierung von Computern, insbesondere die Entwicklung vom Desktop-PC hin zu Smartphones und smarten Haushaltsgeräten. Die Vernetzung von Internet und Mobilfunk ist ebenso maßgeblich. Serverseitig sind die Speicherkosten pro GByte rapide gefallen. Gleichzeitig sind die Möglichkeiten zur Analyse von Daten enorm gewachsen. In dieser Kombination ist es prinzipiell möglich, beliebige Daten zu sammeln, zu übermitteln, zu speichern und zu analysieren. Dies alles geschieht ohne nennenswerte Grenzkosten.

Im Folgenden wird grafisch umgesetzt eine Vielzahl zentraler Möglichkeiten zur Erhebung von Daten beschrieben und in den Kontext eingeordnet. Die Aufzählung dient der Illustration anhand aktueller Entwicklungen. Sie erhebt keinen Anspruch auf Vollständigkeit, sondern versucht zuvorderst, einen breiten Eindruck wichtiger Quellen des Datenschattens zu vermitteln. Basale Möglichkeiten der Datensammlung, z. B. durch Apps, werden als allgemein bekannt angenommen.

Bewegungsprofile

Mittels den in Mobiltelefonen eingebauten GPS-Sensoren lassen sich bekanntermaßen hochpräzise Bewegungsprofile erstellen. Doch auch wenn Nutzer den Zugriff auf GPS-Sensor und Standortdienste deaktiviert haben, verbleiben einige Möglichkeiten, um ihre Position zu bestimmen. Das Mobilfunknetz selbst erlaubt eine Ortung, da bereits die Identität der Basisstation, mit der das Telefon verbunden ist, eine grobe Positionsbestimmung zulässt. Diese kann durch das Einbeziehen der Signalstärke dieser und benachbarter Basisstationen weiter präzisiert werden (Trilateration). Neben dem Mobilfunknetz können weitere Funknetze zur Positionsbestimmung dienen. Die Kombination der verschiedenen verfügbaren Wifi-Netzwerke in Kombination mit der Signalstärke ist im Allgemeinen einzigartig für jede Position. Gerade in urbanen Räumen erlauben die umgebenden Wifi-Netzwerke daher eine hochpräzise Ortung, auch innerhalb von Gebäuden.

IP-basierte Ortsbestimmung

Daneben gibt es Methoden der Ortsbestimmung, die komplett serverseitig funktionieren, indem die geographische Position einer IP-Adresse bestimmt wird (Engl.: geocoding). Da Betreiber von Breitband- und Mobilfunknetzen die IP-Adressen blockweise vergeben, bekommt beispielsweise das Netz einer Stadt einen bestimmten IP-Bereich zugeteilt. Daher lässt die IP-Adresse eines Endgeräts meist auf die Stadt schließen, in der sich das Gerät befindet. In Mobilfunknetzen werden solche IP-Blöcke meist an Basisstationen vergeben, so dass auch hier ein Endgerät anhand der IP-Adresse zumindest grob lokalisiert werden kann. Nicht zuletzt können Angaben oder Fotos von Nutzern in sozialen Medien zur Positionsbestimmung genutzt werden. In Summe lässt sich für jeden Nutzer digitaler Geräte (mobil oder stationär) eine Positionsbestimmung vornehmen. Diese reicht vom reinen Wohnort bis hin zu präzisen Bewegungsprofilen. Ein durchschnittlicher Nutzer kann diese Methoden selber nur ansatzweise einschränken, nicht jedoch völlig wirkungslos werden lassen.

Web- und App-Tracking

Ähnlich wie bei geographischen Bewegungsprofilen bestimmt Web-Tracking die „Bewegung“ des Nutzers im Netz. Handelte es sich dabei früher ausschließlich um Webseiten, kommen nun auch Apps und Seiten in sozialen Netzwerken hinzu. Die Techniken des Web-Tracking sind jedoch in allen drei Fällen verschieden. Klassisches Web-Tracking verfolgt den Browserverlauf eines Nutzers mittels sogenannter Cookies, eigens erzeugter Kennnummern, die im Browser gespeichert werden. Besucht der Nutzer eine Webseite, sendet der Browser den Cookie an den jeweiligen Server. Dieser erkennt die Kennung des Nutzers und übermittelt sie beispielsweise an ein Werbenetzwerk, das das Browsingverhalten des Nutzers somit protokollieren kann. Alternative Ansätze nutzen sogenannte „Unsichtbare Pixel“, um einem Werbenetzwerk mitzuteilen, dass der Nutzer diese spezielle Webseite besucht hat.

Technische Profile

Die Browserkonfiguration eines Nutzers ist im Allgemeinen einzigartig, das heißt beispielweise die Kombination aus Browserversion, Plug-Ins, Zeitzone, Sprachoptionen, Schriftsätzen, Betriebssystem und weiteren Komponenten. Diese einzigartige Signatur wird ebenfalls genutzt, um Web-Verhalten zu protokollieren (Engl.: browser fingerprinting). Bei Apps kann der Betreiber des Services den Nutzer in der Regel an dessen Login erkennen. Parallel enthalten Apps kleine Programmteile (SDKs) von Drittfirmen, zum Beispiel aus der Werbebranche. Diese informieren die jeweilige Drittfirma über den Umgang des Nutzers mit der App. In sozialen Netzwerken protokolliert nicht zuletzt der Betreiber des Netzwerkes, welche Seiten ein Nutzer aufruft.

Shopping

Aus dem Einkaufsverhalten ergibt sich ein detailliertes Profil des Lebensstils eines Nutzers. Hierfür werden sowohl online bestellte Produkte gespeichert, als auch mit Kundenkarten im Einzelhandel erworbene Produkte. Online wird bereits registriert, wenn Kunden entsprechende Produktseiten besuchen oder Produkte auf einer Wunschliste vormerken. Ähnliche Informationen können auch aus den Abrechnungsdaten gewonnen werden (z. B. über PayPal) oder wenn Nutzer die Produkte wieder verkaufen (z. B. auf eBay).

Messenger und Soziale Netzwerke

Auch die Inhalte von Chat-Verläufen oder E-Mail-Konversationen können maschinell analysiert werden. Zum einen können die Inhalte beispielsweise über Keywords oder Markennamen analysiert werden. Des Weiteren lassen sich die sprachlichen Fähigkeiten des Nutzers auswerten, also etwa die Größe des Wortschatzes oder die Nutzung von Hilfsverben zur Vermeidung von Konjugationen. Schließlich könnten sich durch die Analyse von Beginn und Ende von Chat-Verläufen sogar Machtstrukturen der beteiligten Personen ableiten. Die Analysemethoden sind mannigfaltig und hängen von dem konkreten Anwendungsfall ab. Bei kostenlosen E-Mail-Anbietern werden die Postfächer bereits durchsucht, um gezielter Werbung schalten zu können. Bei Anbietern von Messengern ist dies bislang zwar nicht dokumentiert, erscheint jedoch plausibel. In sozialen Netzwerken stellen Nutzer viele Daten freiwillig ein. Diese reichen von scheinbar harmlosen Informationen bis hin zu hoch intimen Angaben. Insbesondere bei Kontaktbörsen geben Nutzer eine Vielzahl brisanter Informationen preis, beispielsweise religiöse, politische oder sexuelle Vorlieben.

Fotos und Videos sowie Sprach- bzw. Stimmerkennung

Objekte und Personen in Fotos lassen sich prinzipiell mit ausreichender Sicherheit erkennen. Insbesondere die Gesichtserkennung hat in den vergangenen Jahren erstaunliche Fortschritte gemacht. Aus Fotos lassen sich beispielsweise anhand der abgebildeten Gegenstände oder Personen ableiten, welche Hobbys ein Nutzer hat oder wie sich sein soziales Umfeld zusammensetzt. In Videos lassen sich darüber hinaus auch Bewegungen erkennen, die beispielsweise zur Identifikation von Sportarten dienen können, aber auch auf den Fitnesslevel oder Alkoholspiegel eines Nutzers schließen lassen. Diese Liste der aus Videos extrahierbaren Erkenntnisse kann um eine beliebige Anzahl weiterer Beispiele ergänzt werden, worauf jedoch im Rahmen dieses Berichts verzichtet wird. Neben Bildern und Videos lassen sich auch Audioaufnahmen analysieren. Aus der Stimmlage lassen sich beispielsweise Rückschlüsse auf Emotionen, wie beispielsweise Freude oder Aggression, oder sogar Krankheiten wie Depression ziehen. Derartige Systeme werden bereits in Callcentern genutzt, um besonders aggressive oder frustrierte Anrufer zu identifizieren.

Eine der zentralen Neuerungen der vergangenen Jahre waren sprachgesteuerte Assistenzsysteme. Lautsprecher als solche, oder integriert in Geräten oder auch in Fahrzeugen, nehmen Befehle entgegen und setzen diese um. Die meisten dieser Geräte kommunizieren mit einem Server. Sie schneiden Gespräche mit und versenden diese an externe Server. Die akustischen Aufnahmen können relativ präzise in Text übertragen und im Anschluss analysiert werden.

Spiele

Computer-Spiele – egal ob am PC, auf der Spielkonsole, im Browser oder auf Smartphones – können wertvolle Informationen über die kognitiven und motorischen Fähigkeiten eines Nutzers liefern. Ganz ähnlich den in der Psychologie eingesetzten Computer-basierten Tests erfassen sie beispielsweise Reaktionsgeschwindigkeit, motorische Koordination und Kombinationsvermögen eines Nutzers. Aus Massen-Online-Gemeinschaftsspielen (MMOG), wie beispielsweise World of Warcraft, können darüber hinaus Rückschlüsse auf die Persönlichkeitsstruktur eines Nutzers gezogen werden.

Tippverhalten auf Tastaturen

Auch das Tippverhalten (keystroke dynamics) eines Nutzers erlaubt, diesen zu identifizieren oder Rückschlüsse auf dessen geistige Verfassung zu ziehen. Tatsächlich sind die Rhythmen, mit denen Menschen Texte eingeben, sehr individuell. So wird die Analyse des Tippverhaltens bereits heute genutzt, um Nutzern Zugang zu Computern zu geben beziehungsweise unberechtigten Personen diesen zu verweigern. Müdigkeit oder Stress beeinflussen das Tippverhalten ebenso und lassen damit Rückschlüsse auf den Nutzer zu.

Eye-Tracking

Diese Technologie registriert den Blick des Nutzers auf einen Bildschirm. Traditionell wurde sie von Psychologen oder Werbefachleuten eingesetzt, um unter Laborbedingungen zu erforschen, auf welchen Teil des Bildschirms ein Nutzer zuerst schaut. Mittlerweile ist die Technik jedoch auch in gängigen Produkten eingebaut und kann beispielsweise auf der Frontkamera eines Smartphones eingesetzt werden. Solche Systeme können prinzipiell messen, welche Objekte auf dem Bildschirm die Aufmerksamkeit des Nutzers binden und erlauben somit Rückschlüsse auf dessen Persönlichkeitsstruktur. Aktuell ist nicht bekannt, dass solche Systeme in signifikantem Umfang zum Einsatz kommen. Anders verhält es sich mit der Erkennung von Stress oder Müdigkeit anhand von Eye-Tracking. Solche Systeme kommen bereits in Autos zum Einsatz, um übermüdete Fahrer zu warnen.

Fitness-Tracker und Gesundheits-Messgeräte

Das vergangene Jahrzehnt hat eine Vielzahl an Geräten und Apps im Gesundheits- und Wellnessbereich hervorgebracht. Dazu gehören Schrittzähler und Jogging-Apps, sowie Zahnbürsten, Waagen und Blutdruckmessgeräte mit Bluetooth- oder Wifi-Schnittstelle. All diese Geräte haben gemein, dass sie Daten erheben, die mehr oder weniger direkt den Gesundheitszustand des Nutzers protokollieren. Die anfallenden Daten werden jedoch im Allgemeinen nicht als besonders schützenswerte Gesundheitsdaten behandelt, sondern wie beliebige andere Nutzerdaten.

IoT: Smart Car, Smart Homes und Smart Fridges

Im Rahmen eines sich langsam manifestierenden IoT sind wir zunehmend von smarten Geräten umgeben. Hierzu zählen Autos, aber auch Heizungsthermostate, Kühlschränke oder Saugroboter. All diese Geräte sammeln Daten über das Verhalten ihrer Besitzer oder deren Lebensumstände. Das smarte Auto kann ein Bewegungsprofil erstellen. Der protokollierte Treibstoffverbrauch, sowie Daten über Beschleunigung und Abbremsvorgänge ermöglichen jedoch auch ein Risiko- oder sogar Persönlichkeitsprofil des Fahrers. Mit Hilfe von smarten Stromzählern und Kühlschränken ist es möglich, bestimmte Handlungen und Vorlieben des Nutzers im geschützten Bereich seiner Wohnung zu dokumentieren. Ob und wie weit bereits ein Markt für solche Daten existiert, ist unklar. Der Versuch eines Herstellers von Saugrobotern, Karten von Privatwohnungen an Dritte zu veräußern, ist nur ein Beispiel für eine mögliche (unzulässige) Verwendung solcher Nutzerdaten.

Handlungen Dritter

Die bisher beschriebenen Methoden haben eines gemeinsam: der Nutzer ist selber Urheber der über ihn erhobenen Daten. Im Zeitalter sozialer Netzwerke und einer vernetzten Wirtschaft greift dieses Modell jedoch zu kurz. Nutzer geben – im Allgemeinen unbewusst – kontinuierlich Informationen über Dritte preis. Angenommen, ein fiktiver Nutzer A verzichtet komplett auf digitale Medien und entzieht sich gezielt. Sobald ein online-affiner Freund ihn in einem sozialen Netzwerk sucht, sind sowohl der Name als auch die Verbindung zu dieser Person dem Netzwerk bekannt. Ebenso kann eine Person, die in keinem sozialen Netzwerk vertreten ist, dort beispielsweise auf Fotos verlinkt werden, was wiederum weitere Informationen, zum Beispiel einen Teil des sozialen Umfelds, zu dieser Person preisgibt. Durch Methoden der Gesichtserkennung können nun theoretisch Millionen anderer Online-Bilder nach dieser Person durchsucht werden. Doch nicht nur durch soziale Netzwerke, auch durch andere Online-Handlungen Dritter, kann die Identifikation von Personen aufgedeckt werden, die selbst versuchen, möglichst wenig Daten im Internet zu hinterlassen. Im Rahmen des Skandals um Facebook und Cambridge Analytica wurden solche Datensätze als „Shadow Profiles“ bekannt.

Besonders problematisch ist, dass Personen in der Regel nicht wissen, welche Informationen über sie von Dritten an welcher Stelle verbreitet werden. Dies macht es unmöglich, sich Auskunft zu den entsprechenden Daten (z. B. im Rahmen einer Datenselbstauskunft) bei allen betroffenen Unternehmen einzuholen. Es verdeutlicht deutlich, dass die (un)bewusste Weitergabe von Daten durch Dritte kaum zu unterbinden ist und zu einer beträchtlichen Quelle des Datenschattens erwachsen kann. Die individuelle Datenaskese zur Minimierung des eigenen Datenschattens greift angesichts der zunehmenden Digitalisierung offensichtlich zu kurz.

5.1.3 Weitergabe und Austausch von Daten mittels Identifikatoren

Die entstandenen Daten werden von dem erhebenden Unternehmen gespeichert und in seltenen Fällen wieder gelöscht (siehe Abbildung 7) – sei es aufgrund rechtlicher Vorschriften oder auf Veranlassung des Nutzers hin (beispielsweise im Rahmen des § 35 BDSG). Für die speichernden Unternehmen selbst gibt es keinen Anreiz, Daten aus eigener Initiative zu löschen. In den meisten Fällen werden Daten an andere Unternehmen weitergegeben. Die speichernden Unternehmen können die Daten wiederum an eine beliebige Anzahl anderer Unternehmen weiterreichen (siehe Abbildung 7). Um das Verhalten eines Nutzers aufzuzeichnen und die entstandenen Daten mit weiteren Unternehmen auszutauschen, bedarf es sogenannter „Identifikatoren“. Dabei handelt es sich um eine Art Seriennummer des Nutzers, anhand derer verschiedene Unternehmen die über ihn bekannten Daten zuordnen können. Diese Nutzerkennungen müssen über die Grenzen der verschiedenen Dienste hinweg konstant bleiben, damit Shopbetreiber, Nachrichtenseiten oder Werbetreibende den Nutzer über denselben Identifikator identifizieren können.

In der analogen Welt reichen meist bereits Name und Adresse um einen Nutzer hinlänglich genau zu identifizieren. Um eine höhere Präzision zu erreichen, werden mitunter zusätzlich noch Geburtstag und -ort erfasst. In der digitalen Welt hingegen werden beinahe ausschließlich digitale Identifikatoren eingesetzt. Diese ergeben sich häufig aus den Koordinaten von Kommunikationsendpunkten, beispielsweise E-Mail-Adressen. Oder es handelt sich um bereits existierende Identifikatoren von Geräten, die dem Nutzer eindeutig zuzuordnen sind. Die E-Mail-Adresse eines Nutzers ist der wohl gängigste Identifikator, da er vielen Diensten aufgrund der Registrierung eines Nutzers mit einer E-Mail-Adresse bekannt ist. Eine ähnliche Funktion übernehmen Accountnamen bei großen Dienst Anbietern wie Facebook, Amazon oder Google. Insbesondere Facebook versucht, eigene Identifikatoren über Funktionen wie „Login über Facebook“ im Markt zu verankern. Wird der Nutzer beispielsweise bei einem Reiseportal Kunde,

muss er dort keinen eigenen Account anlegen, sondern kann sich mit seinem existierenden Facebook-Account einloggen. Er ist also bei all diesen Diensten unter seinem Facebook-Identifikator bekannt. Ein weiterer häufig genutzter Identifikator ist die Telefonnummer des Nutzers. Diese liegt ebenfalls für beinahe jeden Nutzer vor. Aus Sicht eines Diensteanbieters hat sie den Vorteil, dass sie durch Apps direkt auf dem Telefon ausgelesen werden kann. Alternativ können Seriennummern von Geräten, Werbe-IDs von Betriebssystemen oder Cookies als Identifikatoren genutzt werden.

Sollten die oben beschriebenen Methoden versagen, können Nutzer dennoch immer anhand einzigartiger technischer Konstellationen oder anhand ihres einzigartigen Verhaltens identifiziert werden. Beispielsweise ergibt sich für jeden Browser ein Profil, bestehend aus den Versionsnummern von Browser und Betriebssystem, den präferierten Sprachen, installierten Fonds und Plug-Ins. Dieses Profil ist in der Regel einzigartig. Ähnliches gilt beispielsweise für die Liste der auf einem Smartphone installierten Apps. Anhand solcher Profile können Nutzer auch über verschiedene Dienste hinweg eindeutig identifiziert werden. Auch das individuelle Verhalten eines Nutzers (z. B. die Liste seiner besuchten Websites) ist im Allgemeinen einzigartig, und kann zur Identifikation herangezogen werden. Selbiges gilt für sein äußeres Erscheinungsbild.

Identifikatoren werden mitunter nicht im Klartext verwendet, sondern durch sogenanntes „Hashing“ verändert. Vereinfacht gesprochen handelt es sich dabei um eine Funktion, die aus einem Eingabewert eine scheinbar zufällige Zahl berechnet. Ein zentrales Merkmal der Hashfunktion ist, dass minimale Änderungen des Eingabewertes deutliche Änderungen des Hashwertes hervorrufen. Bei den verbreiteten Hashfunktionen ist es im Allgemeinen nicht möglich, aus einem Hashwert rückwirkend den dazugehörigen Eingabewert zu berechnen oder zu einem gegebenen Eingabewert einen zweiten zu finden, der denselben Hashwert hat. Hashfunktionen werden aus unterschiedlichen Gründen eingesetzt. Rein technisch gesehen haben sie den Vorteil, dass sie Werte derselben Länge erzeugen. Es ist mitunter also deut-

lich einfacher, Hashwerte zu versenden oder zu vergleichen als beliebig unterschiedliche Datensätze. Gleichzeitig dient der Einsatz von Hashwerten auch einer Pseudonymisierung. Prinzipiell wäre es möglich, durch den Einsatz sogenannter Schlüsselwerte Sorge zu tragen, dass der Hashwert eines Nutzers von Unternehmen zu Unternehmen unterschiedlich ausfällt. Um einen Austausch personenbezogener Daten zwischen Unternehmen zu ermöglichen, ist es jedoch unabdingbar, dass alle Firmen denselben Identifikator für denselben Nutzer verwenden. Daher werden Schlüsselwerte üblicherweise nicht verwendet. Selbst mit dem Einsatz von Schlüsselwerten leidet Hashing jedoch unter den in Kapitel 5.1.7 aufgeführten Problemen sämtlicher Pseudonymisierungsverfahren.

5.1.4 Analyse und Anreicherung von Daten

Die Daten werden mithilfe statistischer Modelle über andere Nutzer angereichert (siehe Abbildung 7). Meist reichen einige wenige Anhaltspunkte als Anker, um Aussagen zu treffen oder fehlende Informationen über einen Nutzer zu ergänzen. Dies geschieht im fließenden Übergang zur Analyse und Interpretation von Daten. Die Anhaltspunkte sind entweder wenige bekannte Attribute oder ein Ausschnitt aus dem sozialen Umfeld des Nutzers. Je nachdem, welche Daten vorhanden sind, kommen unterschiedliche Methoden zur Anwendung. Die Informatik hat hier eine Vielzahl statistischer Techniken entwickelt. Sie können grob in den Bereich Datamining oder Knowledge Discovery in Databases (KDD) zusammengefasst werden – mit fließendem Übergang zu Methoden der künstlichen Intelligenz (KI). Der folgende – erneut grafisch umgesetzte – Abriss kann nur einen groben Eindruck der Möglichkeiten vermitteln. Dabei sind die mathematischen Zusammenhänge stark vereinfacht. Die Informatik hält ein ganzes Spektrum ergänzender und ähnlicher Methoden bereit. Es gilt jedoch festzuhalten, dass mit mathematischen Methoden der Datenschatten eines Menschen weit über das angereichert werden kann, was dieser direkt preisgegeben hat.

Anreicherung von Datenbruchstücken

Die sogenannte Nächste-Nachbar-Klassifikation vergleicht die Daten der „Zielperson“ mit ähnlichen Personen. Angenommen Person A ist 51 Jahre, männlich und 183cm groß. Die drei ähnlichsten Nutzer sind zwischen 49 und 50 Jahre alt, männlich und zwischen 180 und 184 cm groß. Zwei von ihnen haben Knieprobleme. Die Annahme wäre, dass auch Person A Knieprobleme hat. Alternativ können ähnliche Objekte im Vorfeld zu Gruppen (Clustern) geordnet werden. Einer dieser Cluster könnte beispielsweise Männer von 50 bis 57 Jahren und 180 bis 187cm umfassen. Wenn Nutzer dieses Clusters häufig Knieprobleme haben, dann könnte man dies auch für Person A annehmen. Die Assoziationsanalyse, meist als „Warenkorbanalyse“ bekannt, versucht, aus dem Auftreten eines Attributes auf das anderer zu schließen. Haben viele Nutzer eine bestimmte Kombination von Produkten gekauft, ist die Wahrscheinlichkeit groß, dass auch weitere Nutzer diese Kombination wählen. Dieser Ansatz kann, neben der Anwendung bei Warenkörben, auch für beliebige Attribute (fehlende Datenpunkte) eines Nutzers angewendet werden.

Anreicherung anhand von Netzwerk-Informationen

Sofern keine Informationen über einen Nutzer, aber über einige seiner Bekannten bzw. Freunde existieren, können diese auf den Nutzer übertragen werden. Die hier genutzten Methoden stammen ursprünglich aus der Bibliometrie und wurden zunächst in Suchmaschinen angewendet. Zitiert ein Buch ein anderes, wird davon ausgegangen, dass die beiden inhaltlich enger verwandt sind als zwei beliebige andere Werke aus dem Büchermeer (Engl.: citation). Auf soziale Netzwerke übertragen, bedeutet dies, dass zwei befreundete Personen ähnlicher sind, als zwei beliebige Personen. Zitieren zwei Bücher sich nicht direkt, verweisen aber beide auf dasselbe dritte Buch, dann sind sie ebenso wahrscheinlich inhaltlich enger verwandt als zwei beliebige Bücher (Engl.: co-citation). Diese Annahme ist ebenso auf soziale Netzwerke übertragbar. Citation und Co-citation sind nur zwei von vielen Konstrukten mit denen Informationen in sozialen Netzwerken „verschoben“ werden können, um fehlende Daten zu ergänzen.

Geo-Scoring

Alternativ kann auch die Adresse einer Person genutzt werden, um fehlende Informationen zu ergänzen. Ähnlich dem Clustering ist hier die Annahme, dass Menschen, die in einer Region zusammenwohnen, sich ähnlicher sind als beliebige Menschen aus verschiedenen Regionen. Datenpunkte, wie das Einkommen, der ethnische Hintergrund oder das Wahlverhalten können mit diesen Methoden mitunter recht präzise geschätzt werden.

5.1.5 Automatisierte Entscheidung und die Rückwirkung des Datenschattens auf die Verbraucher

Die Auswirkungen eines Datenschattens auf die Lebenswelt des Nutzers zeigen sich erst bei seiner konkreten Nutzung durch Unternehmen (siehe Abbildung 7 weiter oben). Dies geschieht durch automatisierte Entscheidungen auf Grundlage der erhobenen Daten. Die Bildung von Nutzerprofilen entspricht einer Vorstufe automatischer Entscheidungen, indem verschiedene Nutzer zu Typen zusammengefasst werden. Über diese können in einem zweiten Schritt automatisierte Entscheidungen getroffen werden. Allerdings ist ein Trend zu immer individuelleren Entscheidungsalgorithmen zu konstatieren, die ohne vorherige Profilbildung auskommen. Ähnlich der bereits beschriebenen Logik hinsichtlich des Datenschattens müssen auch hier zwischen (a) heute möglichen, (b) tatsächlich eingesetzten und (c) zukünftig möglichen Technologien unterschieden werden.

Die Techniken zur automatisierten Entscheidungsfindung, unabhängig davon, ob es um einfache statistische Methoden oder den Einsatz künstlicher Intelligenz geht, unterliegen einem ständigen technologischen Fortschritt. Sie werden in verschiedenen Unternehmen in unterschiedlichem Umfang und mit unterschiedlichem Sachverstand eingesetzt. Auch derzeit noch experimentelle Methoden können sukzessive zu Allgemeinwissen werden. Gleichzeitig werden neue, noch bessere Methoden entwickelt werden. Hypothetisch kann in Zukunft prinzipiell jede Beziehung zwischen zwei Vorkommnissen maschinell erkannt werden. Jedes Muster, das von einem Menschen erkannt wird, kann auch automatisch erkannt werden. Und es wird sogar darüber hinausgehen, da ein Mensch nur eine bedingte Anzahl an Datenpunkten und Dimensionen analysieren kann. Nicht erkennen lassen sich hingegen extrem seltene „Muster“, bei denen es sich im eigentlichen Sinne nicht um Muster handelt. Folgt man einer weiteren Hypothese, wird die Technologie zur Mustererkennung sich auch in kleine Geräte integrieren lassen. Automatisierte Entscheidungen werden also nicht zwangsweise

in großen zentralen Serverfarmen stattfinden, sondern durchaus lokal und dezentral.

Erst zum Zeitpunkt der Nutzung des Datenschattens drohen dem Nutzer konkrete Folgen seiner Datenpreisgabe. Die automatisierten Entscheidungen können dabei sowohl zum Wohle des Nutzers als auch zu dessen Schaden ausfallen. Beispielsweise können Daten zum Lebenswandel eines Nutzers zur Berechnung seines Herzinfarkttrisikos verwendet werden. Die Warnung an den Nutzer über sein erhöhtes Risiko wird dieser gegebenenfalls noch positiv wahrnehmen. Die gleichen Systeme, die den Nutzer warnen, können aber auch ein Unternehmen dazu befähigen, dem Nutzer aufgrund des erhöhten Risikos eine Versicherung zu verwehren.

Die Nutzung der Daten durch Unternehmen liegt im Allgemeinen in der Zukunft. Insbesondere zum Zeitpunkt der Datenerfassung ist dem Nutzer in der Regel nicht ersichtlich, für welche Fragestellung diese später verwendet werden. Beispielsweise lässt sich aus der Angabe oder Erfassung von Daten über die Essgewohnheiten eines Nutzers nicht absehen, dass darüber zukünftig möglicherweise auf sein Herzinfarktisiko geschlossen wird. Es lässt sich noch nicht einmal der Lebensbereich abschätzen (z. B. das Versicherungswesen), in dem die Daten zum Einsatz kommen können. Wie im vorangegangenen Beispiel kann der Lebensbereich, in dem die Folgen des Datenschattens für den Nutzer ersichtlich werden, sich vollkommen vom Lebensbereich der Datenerhebung unterscheiden.

Aus Unternehmenssicht muss die Entscheidungsgrundlage für eine solche automatisierte Entscheidung nicht unbedingt korrekt sein, um nutzenbringend zur Anwendung zu kommen. In den meisten Anwendungsfällen reichen beispielsweise 80 Prozent Genauigkeit, um automatisierte Entscheidungen finanziell lukrativ zu gestalten. Der Mehrwert der korrekten Entscheidungen muss lediglich die Verluste durch falsche Entscheidungen überwiegen. Von Anwendungsfall zu Anwendungsfall kann der Einsatz automatisierter Bewertungsschemata bereits bei deutlich geringeren Erfolgsquoten lohnend sein.

Da der Zeitpunkt des Einsatzes solcher Algorithmen zumeist deutlich nach der eigentlichen Datenentstehung liegt, ist die gezielte Angabe falscher Informationen durch Nutzer meist nicht wirksam, um Folgen des Datenschattens zu vermeiden. Der Nutzer kann zum Zeitpunkt des Streuens der Daten gar nicht wissen, unter welchen Gesichtspunkten diese später analysiert werden. Genauso können sich die Rahmenbedingungen, die der Analyse zu Grunde liegen, ändern. Demnach ist auch in der Regel nicht vorherzusehen, welche Art von Falschinformationen lohnenswert für den Nutzer sein könnten. Die automatisierte Entscheidung kann dementsprechend in das Leben des Nutzers eingreifen, unabhängig davon, ob dies aufgrund einer korrekten oder einer falschen Datenlage geschieht.

Zur weiteren Veranschaulichung wird die bis hierhin erläuterte Logik der automatisierten Entscheidungen und Bewertungsschemata anhand von vier Lebensbereichen noch einmal exemplarisch dargestellt.

Lebensbereich: Shopping

Von Interesse für Onlinehändler sind zunächst die Einkommensverhältnisse beziehungsweise die Kreditwürdigkeit des Nutzers. Diese lassen sich zum Beispiel aus dessen bisherigem Einkaufsverhalten, früheren Urlaubsdestinationen, dem Arbeitgeber oder Bildungsstand ableiten. Besonders einfach ist es, das Einkommen aus dem mobilen Endgerät des Nutzers zu ermitteln. Besucht dieser die Webseite eines Onlinehändlers mittels eines 1000€-Smartphones, ist er vermutlich wohlhabend und in der Regel nicht besonders preissensitiv. Die Frage, mit welchen Marken und Produkten ein Nutzer zu bewerben ist, klärt sich wiederum am bisherigen Einkaufsverhalten (oder dem der Freunde oder ähnlicher Nutzer) oder aus der Liste besuchter Webseiten. Die Playlist aus Musikdiensten gibt ebenfalls Aufschluss über Alter und Lebensstil des Nutzers. Derzeit sind solche Kundenprofile noch relativ statisch. Die Entwicklung geht jedoch in Richtung dynamischer und situationsbezogener Bewertungen. Diese werden die momentane Verfassung des Kunden reflektieren und analysieren, ob man ihn beispielsweise zu einem Impulskauf anregen kann.

Lebensbereich: Gesundheit

Viele Gesundheitsdaten werden direkt von privaten Geräten der Endkunden erhoben, beispielsweise durch digitale Waagen, Blutdruckmessgeräte, Blutzuckermessgeräte oder aber durch Fitness-Tracker. Dazu kommen Daten von Gesundheits- und Hygieneprodukten, wie elektrischen Zahnbürsten, Massage- oder Muskelstimulationsgeräte – vorausgesetzt, sie sind digital verbunden. Unfallrisiken, wie beispielsweise durch Extremsportarten, lassen sich unter anderem anhand von Posts und Fotos in sozialen Netzwerken oder der Liste besuchter Websites erkennen. Darüber hinaus lassen sie sich möglicherweise aus dem Verhalten des Freundeskreises ableiten. Bei Risikofaktoren, wie schlechter Ernährung oder Mangel an Bewegung, verhält es sich ähnlich. Diese lassen sich auch mittels Korrelationen aus dem Lebensstil ableiten. Die Methoden der Datenanalyse hängen von dem konkreten Krankheitsbild beziehungsweise dem Risikofaktor ab und können von Fall zu Fall sehr unterschiedlich sein. Depressionen verfügen beispielsweise über vielschichtige Symptome, die sich auch am Smartphone identifizieren lassen. Eine reduzierte Makrobewegung, weil der Patient das Haus beispielsweise nicht mehr verlässt, ergibt sich alleine aus dem mit GPS erstellten Bewegungsprofil.

Sozialer Rückzug kann sich anhand reduzierter Telefonate oder Chat-Kontakte manifestieren. Auch eine Veränderung der Stimmfarbe oder die Änderung des Medien- und Spielekonsums ist direkt auf dem Telefon dokumentierbar. Sofern der Nutzer sein Telefon häufig nutzt, kann es auch Aufschluss darüber geben, welchen Schlafrythmus dieser hat. Eine Vielzahl an personenbezogenen Daten sind geeignet, um Rückschlüsse auf den Gesundheitszustand eines Nutzers zu ziehen. Dies trifft insbesondere auf psychische und neurologische Krankheiten zu. Die Analysemethoden können mitunter relativ simpel sein. Angenommen, ein Nutzer verkauft seine gebrauchte Jeans der Größe 38/34 auf einem Online-Portal. Dann hat er damit gleichzeitig seine Körperproportionen frei zugänglich im Internet veröffentlicht. Der korrespondierende Body-Maß-Index ist eine zentrale Kennzahl für die Abschätzung von Risiken wie Herzinfarkten oder Diabetes bis hin zu orthopädischen Schäden. In Summe sind also streng genommen sämtliche personenbezogenen Daten prinzipiell als Gesundheitsdaten zu behandeln. Die bisherige Trennung von Gesundheitsdaten und „harmlosen“ Daten scheint aufgrund der Verknüpfungsmöglichkeiten technologisch überholt.

Lebensbereich: Politik

Die politische Orientierung eines Nutzers kann sich unmittelbar aus dem regelmäßigen Besuch von Parteiseiten und einschlägigen Medien ergeben. Darüber hinaus lässt sie sich aus politischen Äußerungen und „Likes“ in Sozialen Medien oder den Freunden in sozialen Netzwerken ableiten. Mittels Clustering lässt auch der Lebensstil eines Nutzers Aufschlüsse über dessen politische Orientierung zu: Personen ohne Auto, die viele Bio-Produkte kaufen, wählen beispielsweise mit hoher Wahrscheinlichkeit anders als regelmäßige Konsumenten von Champagner und Luxushandtaschen. Alleine das Einkaufsverhalten kann also Hinweise auf die politische Orientierung eines Bürgers liefern. Neben der politischen Orientierung kann die Beeinflussbarkeit und die Bereitschaft zum Verbreiten radikaler Nachrichten beziehungsweise sogenannter Fake-News ein zentraler Angriffspunkt für die gezielte Einflussnahme von Unternehmen oder sogar Regierungen sein. Dies zeigt auch die aktuelle Diskussion um eine mögliche Einflussnahme Russlands über Facebook-Kanäle bei der Wahl von Donald Trump. Im Gegensatz zur klassischen politischen Propaganda ist es heute möglich, auf Basis von umfangreichen Datenerhebungen und deren Verknüpfung, Personen deutlich gezielter aber auch subtiler zu beeinflussen.

Lebensbereich: Sexualität

Auch das sehr persönliche Merkmal der sexuellen Orientierung eines Menschen lässt sich aus einer Vielzahl von Faktoren herleiten. Besonders einfach ist es, wenn jemand eine Dating-Seite oder App nutzt oder seine sexuelle Orientierung direkt in einem sozialen Netzwerk angibt. Der Konsum von Online-Pornographie spiegelt die sexuelle Orientierung ebenso wider. Aber auch andere, weniger offensichtliche Parameter, wie beispielsweise der Musikgeschmack oder der Konsum bestimmter Körperpflegemittel, können schon zu einer Eingrenzung dieses Merkmals führen. Auch Freunde können hier, genauso wie bei anderen Lebensbereichen, einen Anhaltspunkt liefern. Persönliche Daten können nicht nur Aufschluss über die sexuelle Orientierung liefern, sondern beispielsweise auch über die Dauer und Belastbarkeit von Beziehungen. Die Länge der Beziehung, die Verschiebung von Kommunikationsmustern, der Kontakt zu Dritten – all diese Faktoren lassen sich aus dem Datenschatten der Nutzer bestimmen. Der Datenschatten eines Menschen kann somit auch schnell reichhaltige Indikatoren über intime Lebensbereiche beinhalten.

5.1.6 Größe und Ausmaß des Datenschattens

Aufgrund der geringen Speicherkosten ist davon auszugehen, dass einmal erhobene Daten massiv repliziert und in den meisten Fällen unendlich lange gespeichert bleiben. Der Datenschatten eines Nutzers wächst somit stetig. Zudem ist wie dargestellt prinzipiell davon auszugehen, dass erhobene Daten durch ein Unternehmen im Anschluss mit Partnerfirmen geteilt werden. Vor allem im Werbesektor tauschen Millionen von Webseiten Informationen mit tausenden von Werbenetzwerken. Zentrale Drehkreuze sind dabei wenige Großkonzerne. Für die Größe eines Datenschattens ist es daher irrelevant, von wie vielen Unternehmen Daten ursprünglich erfasst werden. Für einen umfangreichen Datenschatten reicht möglicherweise bereits eine einzige böartige App, die das Nutzerverhalten protokolliert und teilt. Der Datenschatten wächst daher nicht linear, denn verschiedene Datenquellen sind in weiten Teilen redundant. Stellen wir uns eine App vor, die mittels GPS ein Bewegungsprofil bildet und an hunderte Firmen weiterleitet. Würde jetzt eine zweite App ebenfalls ein Bewegungsprofil erheben und versenden, wächst der bereits vorhandene Datenschatten nicht auf das doppelte. Tatsächlich wächst er nur noch minimal, da alle großen werbetreibenden Unternehmen bereits über alle relevanten Informationen verfügen.

Ein zentraler Datensammler, der häufig unterschätzt wird, ist das Betriebssystem eines Mobiltelefons. Es erhebt umfangreich Daten und speist diese in ein weitreichendes Werbenetzwerk ein. Ob ein Nutzer darüber hinaus noch Apps nutzt, die ebenfalls Daten sammeln, ist so beinahe unerheblich. Damit erweist sich auch der Eindruck vieler Nutzer, ihr Mobiltelefon durch die Vermeidung von Apps „datensparsam“ einzusetzen, als trügerisch. Umgekehrt kann eine einzelne Information, die irrelevant scheinen mag, den Datenschatten eines Menschen rapide vergrößern. Dies gilt insbesondere, wenn die darin enthaltene Information bislang noch nicht dokumentiert war und in Kombination mit existierenden Daten umfangreichen Aussagen über den Nutzer zulässt.

Der Datenschatten ist nicht sinnvoll quantifizierbar, zumindest nicht auf eine Weise, die seine Auswirkung auf das Leben eines Nutzers reflektieren würde. Die üblichen Messgrößen aus der Informatik, wie z. B. Bytes, versagen, da sie nicht ansatzweise den Nutzwert der Informationen quantifizieren. Zum Beispiel kann die gesamte Krankengeschichte eines Menschen in wenigen KByte gespeichert werden. Ein belangloses Video des Nutzers kann dagegen schnell ein GByte umfassen. Dennoch ist der Nutzwert der „kleineren“ Krankengeschichte deutlich höher als der des „größeren“ Videos.

Allenfalls lässt sich eine partielle Ordnung feststellen. Liegen von Nutzer A die Daten XYZ bei Firma O und von Nutzer B die Daten XY bei Firma O, hat A einen größeren Datenschatten als B. Liegen von Nutzer C die Daten XYZ bei den Firmen O und P, hat C einen größeren Datenschatten als A. Diese Ordnung ist nur partiell, da sich nicht alle Datenschatten miteinander vergleichen lassen: liegen von Nutzer D die Daten VW bei Firma O, ist es unmöglich zu bestimmen, ob A oder D einen größeren Datenschatten hat, da VW und XY nicht direkt miteinander vergleichbar sind. Aufgrund der Vielzahl an möglichen erhobenen Daten lässt sich für zwei beliebige Nutzer daher im Allgemeinen nicht feststellen, welcher von beiden über den größeren Datenschatten verfügt.

5.1.7 Anonymisierung und Pseudonymisierung – Datenverarbeitung ohne Datenschatten?

Mitunter herrscht die Meinung, man könne Daten ohne Erzeugung eines Datenschattens verarbeiten. Dies basiert meist auf der Annahme, dass die verarbeiteten Daten im engeren Sinne nicht personenbezogen und damit weniger schützenswert sind. Oder es wird davon ausgegangen, dass der Personenbezug durch Anonymisierung und Pseudonymisierung aufzuheben sei. Beide Annahmen sind kritisch zu bewerten. Wie zuvor beschrieben, werden Daten meist dadurch erzeugt, dass ein digitales Gerät eine Handlung erfasst oder die Handlung erfolgt direkt in der digitalen Welt, wo sie protokolliert wird. Selbst Daten, die scheinbar von Maschinen erzeugt wer-

den, sind meist auf menschliches Handeln zurückzuführen oder reflektieren dieses. Beispielsweise lässt die Abnutzung der Bremsklötze am PKW auf den Fahrstil des Fahrers schließen. Daten sind darüber hinaus im Allgemeinen mit personenbezogenen Identifikatoren versehen, etwa einer E-Mail-Adresse, Telefonnummer oder aber der Seriennummer eines persönlichen Gegenstandes. All diese Daten reflektieren menschliches Handeln und sind an einen konkreten Nutzer gekoppelt. Somit können Daten, die das Verhalten, persönliche Merkmale oder Einstellungen eines Nutzers in irgendeiner Weise widerspiegeln, als personenbezogen betrachtet werden.

Pseudonymisierung bezeichnet das Ersetzen eines Identifikators durch einen weiteren, künstlich erzeugten Identifikator, das sogenannte Pseudonym. So werden die Daten eines Nutzers nicht unter seinem Namen geführt, sondern unter einer eigens erzeugten Kennzahl. Bei Anonymisierung hingegen werden sämtliche Identifikatoren gelöscht. Wenn jegliche Identifikatoren fehlen, ist es in anonymisierten Datensätzen nicht möglich, verschiedene einzelne Daten zusammenzuführen. Diese hoch isolierten Daten erlauben fast keine weiterführenden Analysen. Daher werden in der Praxis oftmals wieder künstliche Identifikatoren eingeführt. Die Anonymisierung von Daten bezeichnet daher meist lediglich eine Pseudonymisierung.

Pseudonymisierung reicht in der Regel nicht, um personenbezogene Daten ausreichend zu verschleiern. Menschliches Verhalten ist hoch individuell. Ist auch nur ein Bruchstück dieses Verhaltens bekannt, ist die Möglichkeit hoch, einen Nutzer damit zu identifizieren. In einer Datenbank mit pseudonymisierten Bewegungsprofilen ist es zwar nicht möglich, Daten anhand eines bekannten Identifikators (z. B. Name oder E-Mail) des Nutzers abzufragen. Sind jedoch beispielsweise Wohnort und Arbeitsstädte des Nutzers bekannt, erlaubt es diese Information, den Nutzer zu erkennen. Es reichen daher oftmals wenige bekannte Datenpunkte, um die Pseudonymisierung auszuhebeln.

5.2 Online-Verhalten der Projektteilnehmer

In Kapitel 5.2 werden die Ergebnisse aus den empirischen Untersuchungsmodulen der Online-Befragungen sowie der App Tracy vorgestellt, bevor in Kapitel 5.3 im Detail auf die Resultate der Anträge auf Selbstauskunft eingegangen wird.

Für die Beschreibung und Typisierung von Verbrauchern anhand empirischer Daten wurde untersucht, in welchem Ausmaß und zu welchen Zwecken die Projektteilnehmer üblicherweise online aktiv sind, welche Maßnahmen sie zum Schutz ihrer persönlichen Daten ergreifen, welche Vorstellungen sie über den eigenen Datenschatten haben und in welchem Umfang sie in Kundenbeziehungen zu den 30 ausgewählten Unternehmen der Selbstauskunft stehen. Hintergrund dieser Fragen ist das Bemühen, zu erkunden, in welchem Ausmaß die Befragten selbst durch ihr Verhalten Daten ins Netz geben und so den Datenschatten nähren.

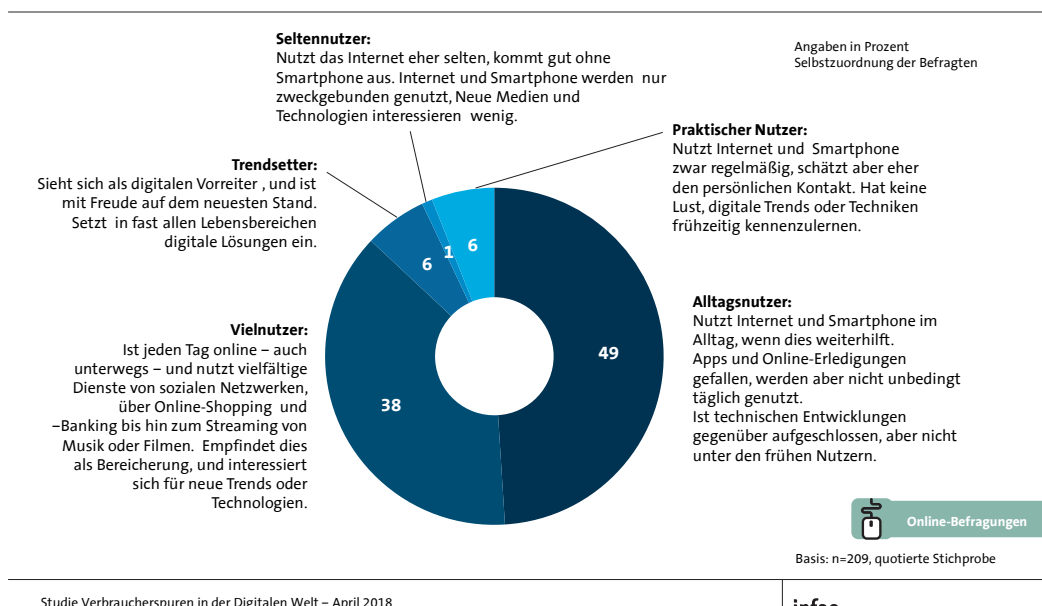
Die nachfolgend dargestellten Ergebnisse zur Charakterisierung der Projektteilnehmer basieren im Wesentlichen auf den unterschiedlichen Online-Befragungsmodulen des Projekts und den objektivierbaren Daten, die durch die App Tracy aufgezeichnet wurden.

Für das Modell des Datenschattens handelt es sich hierbei um erklärende Merkmale, die ursprünglich zu dem Zweck erhoben wurden, den aus den Selbstauskünften abzuleitenden Datenschatten daraufhin zu untersuchen, ob sich spezifische Einflussgrößen auf diesen identifizieren lassen. Eine empirische Ableitung des individuellen Datenschattens, in die Teile der vorgestellten Angaben der Befragten einfließen, nimmt später Abschnitt 5.4 vor.

5.2.1 Online-Aktivitäten

Bevor im Detail auf die unterschiedlichen Facetten der Online-Nutzung als Quelle der Herausgabe von persönlichen Daten eingegangen wird, soll ein Blick auf die Selbstzuschreibung der Befragten erfolgen: Wie sehen sich die Befragten, welchem der vorgestellten Online-Nutzungstypen vom Seltennutzer bis hin zum Trendsetter ordnen sie sich selbst zu? Die Ergebnisse weisen auf eine für diese Zielgruppe bereits starke Online-Durchdringung hin. Immerhin 9 von 10 Befragten bezeichnen sich als Alltags- oder Vielnutzer. Eingedenk der Vermutung, dass es sich bei der Gruppe der Projektteilnehmer eher um aufgeklärte und interessierte Personen handeln dürfte, können die Resultate als Ausblick auf zukünftige Entwicklungen der Gesamtbevölkerung verstanden werden.

Abbildung 8 Selbstzuschreibung Online-Typ

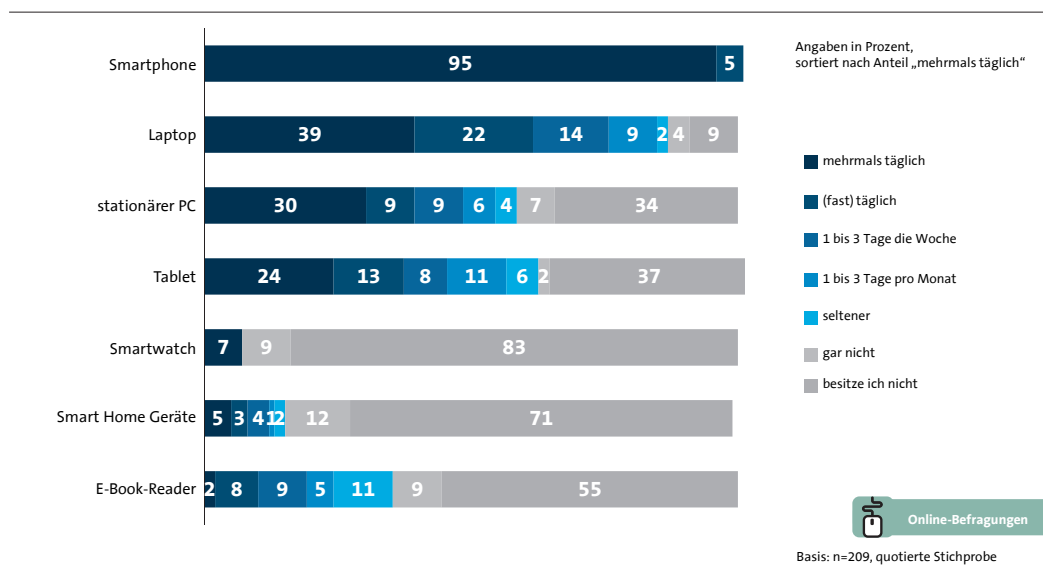


Ein zentrales Stichprobenmerkmal betraf die Smartphone-Nutzung. Alle Teilnehmer der Studie sollten ein solches Gerät mit Betriebssystem Android von Google (mindestens Version 4.1.x) oder iOS von Apple (mindestens

Version iOS-Version 9.0) besitzen und nutzen, unabhängig davon, wie intensiv und zu welchen Zwecken dies der Fall war²⁸. Im Überblick über die genutzten technischen Geräte dieser Stichprobe zeigt sich dementsprechend, dass das Smartphone als ständiger Begleiter ganz überwiegend auch mehrmals täglich privat genutzt wird. Dies gilt im Übrigen geschlechterübergreifend. Im Blick auf das Alter bestätigt sich der noch immer erwartbare Trend einer leicht geringeren Smartphone-Nutzung mit steigendem Lebensalter.

Alle übrigen internetfähigen Geräte von Laptop hin zu Tablet PC werden erheblich seltener genutzt. Weitere technische Optionen des vernetzten Lebens werden nicht täglich, aber wie im Falle des E-Book-Readers doch gelegentlich wahrgenommen, sofern sich ein solches Gerät im Besitz befindet.

Abbildung 9 Private Nutzung technischer Geräte



Ergebnispräsentation Studie Verbraucherspurten in der Digitalen Welt - April 2018

infas

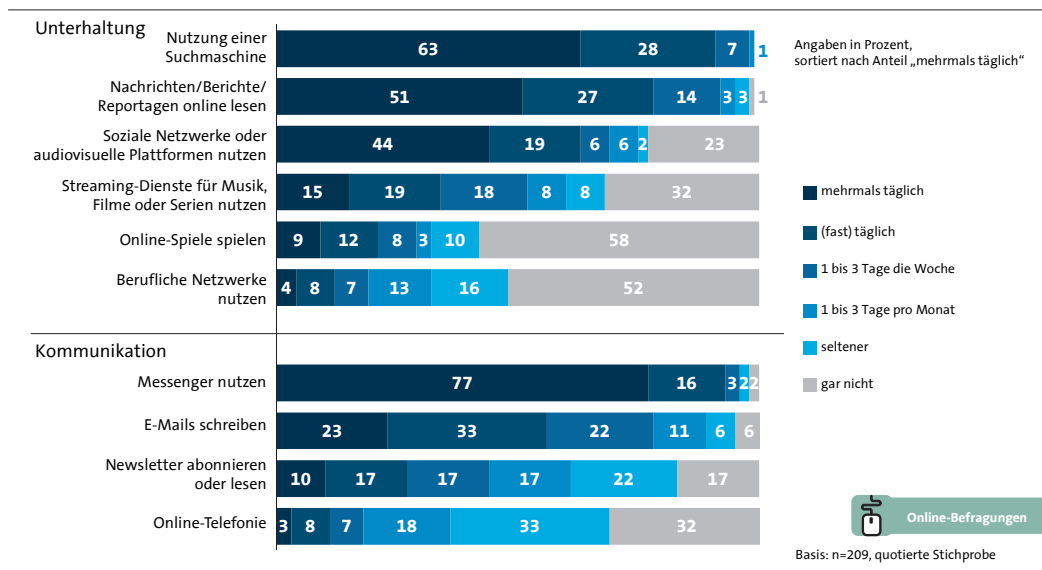
Die folgenden Abbildungen stellen ausgewählte private Online-Aktivitäten der Befragten nach inhaltlichen Themenbereichen vor. Unterhaltung und

²⁸ Zusätzlich waren aus Vergleichsgründen bewusst vier „analoge“ Probanden ohne ein Smartphone im Sample.

Kommunikation stellen hierbei die wichtigsten Anlässe für Online-Aktivitäten dar. Insbesondere den Suchmaschinen und Messenger-Diensten kann ein hoher Aufforderungscharakter zugeschrieben werden. Dement-sprechend berichten jeweils mehr als 90 Prozent der Befragten über eine etwa tägliche Nutzungsfrequenz – darunter ist die überwiegende Mehrheit mehrmals täglich aktiv. Aber auch soziale Netzwerke und audiovisuelle Medien werden von 70 Prozent der Befragten immerhin noch wöchentlich oder öfter genutzt.

In den Kommunikationsmöglichkeiten haben Messenger allen anderen Kanälen den Rang abgelaufen. Immerhin etwa die Hälfte der Befragten schreibt dennoch fast täglich private E-Mails.

Abbildung 10 Private Online-Aktivitäten zu Unterhaltung und Kommunikation



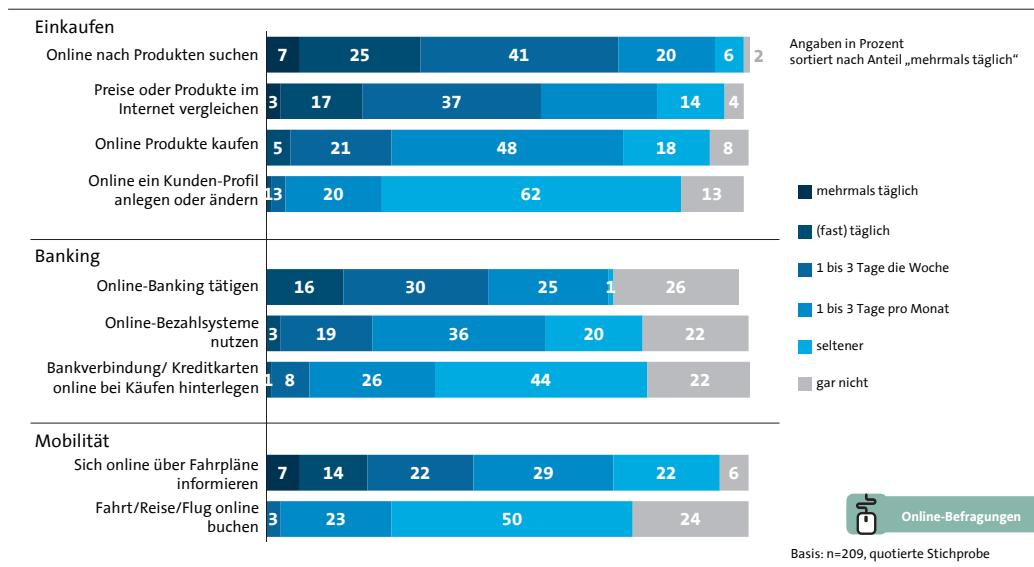
Ergebnispräsentation Studie Verbraucherspurten in der Digitalen Welt - April 2018

infas

Weitere Online-Aktivitäten zu Einkaufen, Banking und Mobilität sind weniger relevant, werden jedoch nur von jeweils kleinen Minderheiten gar nicht in Anspruch genommen. Der grundsätzliche Durchdringungsgrad ist also hoch, wenngleich die Nutzungshäufigkeit (derzeit noch) geringer ausfällt. Im Vergleich der mindestens wöchentlichen Nutzung sticht vor allem die Online-Produktsuche heraus, die 7 von 10 Befragten wöchentlich oder

öfter betreiben. Online-Banking wiederum ist nicht täglich relevant, aber substantiell für etwa 70 Prozent der Befragten mindestens an einigen Tagen im Monat aktuell. Online-Fahrplaninformationen schließlich sind ebenfalls weit verbreitet, aber entsprechend der Entscheidungen zur Verkehrsmittelwahl eher nicht zwingend täglich erforderlich.

Abbildung 11 Private Online-Aktivitäten zu Einkauf, Banking und Mobilität

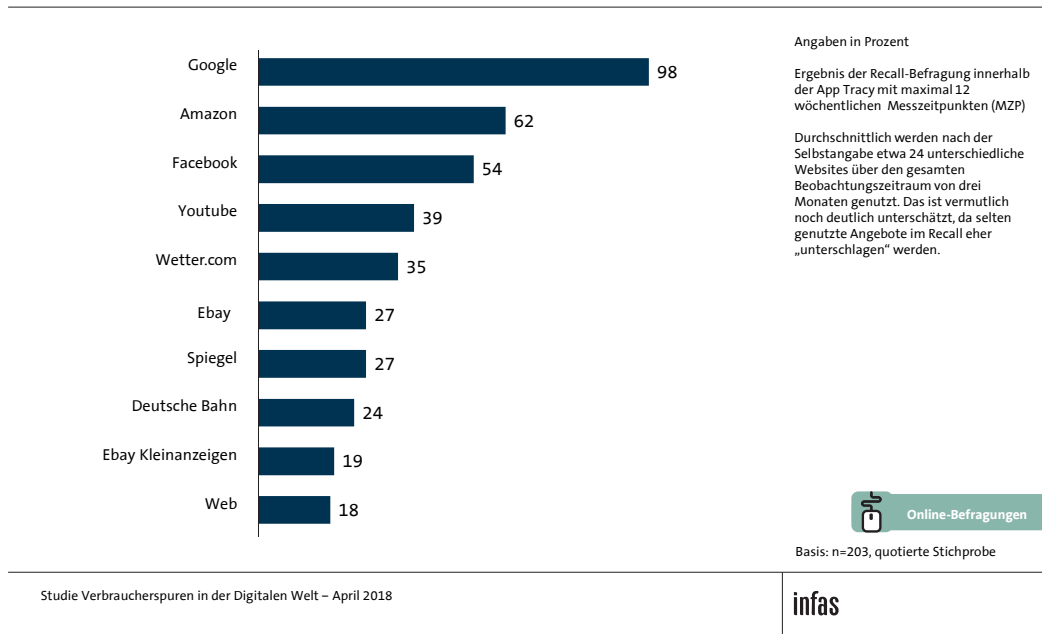


Ergebnispräsentation Studie Verbraucherspuren in der Digitalen Welt - April 2018

infas

Die untersuchten Internetaktivitäten – geräteunabhängig erfragt, also sowohl per Smartphone als auch mittels Laptop, Tablet etc. – weisen einzelne Anbieter mit stark überdurchschnittlicher Nutzung aus, wie folgende Abbildung zeigt. Nahezu unvermeidbar erscheint die Inanspruchnahme von Google, aber Amazon und Facebook stehen bei den Befragten ebenfalls hoch im Ranking der genutzten Internetseiten. Dargestellt ist das Ergebnis der Recall-Befragungen aus der App Tracy heraus, die unter anderem aus einer wöchentlichen Abfrage der im Laufe der jeweils letzten 24 Stunden genutzten Websites bestand.

Abbildung 12 Genutzte Internetseiten (Basis: Recall-Befragung)



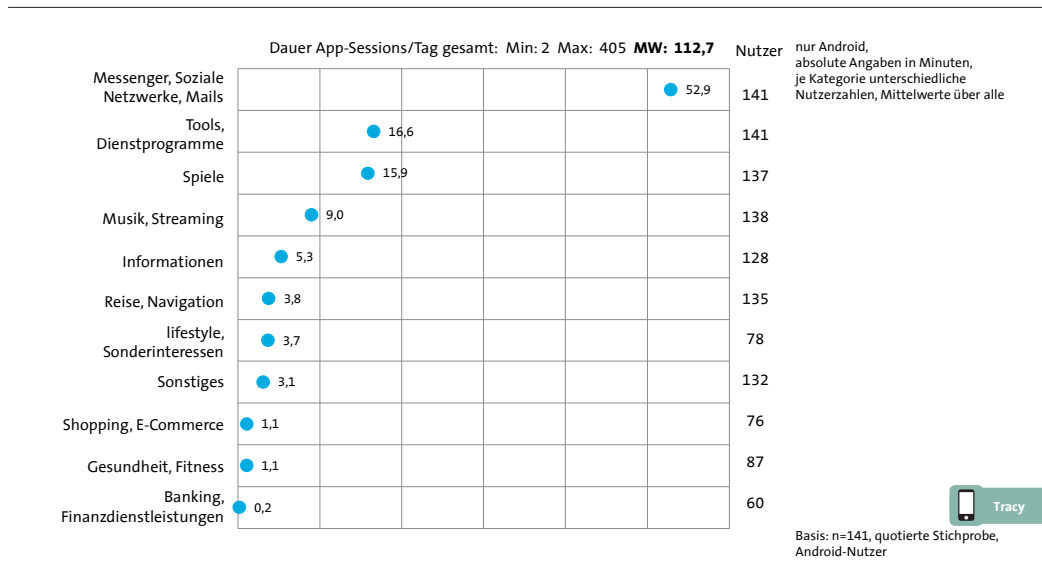
Ebenfalls aus der App Tracy resultieren Erkenntnisse über die App-Prozesse, die auf den Smartphones der Beteiligten registriert wurden. Die App Tracy misst als „App-Prozesse“ technisch bedingt alle Anwendungen und Dienstprogramme auf dem Smartphone – auch solche, die nicht über eine direkte Nutzerschnittstelle verfügen. Hinzu kommt eine größere Zahl ständig aktiver Basisprozesse. Dabei ist die Auslese aus technischen Gründen für iOS und für Android in sehr unterschiedlichem Umfang möglich bzw. führt im Einzelfall zu sehr unterschiedlichen Ergebnissen.

Im Schnitt finden auf den Smartphones der Befragten mit Android-Betriebssystem täglich etwa 113 Minuten lang App-Prozesse statt. Männer und Frauen zeigen hier die gleiche Nutzungsdauer. Im Vergleich der Altersgruppen sinkt die Nutzung von 124 Minuten unter den 18- bis 25-Jährigen bis hin zu 80 Minuten bei den ab-70-Jährigen.

Im Schnitt entfällt fast die Hälfte dieser Dauer, nämlich 53 Minuten, bereits auf die Nutzung von Messenger-Diensten, sozialen Netzwerken und E-Mails. Weitere relevante Anlässe der App-Prozesse sind beispielsweise Tools und Dienstprogramme oder Spiele – deren Nutzungsdauern mit im

Schnitt lediglich etwa einer Viertelstunde täglich deutlich niedriger anzu-
setzen sind.

Abbildung 13 Tägliche Dauer der App-Prozesse nach Inhalten



Studie Verbraucherspuren in der Digitalen Welt – April 2018

infas

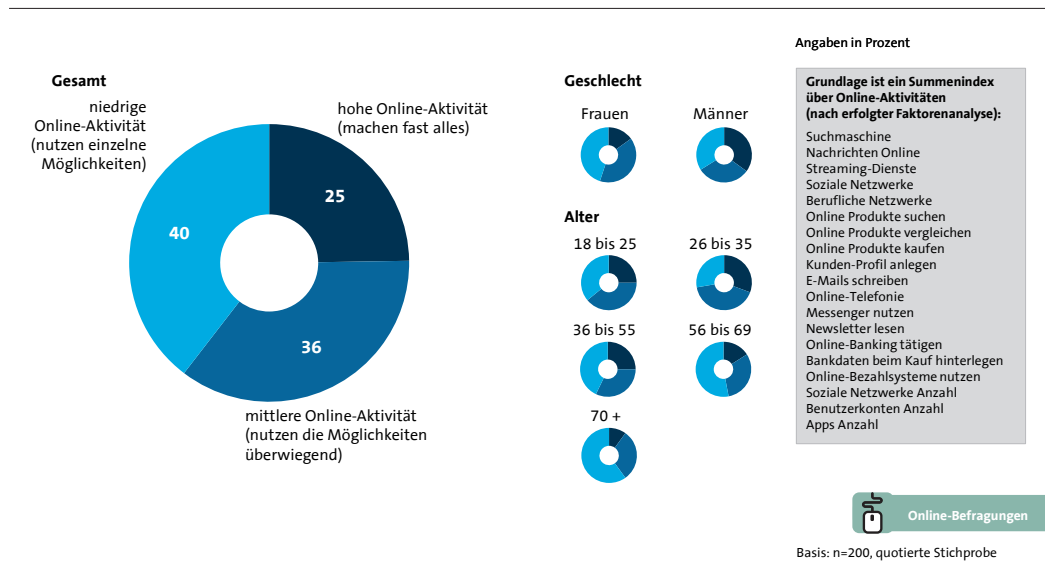
Ein interessantes Zusatzergebnis spannt sich auf, wenn die Anzahl der genutzten App-Prozesse in die Betrachtung einbezogen wird: Unter dieser Perspektive zeigt sich, dass auf den Smartphones der befragten Männer offenbar durchschnittlich mehr App-Prozesse stattfinden als bei den Frauen, dass dies aber nicht zu einer längeren Gesamtdauer der Smartphone-Prozesse führt. Insbesondere im Bereich der Kommunikation erscheint vorstellbar, dass auf den Smartphones der Frauen im Schnitt weniger, aber längere Prozesse stattfinden als auf den Smartphones der Männer.

Die Anzahl der genutzten Apps geht zusammen mit anderen Kriterien in die nachfolgend beschriebene **Profilbildung zu Online-Aktivitäten** ein. Sie stellt ein zusammenfassendes und verdichtetes Maß für die Intensität der Online-Aktivitäten dar. Nach erfolgter multivariater Analyse fließen die berichteten Aktivitäten zu Unterhaltung, Kommunikation, Einkaufen, Banking und Mobilität ein, aber auch die Anzahl der sozialen Netzwerke, Benutzerkonten und genutzter Apps – letztere aus der App Tracy.

Zur Abbildung der Profile wurden die Befragten nach ihrer Nutzungsintensität bei den letztgenannten Maßnahmen in drei Gruppen klassifiziert: in eine hohe, mittlere und niedrige Online-Aktivität, gemessen an der Referenz der durchschnittlichen Aktivität der hier zugrunde liegenden Stichprobe. Diese Gruppierung dient dem Vergleich bezogen auf inhaltliche Merkmale und erlaubt eine leichtere Gegenüberstellung unterschiedlicher soziodemografischer Gruppen – in der nachfolgenden Abbildung exemplarisch dargestellt für Geschlecht und Alter.

Unter den Befragten lässt sich etwa ein Viertel der Gruppe mit hoher Online-Aktivität zuordnen. Ein gutes Drittel weist einen mittleren Aktivitätsgrad auf. Die restlichen 40 Prozent sind schließlich den online wenig Aktiven zuzurechnen. Männer weisen in dieser Stichprobe einen deutlich höheren Online-Aktivierungsgrad auf als Frauen. Im Sample sind Personen ab dem Alter von etwa Mitte 50 weniger aktiv, wobei die Altersgruppe mit der höchsten Aktivität nicht unter den Jüngsten, sondern unter den 26- bis 35-Jährigen zu finden ist.

Abbildung 14 Profilbildung zu den Online-Aktivitäten

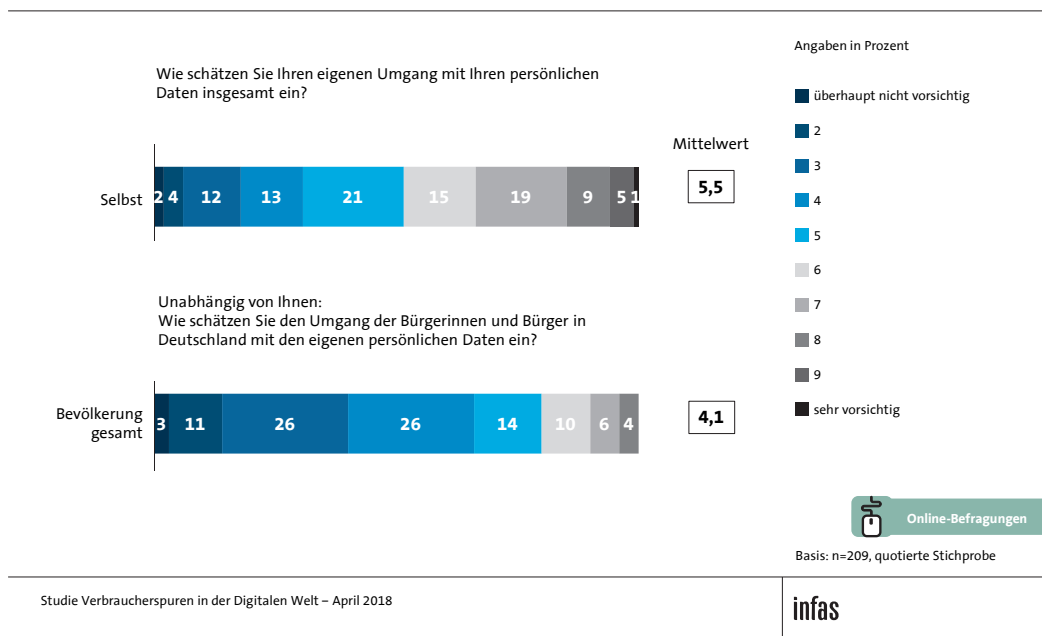


5.2.2 Datenschutzaktivitäten

Im Folgenden sind Auszüge der Ergebnisse dargestellt, die sich unter anderem mit den Einschätzungen der Teilnehmer zum eigenen Umgang mit persönlichen Daten sowie mit ihren Einstellungen im Hinblick auf den Datenschutz beschäftigen.

In einer ersten Bewertung schätzen die Teilnehmer ihre eigene Datenschutzsensibilität höher ein als die Datenschutzsensibilität ihrer Mitmenschen. Ein Drittel empfindet den eigenen Umgang als (sehr) vorsichtig (Werte 7 bis 10 auf einer 10er-Skala), während nur 10 Prozent der Befragten den Umgang der Gesamtbevölkerung in Deutschland als ähnlich vorsichtig bezeichnen. Damit bestätigt sich die Tendenz zu einer positiv verschobenen Eigenwahrnehmung. Nicht auszuschließen ist allerdings, dass die selbstselektierte Stichprobe tatsächlich sensibler ist als der Bevölkerungsschnitt.

Abbildung 15 Selbst- und Bevölkerungswahrnehmung zur Datenschutzsensibilität

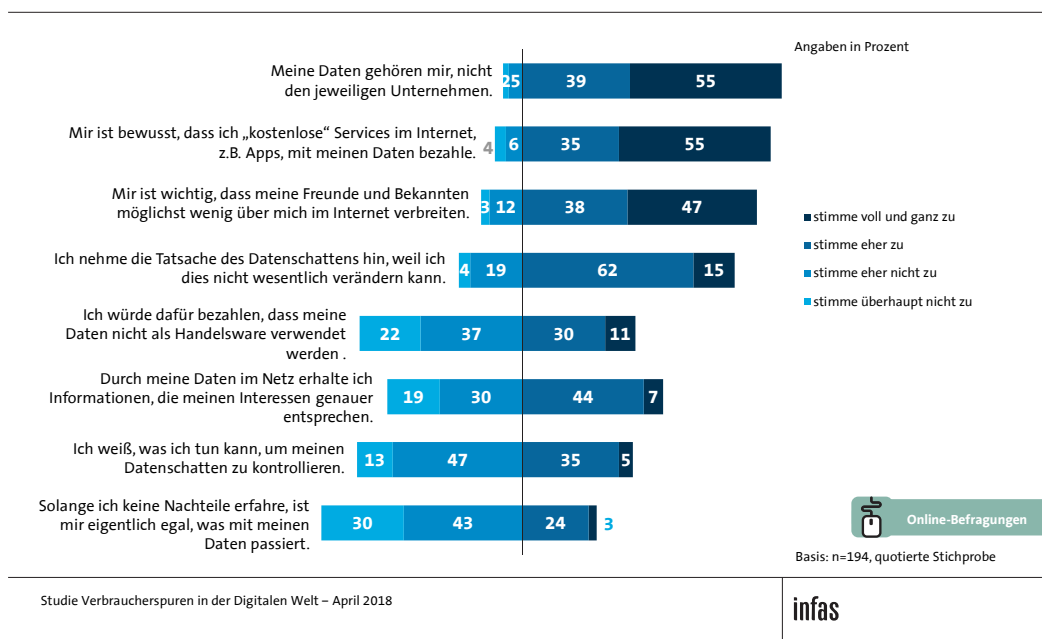


Dabei betragen die Unterschiede im Mittelwert lediglich einen Punktwert Differenz auf der 10er-Skala von „überhaupt nicht vorsichtig“ bis „sehr vorsichtig“. Auch der Gesamteindruck der Selbsteinschätzungen fällt mit

einem Mittelwert von 5,5 noch eher moderat aus. Demnach schreibt sich selbst diese in Datenschutz vermutlich eher aufmerksame Zielgruppe nur mittlere Datenschuttsensibilität zu.

Trotzdem drückt sich in den Zustimmungsraten zu den folgenden Statements durchaus ein überwiegendes Bewusstsein (90 Prozent) für die Mechanismen der Transaktion von Daten und Services im Netz aus. Die Tatsache des Datenschattens selbst wird von der Mehrheit (77 Prozent) indes als unbeeinflussbar hingenommen, zumal eher Unwissenheit über mögliche eigene Maßnahmen zur Reduktion des Datenschattens zu konstatieren ist (60 Prozent).

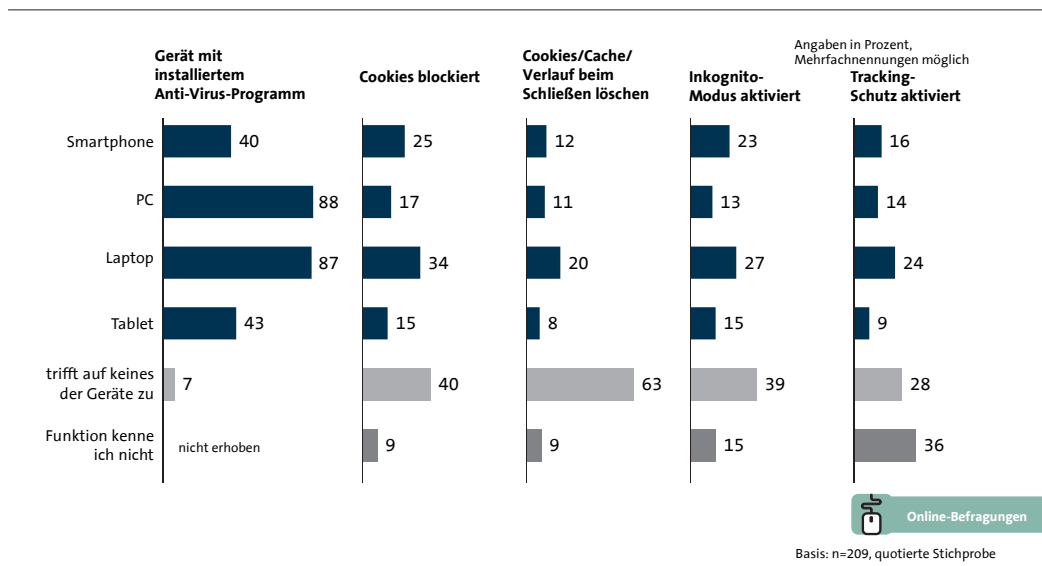
Abbildung 16 Meinungen zu den eigenen Daten im Netz



Fast alle Befragten verwehren sich gegen eine Überlassung des Besitzanspruchs ihrer Daten an Unternehmen (94 Prozent) – es ist wiederum aber auch für drei Viertel der Befragten nicht relevant, was mit den Daten passiert, solange sie keine Nachteile erfahren. Eine Zahlungsbereitschaft für die mögliche Verhinderung der kommerziellen Nutzung eigener Daten durch Dritte ist derzeit bei immerhin etwa 40 Prozent vorhanden.

Im Hinblick auf das konkrete **eigene Datenschutzverhalten** wurde eine Vielzahl an Möglichkeiten im Projekt untersucht. Bezogen auf bekannte und genutzte Browser- und Sicherheitseinstellungen bei privat genutzten Geräten zeigt sich die enorme Bandbreite der Aktivitäten von den fast durchgängig genutzten Anti-Virus-Programmen bei PC und Laptop über die noch vergleichsweise häufig praktizierte Blockierung von Cookies bis hin zu den wenig genutzten Funktionen des Löschens von Cookies bzw. Verlaufsdaten, der Aktivierung von Inkognito-Modus oder Tracking-Schutz. Letztere Option ist zudem zur relativ Wenigen bekannt.

Abbildung 17 Browser- und Sicherheitseinstellungen bei privat genutzten Geräten



Studie Verbraucherspuren in der Digitalen Welt – April 2018

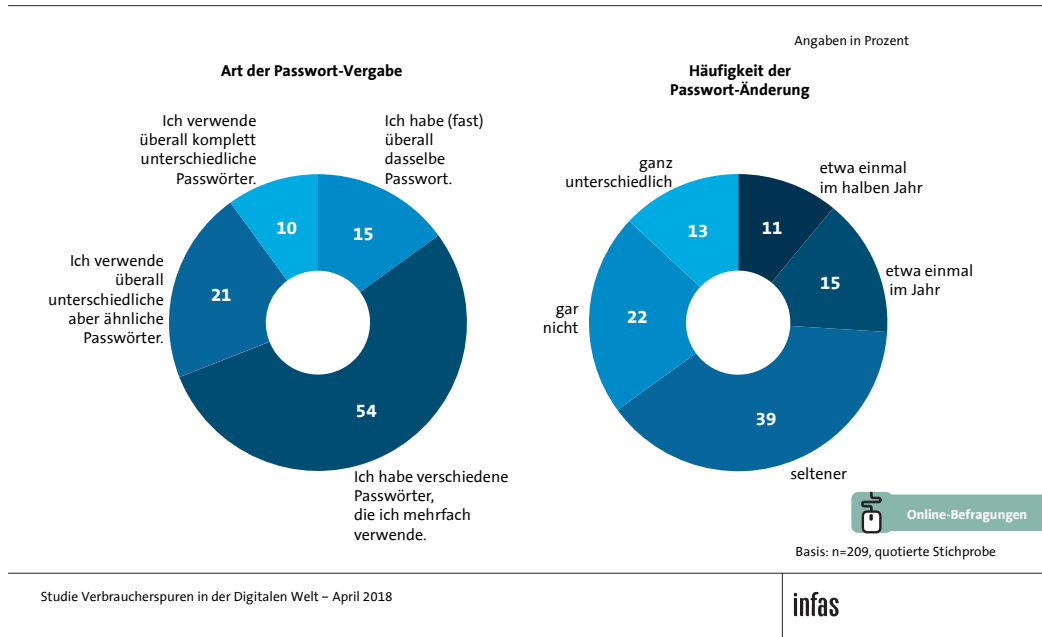
infas

Die meisten Maßnahmen werden, wie dargestellt, bezüglich des Laptops ergriffen. Anti-Virenprogramme werden darüber hinaus auch beim PC häufig in Anspruch genommen. Das Smartphone wird mit Ausnahme der Anti-Virenprogramme aktuell ähnlich geschützt wie der Laptop. Auf das Tablet beziehen sich die wenigsten entsprechenden Bemühungen.

Die **Passwortvergabe** als mögliches Zugangstor zur unbefugten Datenauslese wird mit verhältnismäßig geringer Aufmerksamkeit belegt, sowohl im Hinblick auf die Art der Vergabe als auch auf die Häufigkeit der Änderung.

Nur etwa ein knappes Viertel der Befragten verwendet überall unterschiedliche Passwörter. Ebenfalls nur ein etwa gutes Viertel ändert jährlich oder halbjährlich das Passwort.

Abbildung 18 Art und Häufigkeit der Passwort-Vergabe

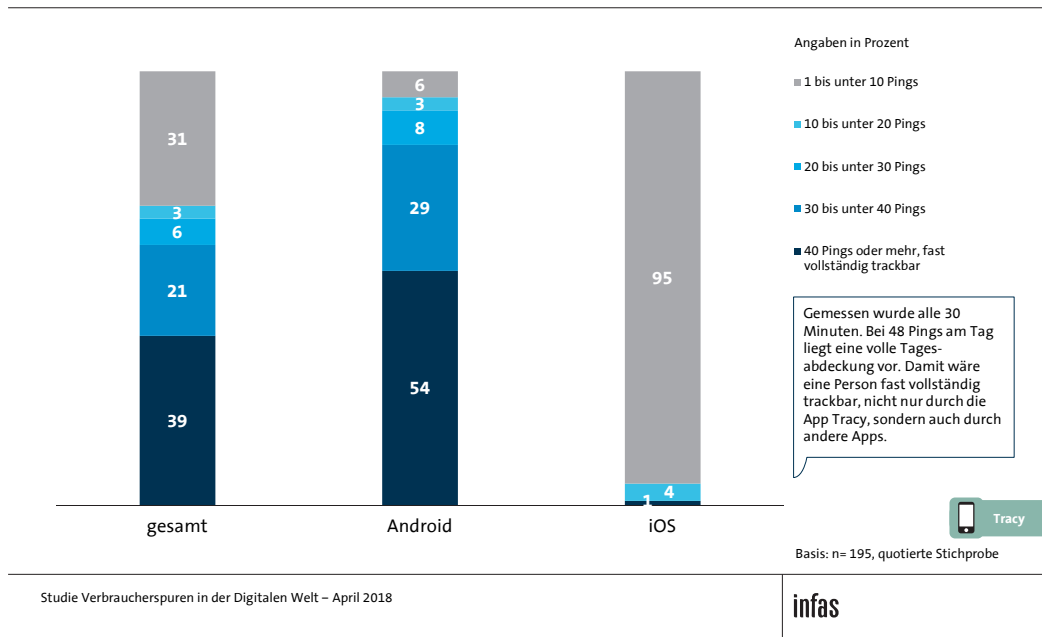


Auch die Art der Nutzung von **W-LANs** ist als ein Indikator für die Bemühungen um Datenschutz zu werten. So gaben etwa zwei Drittel der Befragten in der Online-Befragung an, öffentlich verfügbare, offene W-LANs zu nutzen. Im Detail gibt hierüber die App Tracy zusätzliche Informationen: Im Schnitt werden durchschnittlich 5,4 W-LANs über die gesamte Laufzeit der App Tracy genutzt. Nur unwesentlich geringer fällt mit einem Durchschnittswert von 4,4 die Anzahl der genutzten offenen W-LANS aus. Im Schluss deutet dies darauf hin, dass der Anteil der genutzten offenen W-LANS als sehr hoch zu betrachten ist.

Die Resultate der App Tracy zeigen ebenfalls eine hohe Trackbarkeit der Befragten über die **GPS-Locations** ihrer Smartphones. Gemessen wurden durch die App Tracy die sogenannten „Pings“ (Datenübertragungen, die Auskunft darüber geben, ob ein bestimmter Host in einem IP-Netzwerk erreichbar ist) in einem zeitlichen Abstand von jeweils 30 Minuten.

Bei 48 Pings am Tag liegt theoretisch eine volle Tagesabdeckung vor. Damit wäre eine Person fast vollständig durch die App Tracy trackbar, gleichermaßen auch durch andere Apps. Auffällig sind der hohe Anteil der weitgehend trackbaren Zielpersonen und die Unterschiede in der Nachverfolgbarkeit zwischen den Systemen iOS und Android.

Abbildung 19 Durchschnittliche Anzahl der GPS-Locations am Tag

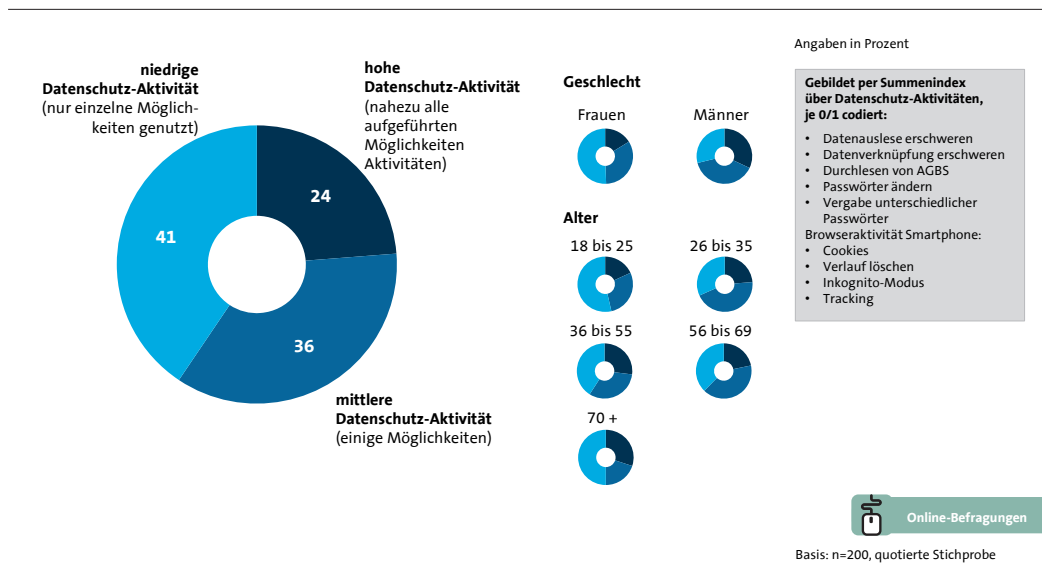


In analoger Vorgehensweise zur Bildung des Indexes bezüglich der Online-Aktivitäten wurde zusammenfassend ein weiteres Profil bezogen auf die **Datenschutzaktivitäten** gebildet. Dieser umfasst die vorgestellten Browser- und Sicherheitsaktivitäten bezüglich des Smartphones als dem in dieser Studie relevantesten Endgerät, sowie weitere Aktivitäten zur Reduktion des Datenschattens durch Erschwerung von Datenauslese und -verknüpfung.

In der Segmentierung ablesbar sind eher geringe bis mittlere Datenschutzaktivitäten. Nur das Viertel der Befragten in der Gruppe mit hoher Datenschutzaktivität nutzt nahezu alle der aufgeführten Optionen. Der aus den Online-Aktivitäten erwartbare Geschlechtereffekt tritt im Hinblick auf den Datenschutz noch einmal verschärfter zutage und weist die befragten Männer wiederum deutlich datenschutzaktiver aus als die teilnehmenden

Frauen. Im Vergleich der Altersgruppen lässt sich festhalten, dass mit steigendem Alter tendenziell höhere Datenschutzaktivitäten berichtet werden – also offenbar eine etwas höhere „Besorgnis“ besteht, obwohl diese Gruppe insgesamt weniger online aktiv ist.

Abbildung 20 Profilbildung zu Datenschutzaktivitäten



Studie Verbraucherspuren in der Digitalen Welt – April 2018

infas

In einer Gegenüberstellung von Datenschutzaktivitäten einerseits und Online-Aktivitäten anhand der jeweiligen Indizes andererseits zeigt sich, dass die Aktivitäten zum verstärkten Schutz der persönlichen Daten im Netz durchaus einhergehen mit der grundsätzlichen Online-Aktivität.

Tabelle 1 Datenschutzaktivitäten nach Online-Aktivitäten

Zeilen %		Datenschutzaktivitäten		
		niedrig (n = 81)	mittel (n = 72)	hoch (n = 47)
Online-Aktivitäten	niedrig (n = 79)	46%	37%	18%
	mittel (n = 72)	46%	33%	21%
	hoch (n = 49)	24%	39%	37%

Befragte mit eher geringen Online-Aktivitäten weisen auch vergleichsweise geringere Bemühungen um Datenschutz auf. In dieser Gruppe entfalten 46 Prozent der Befragten niedrige Datenschutzaktivitäten und lediglich 18 Prozent hohe Aktivitäten.

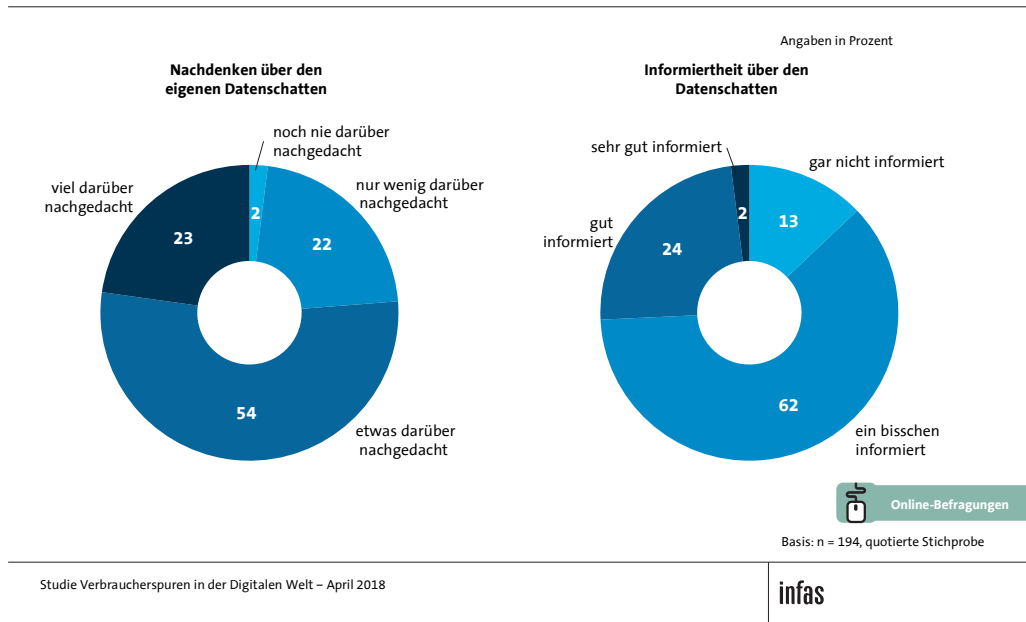
Befragte mit hoher Online-Aktivität tendieren wiederum auch zu stärkeren Datenschutzbemühungen. Hier sind lediglich 24 Prozent gering im Datenschutz aktiv und 37 Prozent weisen hohe Datenschutzaktivitäten auf. Die Gruppen mit den mittleren Profilausprägungen nehmen auch inhaltliche Mittelstellungen ein.

5.2.3 Vorstellungen vom Datenschatten

Das folgende Kapitel befasst sich mit den Fragen, welche Vorstellungen und Meinungen die Studienteilnehmer von ihrem und dem allgemeinen Datenschatten sowie von den diesbezüglichen Einflussmöglichkeiten haben. Wie in den vorangegangenen Kapiteln geben die Antworten der Projektteilnehmer Einblicke in die Befindlichkeiten eines thematisch besonders interessierten Samples, ohne Anspruch auf Repräsentativität für die Gesamtbevölkerung.

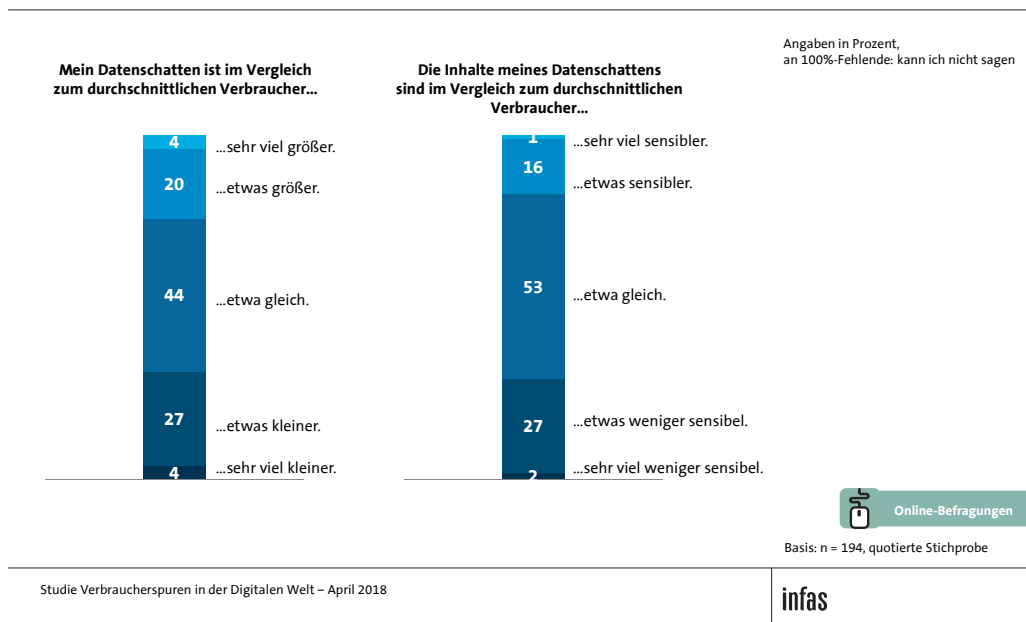
Für drei von vier Befragten war das Thema Datenschatten auch bereits vor der Studie gedanklich präsent. Nur etwa ein Viertel hat sich noch wenig oder nie damit befasst. Etwa die Hälfte der Befragten haben dies „etwas“ getan und ein weiteres Viertel hat sich bereits ausgiebiger damit beschäftigt. Ebenfalls drei von vier Befragten halten sich hierzu jedoch für lediglich „ein bisschen“ bis „gar nicht“ informiert. Unter dem übrigen Viertel der gut Informierten betrachten sich schließlich nur 2 Prozent als sehr gut informiert. Wenn das Nachdenken und die subjektive Informiertheit über den Datenschatten als Indikatoren für die thematische Nähe betrachtet werden, scheint ein adäquater Überblick über das Thema Datenschatten demnach selbst bei intensiverer Beschäftigung nicht leicht zugänglich.

Abbildung 21 Nachdenken und Informiertheit über den eigenen Datenschatten



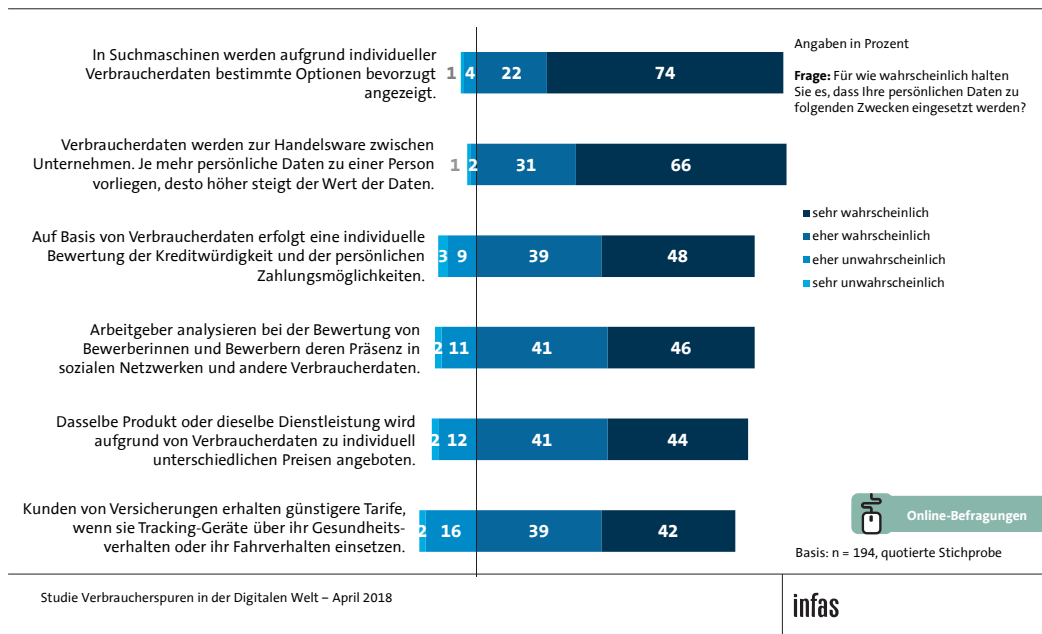
In Vergleich zu einem hypothetischen Durchschnittsverbraucher stellt sich – so die folgende Abbildung – die Mehrheit der Befragten ihren Datenschatten als genauso groß oder kleiner vor bzw. als genauso sensibel oder weniger sensibel. Nur jeweilige Minderheiten vermuten ihren Datenschatten überdurchschnittlich groß oder halten dessen Inhalte für eher sensibel.

Abbildung 22 Vorstellungen von Größe und Inhalt des Datenschattens



Die geäußerten Vermutungen der Teilnehmer gehen dahin, dass ihre persönlichen Daten bereits jetzt umfassend eingesetzt werden und vielfältige Wirkungen zeigen. Dabei verschieben sich die Wahrscheinlichkeitsratings, mit der die untersuchten Folgewirkungen des Datenschattens als gegeben angenommen werden, nur marginal. So wird beispielsweise der Einsatz der eigenen Daten als Filterfunktion in Suchmaschinen von drei Vierteln der Befragten als sehr wahrscheinlich angenommen. In fast ähnlicher Größenordnung halten die Befragten den Austausch von personenbezogenen Daten als Handelsware zwischen Unternehmen für gegeben. Auch andere Einsatzmöglichkeiten des Datenschattens etwa in den Bereichen Bonitätsprüfung, Arbeitsplatzsuche, Preisgestaltung und Versicherungstarifen werden von den Befragten zu hohen Anteilen als gegeben unterstellt.

Abbildung 23 Wissen über die Verwendung der eigenen Daten

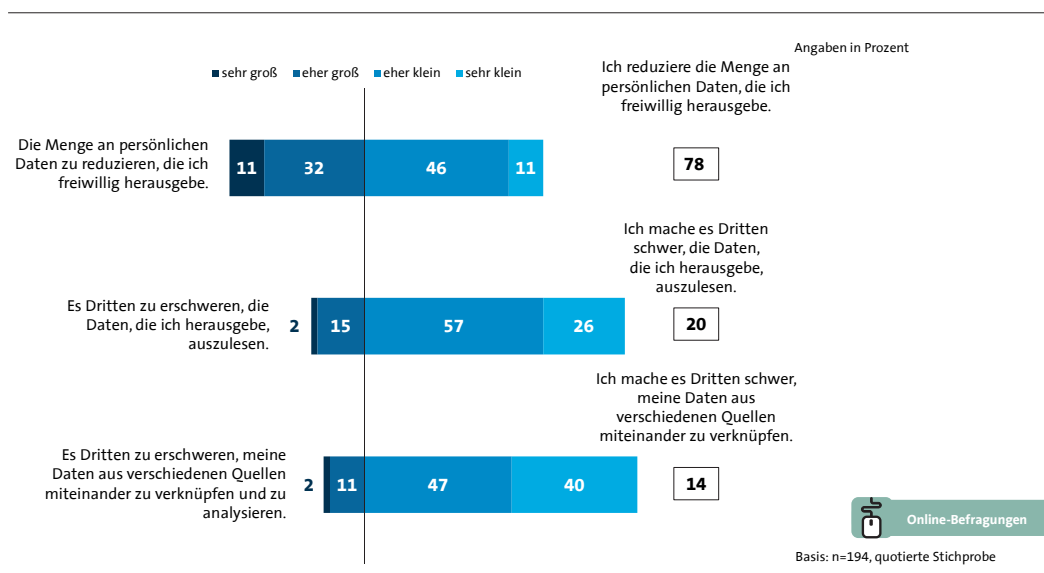


In der Summe weisen die geäußerten Vermutungen der Befragten auf einen ahnungsvollen Blick auf den eigenen Datenschatten und dessen Konsequenzen.

Können Verbraucher diese Informationen jedoch umsetzen, bzw. geeignete Maßnahmen zur Eindämmung ihres Datenschattens ergreifen? In diesem Zusammenhang wurden den Befragten die drei zentralen Ansatzpunkte für mögliche Maßnahmen – Datenherausgabe, Datenauslese und Datenverknüpfung – vorgestellt (siehe Abbildung 24). Sie wurden dazu aufgefordert, den jeweiligen Handlungsspielraum einzuschätzen und ihr tatsächliches Verhalten zu benennen. Insgesamt zeigt sich, dass die Befragten – wenn überhaupt – fast nur im Hinblick auf die Herausgabe der eigenen Daten Spielraum erkennen. Die Datenauslese oder -verknüpfung scheint sich der persönlichen Kontrolle nach ihrer Einschätzung komplett zu entziehen.

Wo größerer Spielraum gesehen wird, wird dieser auch häufiger umgesetzt. So berichten fast 80 Prozent, dass sie bereits die Daten reduzieren, die sie über sich persönlich herausgeben. Zu den übrigen Ansatzpunkten findet jedoch kaum eine Umsetzung statt. Hier liegen die Anteilswerte lediglich bei 20 bzw. 14 Prozent.

Abbildung 24 Einflussmöglichkeiten auf den eigenen Datenschatten



Schließlich lassen sich die Angaben der Befragten aus einer ergänzend gestellten offenen Nachfrage zum Thema Datenreduktion zu folgenden Maßnahmenbündeln zusammenfassen:

- die radikale Vermeidung von sozialen Netzwerken, Messengern und Online-Käufen,
- die weitgehende Reduktion von Nutzer- und Kundenprofilen,
- das Anlegen von Profilen mit Fake-Daten,
- der verbesserte Schutz der Privatsphäre durch Browser- und Einstellungsoptionen.

Damit werden eher alltags- und sozialfeindliche Vorstellungen formuliert. Sie erscheinen zudem nur bedingt praxistauglich und lassen auf Verbraucherseite deutliche Unsicherheiten in der Frage nach adäquaten, nicht interaktionsmindernden Maßnahmen vermuten.

5.2.4 Kontakte zu den ausgewählten Unternehmen der Selbstauskunft

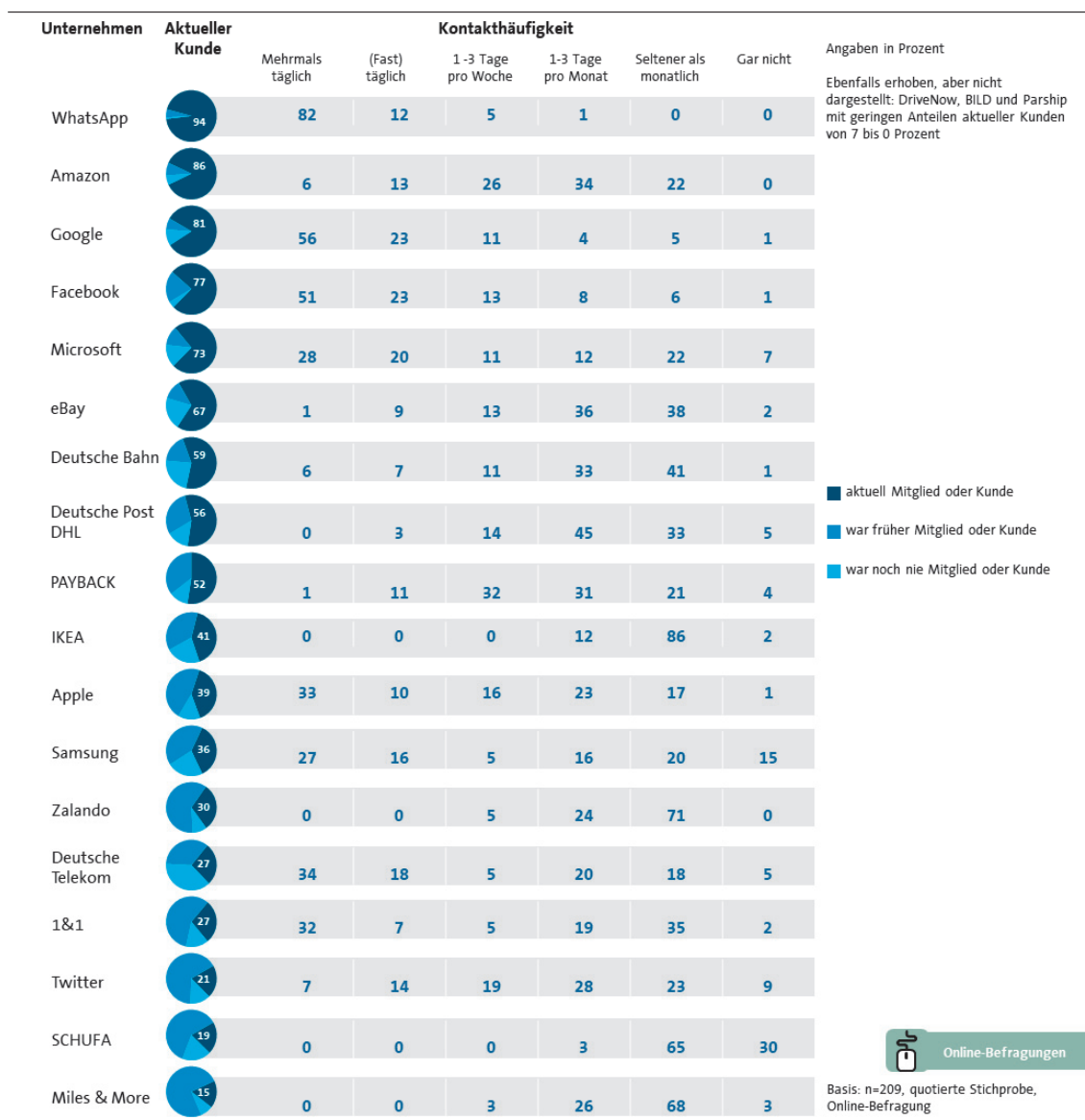
Um einen analytischen Zusammenhang zwischen Verbraucherverhalten und beantragten Selbstauskünften herzustellen, wurde bei den Probanden unter anderem erhoben, zu welchen der ausgewählten Unternehmen eine aktuelle oder frühere Kundenbeziehung bzw. ein Mitgliedsstatus vorlag. Dies wurde für eine Auswahl von 21 der 30 Unternehmen untersucht, zu denen derartige Kundenbeziehungen aktiv und für den Verbraucher auch wahrnehmbar eingegangen werden können.²⁹

Zu WhatsApp, Amazon, Google und Facebook befinden sich je etwa 80 bis über 90 Prozent der Teilnehmer sowohl in aktuellen Kundenbeziehungen als auch hoher Kontakthäufigkeit (siehe Abbildung 25). Nur geringe Anteile der Befragten sind wissentlich noch nie eine Kundenbeziehung zu diesen

²⁹ Nicht in diese Erhebung des aktiven Kundenkontakts einbezogen wurden die Unternehmen Acxiom, AdAudience, AZ Direct, ComScore, Creditreform Boniversum, Deutsche Post Direkt, informa HIS, Schober und Semasio.

Unternehmen eingegangen. Auch zu Microsoft und eBay berichten beispielsweise je etwa drei Viertel der Befragten über aktuelle Kundenbeziehungen. Zur Deutschen Bahn, Deutschen Post und zu Payback stehen immerhin noch mehr als die Hälfte der Befragten nach ihrer Einschätzung in aktueller Beziehung.

Abbildung 25 Kontakt zu ausgewählten Unternehmen



Studie Verbraucherspuren in der Digitalen Welt – April 2018

infas

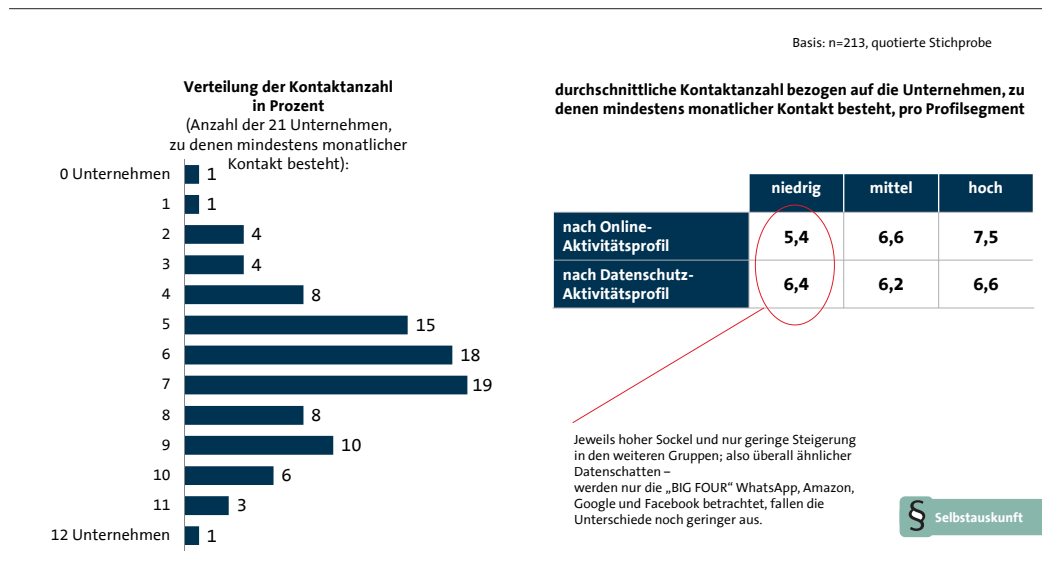
Im Sinne des Beitrags für den individuellen Datenschatten ist jeweils nicht nur der Anteil der aktuellen, sondern auch früherer Kundenbeziehungen in

die Betrachtung einzubeziehen. In der Summe wird zu nahezu allen Unternehmen von jeweils mindestens drei Vierteln der Befragten eine aktuelle oder frühere Kundenbeziehung angegeben. Einzige Ausnahmen – gemäß der Selbsteinschätzung – sind die Deutsche Telekom, sowie die in der vorangehenden Abbildung 25 nicht dargestellten Unternehmen mit sehr kleinen Kundenanteilen (Drive Now, Bild und Parship).

Damit bestätigte sich nochmals die Auswahl der Unternehmen für die Anträge auf Selbstauskunft. Ihre Zusammenstellung ließ umfassende Datenlieferungen erwarten (siehe hierzu Kapitel 5.3).

Vor einer Betrachtung der Unternehmensauskünfte selbst sollen abschließend die berichteten aktuellen Unternehmenskontakte in einer weiteren Zusammenfassung verdichtet werden.

Abbildung 26 Profilbildung zur Kundenbeziehung



Studie Verbraucherspurten in der Digitalen Welt – April 2018

infas

Gezählt wurden Unternehmen, zu denen nach Angaben der Probanden ein mindestens monatlicher Kontakt besteht. Wie die nachfolgende Abbildung zeigt, ist die durchschnittliche Anzahl der kontaktierten Unternehmen durchaus abhängig von der Intensität der Online-Aktivitäten, und steigt mit steigenden Online-Aktivitäten signifikant von einem Wert von 5,4 über 6,6

bis hin zu einem Wert von 7,5. Bei höherer Online-Aktivität stehen die Befragten üblicherweise also zu einer größeren Anzahl der 21 untersuchten Unternehmen in Kontakt.

Auffälliger ist jedoch, dass bereits Probanden mit geringer Online-Aktivität eine relativ hohe Kontaktintensität aufweisen – ebenso, wie ins Auge fällt, dass Befragte mit hohen Datenschutzaktivitäten keinen niedrigen Wert aufweisen.

Werden schließlich über alle Gruppen alleine die BIG FOUR Amazon, WhatsApp, Google und Facebook betrachtet, verschwinden die Unterschiede zwischen den Probanden fast ganz. Im Umkehrschluss bedeutet dies, dass diese Unternehmen den Datenschatten vermutlich in hohem Ausmaß dominieren.

Im Hinblick auf die Intensität der individuellen Datenschutzaktivitäten lässt sich indes kein vergleichbarer signifikanter Zusammenhang nachweisen; dort schwankt der Mittelwert in einem engen Wertebereich zwischen 6,2 und 6,6 und stellt sich weitgehend unabhängig von den Datenschutzaktivitäten dar. Für die Unternehmenskontakte scheint daher relativ unerheblich, ob die Befragten über hohe oder niedrige Bemühungen um den Schutz ihrer persönlichen Daten im Netz berichten. Dieses Ergebnis bestätigt sich in der empirischen Ableitung zum Datenschatten, die in Kapitel 5.4 vorgestellt wird.

5.3 Datensammlung und -anreicherung: Ergebnisse der Selbstauskünfte

5.3.1 Grundsätzliche Hinweise

In Kapitel 4.4 wurde die Konzeption und Anlage des Moduls zur Selbstauskunft bereits ausführlich erläutert. Die Resultate dieser insgesamt knapp 5.000 Anträge (Probanden x jeweils einbezogene Unternehmen) bei 30 Unternehmen fielen nicht wie erwartet aus. Insgesamt entsprachen 17 der 30 Unternehmen den Anträgen auf Selbstauskunft nicht, und verweigerten sich der Teilnahme an diesem zentralen Untersuchungsschritt. Sie entsprachen

damit nach unserer Interpretation sowie der des LDI NRW (siehe auch Kapitel 4.4) vermutlich nicht den gesetzlichen Verpflichtungen. Wo die Unternehmen Gründe nannten, bezogen sich diese in unterschiedlichen Facetten auf eine vermeintlich unzureichende Anwendbarkeit des § 34 BDSG.

Um mögliche juristische Hemmnisse zu umgehen, fand im Vorfeld sowie während des Verlaufs dieses Projektmoduls eine umfassende juristische Beratung statt. Hinzu kam im Projektverlauf wie schon berichtet eine Abstimmung mit dem LDI NRW, in deren Folge infas die Unterstützung für eine erneute Kontaktierung der Unternehmen erhielt. Der Auftraggeber war zu jedem Zeitpunkt über die Rückmeldungen der Unternehmen und die kontroverse Argumentationslage informiert. Details dazu wurden bereits in Kapitel 4.4 vorgestellt.

Auch die schriftliche Nachfassaktion im Januar 2018 bei jenen Unternehmen, die die Bearbeitung der Anträge zurückgewiesen hatten, erbrachte keinen Erfolg. In der Bilanz bearbeiteten lediglich 13 Unternehmen die Auskünfte in irgendeiner Form. Die erteilten Auskünfte wurden von den Unternehmen in den meisten Fällen, wie erbeten und mit den Probanden abgestimmt, an infas übersendet. Nur zu einem geringeren Anteil wurden diese entsprechend der ebenfalls angebotenen Möglichkeit direkt an die Teilnehmer verschickt, die die erhaltenen Unterlagen ihrerseits in den meisten Fällen infas zur Verfügung stellten.

Insgesamt ergab sich die aus Projektsicht bedauerliche, aber inhaltlich aufschlussreiche Einsicht, dass Quantität und Qualität des Datenschattens mit dem Instrument der informationellen Selbstauskunft nicht verlässlich empirisch ermittelt werden konnten. Ausschlaggebend dafür sind die hohe Verweigerungsrate der Unternehmen sowie vermutlich häufig lückenhafte bzw. erkennbar unzuverlässige Auskünfte, wie noch gezeigt wird.

Somit ist die Zielsetzung des Forschungsprojekts in dem Punkt gescheitert, den Datenschatten aufgrund gesetzlicher Auskunftsrechte bei 30 exemplarischen Unternehmen auf einer „gültigen“ Basis empirisch zu untersuchen. Die Summe möglichst aller Daten aus den 30 Unternehmen sollte den Da-

tenschatten abbilden. Inhaltlich sollte dieser in zahlreichen Fällen bestimmbar sein durch eine inhaltsanalytische Klassifikation und Auswertung der vorliegenden Daten. Der erzielte Rücklauf ließ aber weder valide noch reliable Auskünfte vermuten. Hinzu kamen zentrale Komplettausfälle wie etwa Google oder WhatsApp.

Für die vorliegenden Auskünfte wurden die Schritte der inhaltlichen Datenauswertung trotz der schmalen Datenlage in Teilen durchgeführt – nicht zuletzt, um transparent zu machen, dass die erlangte Datenqualität nicht für eine zuverlässige individuelle Ableitung des Datenschattens genügt. Die entsprechenden Ergebnisse werden im Folgenden dargestellt, obwohl die Auswertungen sich lediglich auf eine Teilmenge der benötigten Daten stützen konnten und daher in der Summe nicht die beabsichtigten Erkenntnisse lieferten.

In Folge der Unvollständigkeit kann allerdings nicht nur der bei 30 ausgewählten Unternehmen entstehende Datenschatten nicht adäquat beschrieben werden, sondern es entfallen auch potentielle weitere Analysemöglichkeiten. Insbesondere musste auf eine anonymisierte individuelle Auswertung pro Proband verzichtet werden, um nicht falsche oder unvollständige Ergebnisse als Datenschatten zu berichten. So kann weder die Frage verlässlich beantwortet werden, ob soziodemografische Merkmale signifikante Einflussgrößen darstellen, noch kann untersucht werden, ob das Verhalten der Befragten selbst im Hinblick auf Online- oder Datenschutzaktivitäten sich maßgeblich auf den Datenschatten auswirken.

5.3.2 Beschreibung der gelieferten Auskünfte

Grundsätzlich wurde infas von 176 Teilnehmern sehr umfassend zur Einholung der Selbstauskünfte bevollmächtigt. Von den insgesamt 4.959 gestellten Ersuchen entfielen aufgrund der individuellen Wahlmöglichkeiten der Teilnehmer zwischen 158 und 174 Ersuchen auf die einzelnen Unternehmen. Die folgende Tabelle 2 stellt im Überblick die Anzahl der jeweils

übermittelten und der vorliegenden bearbeiteten Ersuchen sowie den Versandmodus dar, über den die Rückmeldungen infas erreicht haben³⁰.

In nahezu vollständiger Form haben alle drei Auskunftsteien (Creditreform Boniversum GmbH, informa HIS GmbH, Schufa Holding AG) die Ersuchen bearbeitet. Hier entspricht die jeweilige Anzahl der bearbeiteten Anträge bis auf wenige Ausfälle derjenigen der gestellten Anträge. Ähnlich verhält es sich bei der United Internet AG, sowie bei der AdAudience GmbH, der Google Germany GmbH und der DriveNow GmbH & Co. KG. Wie später zu sehen ist, stehen dahinter jedoch sehr unterschiedliche inhaltliche Ausprägungen der bearbeiteten Ersuchen bis hin zu möglicherweise irreführenden Auskünften.

Bearbeitete Auskunftersuchen, die infas über die Teilnehmer erreichten, liegen häufig nicht in der vollständigen Anzahl vor, wie die direkt an infas versendeten Auskünfte. Dahinter kann im Einzelfall stehen, dass weitere Auskünfte an die Petenten versendet, aber nicht (mehr) an infas weitergeleitet wurden. Es kann jedoch auch der Fall sein, dass nur einzelne, positive Auskünfte versendet wurden, z. B. bei Vorliegen einer Kundenbeziehung oder anderen zutreffenden Merkmalen. Eine genaue Differenzierung der nicht spezifizierten Ausfälle entzieht sich bedingt durch die Abläufe und die freiwillige Partizipation der Probanden unserer vollständigen Kenntnis.

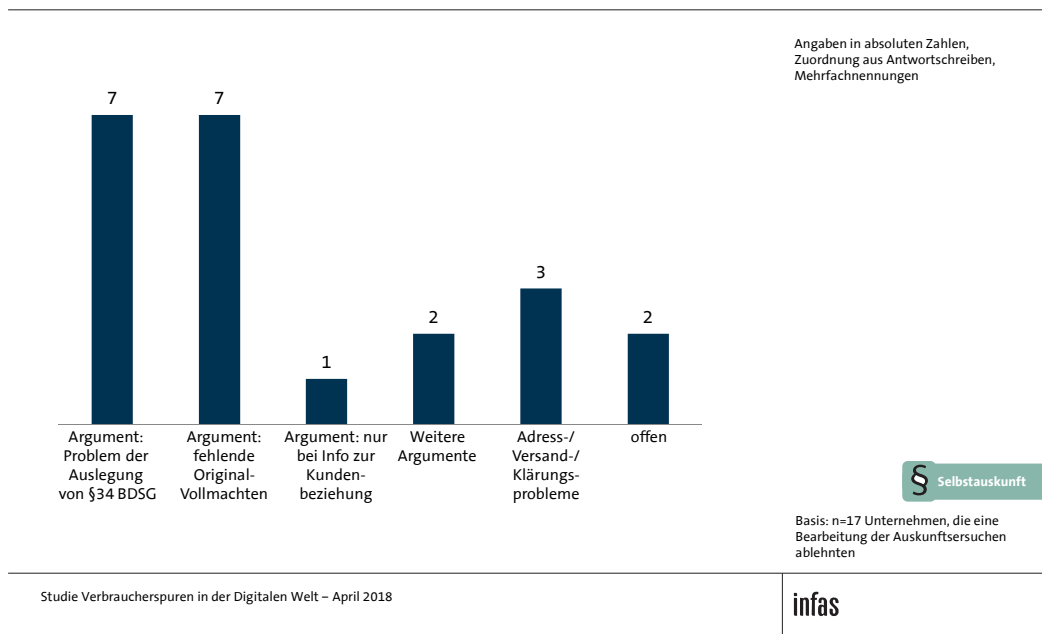
³⁰ In dieser Berichtsfassung, die zunächst als nicht zu Veröffentlichung bestimmte Version dem Auftraggeber übergeben wird, werden zur Transparenz die Unternehmensnamen genannt. Im Fall einer Veröffentlichung ist der Umgang hiermit noch abzustimmen und gegebenenfalls zu ändern, indem beispielsweise nur die Branchenbezeichnungen der jeweiligen Unternehmen angegeben werden.

Tabelle 2 Gestellte Auskunftersuchen und Resonanz

alle angefragten Unternehmen	Branche	gestellte Ersuchen durch infas	Absagen durch Unter- nehmen	bei infas vorliegende bearbeitete Ersuchen	Lieferung der bearbei- teten Ersu- chen an infas oder Teilnehmer
		abs.	abs.	abs.	
Gesamt		4.959	17 Unter- nehmen	1.346	
Creditreform Boniversum GmbH	Auskunftei	159		156	infas
informa HIS GmbH	Auskunftei	159		154	infas
Schufa Holding AG	Auskunftei	163		155	infas
United Internet AG (1&1)	Kommunikation/ Internet	167		160	infas
Telekom Deutschland GmbH	Kommunikation/ Internet	169		80	Teilnehmer
Deutsche Post DHL Group	Logistik, Post	172	x		infas
Acxiom Deutschland GmbH	Adress-/Datenhändler	158	x		infas
AdAudience GmbH	Adress-/Datenhändler	157		157	infas
AZ Direct GmbH	Adress-/Datenhändler	162	x		infas
Deutsche Post Direkt GmbH	Adress-/Datenhändler	169	x		infas
Schober Information Group D.	Adress-/Datenhändler	159	x		infas
comScore, Inc.	Marktforschung	162	x		infas
Semasio GmbH	Marktforschung	160	x		infas
BILD GmbH & Co. KG	Medien	160		59	Teilnehmer
Google Germany GmbH	Online-/Internetdienste	173		170	infas
Miles & More GmbH	Prämienprogramm	162	x		infas
Payback GmbH	Prämienprogramm	171		10	Teilnehmer
Amazon EU SARL (Niederlassg. D)	Shopping	173	x		infas
eBay International AG	Shopping	171		33	infas
IKEA Deutschland GmbH & Co. KG	Shopping	171	x		infas
Zalando SE	Shopping	163	x		infas
Facebook Germany GmbH	Soziales Netzwerk	168	x		infas
PE Digital GmbH (Parship)	Soziales Netzwerk	160		5	Teilnehmer
Twitter Germany GmbH	Soziales Netzwerk	158	x		infas
WhatsApp Inc.	Soziales Netzwerk	170	x		infas
Apple Retail Germany B.V. & Co. KG	Technologie	163	x		infas
Microsoft Deutschland GmbH	Technologie	174	x		infas
Samsung Electronics GmbH	Technologie	170	x		infas
Deutsche Bahn AG	Verkehr	173		53	Teilnehmer
DriveNow GmbH & Co. KG	Verkehr	163		154	infas

Wie dargestellt, bearbeiteten 17 Unternehmen die Ersuchen offenbar gar nicht. Als Grund der Nichtteilnahme wurde von den Unternehmen vor allen Dingen angeführt, dass § 34 BDSG im Projektbezug nicht anwendbar sei. Sieben Unternehmen bezogen sich sinngemäß darauf, dass § 34 BDSG nicht für die Forschung gelte, sondern lediglich für die Einholung von Auskünften zugunsten der Betroffenen. Da die Daten primär für ein Forschungsprojekt zusammengetragen würden, sei § 34 BDSG trotz der vorliegenden Vollmachten nicht wirksam anwendbar. In sieben Fällen wurde darauf verwiesen, dass (gemäß § 174 BGB) die Vorlage von Originalvollmachten der Petenten für die Geltendmachung der Auskunftsansprüche zwingend notwendig sei³¹.

Abbildung 27 Gründe der Nicht-Teilnahme bei 17 Unternehmen



Ein weiteres Argument betraf den Umstand, dass Auskünfte nur bei nachweislicher Kundenbeziehung erteilt werden könnten, über die die vorliegenden Ersuchen keine Information enthielten. Dieses Argument wurde

³¹ Dieses Verfahren wurde auch auf Basis juristischer Beratung mit Rücksicht auf die Probanden bewusst gewählt, um diesen nicht jeweils bis zu 30 Einzelvollmachten zuzumuten. Zudem wurde so für die Unternehmen transparent, in welchem Umfang die Auskunftersuchen gestellt wurden.

unter den ablehnenden Unternehmen nur einmal explizit angeführt. Es wurde jedoch in weiteren Einzelfällen von Unternehmen geäußert, die schließlich doch Auskünfte an Petenten aussandten oder mit denen mündlicher Kontakt im Rahmen gestellter Nachfragen bestand.

In drei Fällen konnten die Auskunftersuchen den adressierten Unternehmen trotz mehrmaligen Versandbemühungen nicht zugestellt werden, da von diesen die Annahme verweigert wurde oder unlösbare Adressprobleme eine Auslieferung verhinderten. Aus zwei weiteren Unternehmen erfolgte schließlich keinerlei Rückmeldung, so dass die Gründe für die Nicht-Teilnahme unbekannt bleiben müssen.

Wie oben ausführlich dargestellt, wurden die von den Unternehmen angeführten Gründe, den Auskunftersuchen nicht nachzukommen, im Projektverlauf mit dem Auftraggeber und dem LDI NRW vielfältig diskutiert. infas erhielt dabei für sein inhaltliches und organisatorisches Vorgehen volle Unterstützung der genannten Beteiligten.

Etwaige juristische Maßnahmen zur weiteren Durchsetzung der Auskunftersuchen gehörten nicht zum Projektauftrag, zumal der nur schwer einschätzbare Ausgang solcher Schritte inhaltlich, zeitlich und monetär dem Forschungsauftrag nicht entsprochen hätte. Somit stellen die erzielten Rückmeldungen der teilnahmebereiten Unternehmen die empirische Basis dar, die im Folgenden genauer vorgestellt werden soll.

Im ersten Auswertungsschritt ist dokumentiert, ob überhaupt eine Resonanz erfolgt ist, die sich auf den einzelnen Petenten bezieht. Ob diese inhaltlich auswertbare Daten enthält, wird in Tabelle 3 dargestellt. Welche Art der Daten im Detail vorliegen, kann Tabelle 4 und folgenden entnommen werden.

Tabelle 3 weist die Art der in den Auskünften beinhalteten Informationen aus. Dabei wurde unterschieden zwischen der Information, dass und welche personenbezogenen Angaben vorliegen und der Information, dass keine solchen Angaben vorliegen – sowie letztlich dem Hinweis, dass solche

Auskünfte durch die Petenten auf der Unternehmenswebsite recherchiert werden können. Grundsätzlich ist darauf hinzuweisen, dass die Glaubwürdigkeit der erhaltenen Auskünfte durch das Projektteam nicht abschließend beurteilt werden kann.

Tabelle 3 Bearbeitete Auskunftersuchen und vorliegende Daten

Unternehmen, die die Auskunftersuchen in irgendeiner Form bearbeitet haben	Branche	bearbeitete Ersuchen durch die Unternehmen	darunter: enthalten personenbezogene Angaben	darunter: enthalten <u>keine</u> personenbezogenen Angaben
		abs.	abs.	abs..
Gesamt		1.346	433	913
Creditreform Boniversum GmbH	<i>Auskunftei</i>	156	156	0
informa HIS GmbH	<i>Auskunftei</i>	154	12	142
Schufa Holding AG	<i>Auskunftei</i>	155	153	2
United Internet AG (1&1)	<i>Kommunikation/ Internet</i>	160	54	106
Telekom Deutschland GmbH	<i>Kommunikation/ Internet</i>	80	31	49
AdAudience GmbH	<i>Adress-/Datenhändler</i>	157	0	157
BILD GmbH & Co. KG	<i>Medien</i>	59	4	55
Google Germany GmbH	<i>Online-/Internetdienste</i>	170	0	170
Payback GmbH	<i>Prämienprogramm</i>	10	5	5
eBay International AG	<i>Shopping</i>	33	0	33
PE Digital GmbH (Parship)	<i>Soziales Netzwerk</i>	5	5	0
Deutsche Bahn AG	<i>Verkehr</i>	53	5	48
DriveNow GmbH & Co. KG	<i>Verkehr</i>	154	8	146

Unter den 1.346 insgesamt bearbeiteten Ersuchen enthalten nur 413 Auskünfte konkrete inhaltliche Angaben über vorliegende Daten der Antragsteller. Damit werden in weniger als einem Drittel der Auskünfte verwertbare inhaltliche Informationen zur Bestimmung der Qualität und Quantität des Datenschattens geliefert. Hinzu kommt, dass sich allein etwa 300 dieser Angaben aus den Resonanzen zweier Auskunftsteilen speisen, und somit ein realistisches Gesamtbild über alle untersuchten Branchenvertreter nicht abgeleitet werden kann.

Die übrigen 913 Auskünfte enthalten keine personenbezogenen Daten zu den Petenten. Damit sind sie ebenfalls zu den bearbeiteten Auskünften zu

rechnen, die jedoch keine Daten enthalten. Die Tatsache, dass diese Auskünften keine personenbezogenen Daten enthalten, wird teilweise damit erklärt, dass kein Vertragsverhältnis zu den genannten Personen besteht (z. B. Telekom Deutschland GmbH), oder dass zu den genannten Identifikatoren keine internen Datensätze recherchiert werden konnten (z. B. Payback GmbH) oder dass vorab weitere, eindeutige Informationen zu einem möglichen Vertragsverhältnis geliefert werden müssten (z. B. Deutsche Bahn AG).

Google Germany GmbH und Ebay International AG geben in ihren Auskünften zusätzliche Hinweise, wie die Auskünfte über gespeicherte Daten für persönliche Nutzerkonten auf der Unternehmenswebsite abgerufen werden können. Im Einzelfall können demnach durchaus Daten vorliegen; es wurden allerdings keine Auskünfte hierüber erteilt, da diese individuell über bereitgestellte Tools auslesbar seien.

Im Falle Google wurde das in Deutschland ansässige Unternehmen Google Germany GmbH angeschrieben, das nach eigener Auskunft über keinerlei personenbezogene Daten der Petenten verfügt, da es nicht gleichzusetzen sei mit dem in den USA angesiedelten Internet-Anbieter Google Inc.

Die Gesamtbilanz der Unternehmensrückmeldungen steht gewissermaßen im Widerspruch zu den in Kapitel 5.2.4 berichteten Kunden- oder Nutzerbeziehungen der Petenten. So betreffen einige der Komplettausfälle bedauerlicherweise jene Unternehmen, zu denen die meisten Befragten über aktuelle oder frühere Kundenbeziehungen berichten. Aber auch von einigen der antwortenden Unternehmen durften insgesamt vollständigere Auskünfte erwartet werden.

In Tabelle 4 bis 6 werden alle 1.346 bei infas vorliegenden Auskünfte differenzierter betrachtet und die Ergebnisse in prozentualen Verteilungen dargestellt. Diese Art der Darstellung auf Basis sämtlicher vorliegenden Auskünfte erlaubt am ehesten den Vergleich der Resultate zwischen den Unternehmen.

Tabelle 4 enthält zunächst die Information über Weitergabe und Speicherung von Daten. Über alle Unternehmen hinweg enthalten 32 Prozent der bearbeiteten Ersuchen Informationen zur Weitergabe der Petenten-Daten an bestimmte Personen bzw. Institutionen. In 30 Prozent der bearbeiteten Ersuchen sind Informationen zum Zweck der Speicherung enthalten. In 25 Prozent der bearbeiteten Ersuchen finden sich schließlich Informationen zu Speicherfristen. Informationen zu Weitergabe und Speicherung liegen demnach insgesamt relativ häufig vor.

Tabelle 4 Bearbeitete Ersuchen: Informationen zu Weitergabe und Speicherung

Auskünfte, die infas vorliegen	Branche	bearbeitete Ersuchen	enthält Information zur Datenweitergabe	enthält Information zum Zweck der Speicherung	enthält Information zu Speicherfristen
		abs.	%	%	%
Gesamt		1.346	32	30	25
Creditreform Boniversum GmbH	Auskunftei	156	100	100	100
informa HIS GmbH	Auskunftei	154	8	8	8
Schufa Holding AG	Auskunftei	155	99	99	99
United Internet AG (I&I)	Kommunikation/ Internet	160	30	23	0
Telekom Deutschland GmbH	Kommunikation/ Internet	80	38	30	20
AdAudience GmbH	Adress-/Datenhändler	157	0	0	0
BILD GmbH & Co. KG	Medien	59	3	0	0
Google Germany GmbH	Online-/Internetdienste	170	0	0	0
Payback GmbH	Prämienprogramm	10	50	50	0
eBay International AG	Shopping	33	0	0	0
PE Digital GmbH (Parship)	Soziales Netzwerk	5	100	100	0
Deutsche Bahn AG	Verkehr	53	9	6	0
DriveNow GmbH & Co. KG	Verkehr	154	5	5	0

Nach Unternehmen aufgeschlüsselt zeichnen sich einige durch fast vollständige Aufklärung für alle Petenten aus. Hierzu zählen wie schon ausgeführt vor allem die antwortenden Auskunfteien. Auch aus anderen Unternehmen werden vorliegende Daten berichtet, jedoch für deutlich weniger Petenten. Ein Beispiel ist hier die PE Digital GmbH, die sehr vollständige Auskünfte erteilte, allerdings für lediglich 5 Mitglieder bzw. Kunden.

Im Folgenden werden die inhaltlichen Dimensionen der Auskünfte vorgestellt. Mit einem Anteil von 32 Prozent enthalten die Auskünfte am häufigsten Informationen zu Adress- und weiteren Daten, die den im Ersuch übermittelten Daten entsprechen. In ähnlicher Größenordnung (31 Prozent) sind Informationen zur Kundenbeziehung enthalten. Mit deutlichem Abstand, aber ebenfalls noch substantieller Größenordnung umfassen 23 Prozent der Auskünfte Informationen zu Bonitätsanfragen, 17 Prozent zu Zahlungsdaten und 16 Prozent zu Vertragsdaten.

Hervorzuheben ist erneut, dass die Auskunftsteile diejenigen Unternehmen darstellen, die am stärksten zu den Ergebnissen beitragen. Insgesamt darf vermutet werden, dass ein vollständiger Überblick der vorliegenden Daten aus sämtlichen untersuchten Branchen zu höheren Anteilswerten und anderen Verteilungen geführt hätte.

Tabelle 5 Bearbeitete Ersuchen: Häufig genannte Inhalte

Auskünfte, die infas vorliegen	bearbeitete Ersuchen	enthält (Adress-) Daten aus Auskunftsersuch	enthält Daten zur Kundenbeziehung	enthält Bonitätsanfragen	enthält Zahlungsdaten	enthält Vertragsdaten
	abs.	%	%	%	%	%
Gesamt	1.346	32	31	23	17	16
Creditreform Boniversum GmbH	156	100	100	0	0	100
informa HIS GmbH	154	8	0	0	0	1
Schufa Holding AG	155	100	98	95	94	66
United Internet AG (I&I)	160	34	34	0	23	33
Telekom Deutschland GmbH	80	39	39	0	36	39
AdAudience GmbH	157	0	0	0	0	0
BILD GmbH & Co. KG	59	7	3	0	0	2
Google Germany GmbH	170	0	0	0	0	0
Payback GmbH	10	50	50	0	50	50
eBay International AG	33	0	0	0	0	0
PE Digital GmbH (Parship)	5	100	100	0	0	100
Deutsche Bahn AG	53	9	9	0	6	8
DriveNow GmbH & Co. KG	154	5	5	0	5	5

Zu allen übrigen untersuchten inhaltlichen Dimensionen liegen nur in Bruchteilen der Auskünfte Informationen vor. So werden nur in 5 Prozent Kommunikationsdaten geführt, in 2 Prozent Verhaltensdaten, wie etwa Einkäufe oder Wege bzw. Orte, sowie in je einem Prozent Versicherungsanfragen oder technische Daten. Personenbezogene Daten Dritter wurden schließlich ebenfalls in lediglich 2 Prozent der Auskünfte benannt.

Nach den vorliegenden Ergebnissen verfügen die Auskunftsteile über derartige Informationen nahezu gar nicht. Wo die genannten inhaltlichen Datenkategorien vorliegen, entsprechen sie der Geschäftstätigkeit und den angebotenen Services der jeweiligen Unternehmen, im Sinne von Kommunikation und Internet, Prämienprogramm und Dating-Services.

Tabelle 6 Bearbeitete Ersuchen: Selten genannte Inhalte

Auskünfte, die infas vorliegen	bearbeitete Ersuchen	enthält Kommunikationsdaten	enthält Verhaltensdaten	enthält Versicherungsanfragen	enthält technische Daten	enthält personenbez. Daten Dritter
	abs.	%	%	%	%	%
Gesamt	1.346	5	2	1	1	2
Creditreform Boniversum GmbH	156	0	0	0	0	0
informa HIS GmbH	154	0	0	7	0	0
Schufa Holding AG	155	0	1	0	0	0
United Internet AG (1&1)	160	20	0	0	0	6
Telekom Deutschland GmbH	80	23	1	0	0	4
AdAudience GmbH	157	0	0	0	0	0
BILD GmbH & Co. KG	59	3	2	0	2	0
Google Germany GmbH	170	0	0	0	0	0
Payback GmbH	10	0	50	0	0	20
eBay International AG	33	0	0	0	0	0
PE Digital GmbH (Parship)	5	100	40	0	0	100
Deutsche Bahn AG	53	0	6	0	0	0
DriveNow GmbH & Co. KG	154	5	5	0	5	5

Tabelle 7 greift noch einmal die Subkategorien auf, die zur Klassifikation der berichteten inhaltlichen Dimensionen gebildet und verwendet wurden. Diese Darstellung beantwortet die Frage nach der inhaltlichen Spezifizierung der Dimensionen und liefert Erkenntnisse darüber, welche Daten konkret vorlagen, wenn beispielsweise zu einer Kundenbeziehung berichtet wurde, oder wenn Adressdaten vorlagen.

Tabelle 7 Codeplan und Inhalte der inhaltsanalytischen Auswertung

Dimensionen, zu denen Daten vorliegen	Anteil der bearbeiteten Ersuchen, die Angaben zu den Di- mensionen enthalten	inhaltliche Subkategorien, zu denen Daten vorliegen, sortiert nach Häufigkeit absteigend
Weitergabe und Speicherung		
Weitergabe der Daten an	32%	Zielperson (30%), Vertragspartner aller Branchen (23%), Werbetreibende (4%), Banken (3%), Strafverfolgungsbehörde (2%), Handel (2%), Versicherungen (1%), sonstige Unternehmen (1%)
Zweck der Speicherung	30%	Bonität/Zahlungsverhalten (23%), Vertragsdurchführung (6%), Informationen zu Versicherungen (1%)
Information zu Speicherfristen	25%	vorliegend (25%)
inhaltliche Dimensionen		
Daten laut Auskunftersuch	32%	Vorname (32%), Nachname (32%), aktuelle Adresse (32%), Geburtsdatum (31%), alte Adressen (8%), E-Mail (8%), Mobilfunknummer (6%), Telefon (3%), IP-Adresse (2%), Führerscheindaten (1%), Personalausweisdaten (1%),
Kundenbeziehung	31%	Art (31%), Dauer (16%), Kundensegment (1%)
Bonitätsanfragen	23%	berechnete Score-Werte (22%), angefragte Score-Werte (10%), positive Zahlungsinformationen (1%),
Zahlungsdaten	17%	Kontodaten (15%), Kreditkartendaten (8%), laufende Kredite/Leasing Geschäfte (4%), Zahlungen (Rechnungen, Raten) (4%), Identitäts- und Altersverifizierung durch Schufa (4%)
Vertragsdaten	16%	aktive Vertragsdaten (15%), inaktive/alte Vertragsdaten (3%)
Kommunikationsdaten	5%	Login (Zeiten online) (3%), Telefonate (1%), Kontakte (2%)
Verhaltensdaten	2%	Einkäufe (1%), Wege/Orte (1%)
Versicherungsanfragen	1%	Schadensfälle angefragt (1%),
weitere technische Daten	1%	sonstiges (1%)
personenbezogene Daten Dritter	2%	Namen/Adressen/E-Mail (2%)

Nach diesen Ergebnissen werden im Wesentlichen die klassischen Identifikatoren wie Name und Vorname, aktuelle Adresse und Geburtsdatum gespeichert, sowie Informationen zur Art der Kundenbeziehung, aber auch Score-Werte (allerdings nur der Auskunfteien), Konto- und Vertragsdaten. Auf die zahlreichen weiteren Identifikatoren, die im Auskunftersuchen geführt waren, wird mit Ausnahme von E-Mail und Telefonnummern nahezu kein Bezug genommen. Accountkennungen für Facebook, Apple oder Google tauchen gar nicht auf. Gleiches gilt für Kennungen zum Smartphone (IMEI, MAC-Adresse, AD-ID).

In der Bewertung der vorliegenden Inhalte lässt sich zusammenfassend festhalten, dass die Unternehmen zwar teilweise „statische“ Daten berichten, aber kaum bis keine „dynamischen“ Daten – also als Eingangsdaten gespeicherte Angaben, aber keine komplexen aus einer Weiterverarbeitung oder laufenden Prozessen entstandenen Daten. Damit entsteht der Eindruck eines gewissermaßen „analogen Datenschattens“ vor einer möglichen Weiterverarbeitung und dem Einsatz von Algorithmen. Bezogen auf die eingangs vorgestellte Abbildung zum Modell des Datenschattens (siehe Abbildung 7) beziehen sich die Auskünfte fast ausschließlich auf Phase 2 und Daten aus allen darüber hinaus gehenden Prozesse, zu denen jedoch häufig Angaben verfügbar sein müssen, werden ganz überwiegend nicht berichtet. Vor dem Hintergrund dieser Erkenntnis erscheint zweifelhaft, ob Verbraucher über den Hebel des § 34 BDSG tatsächlich ein realistisches Bild ihres Datenschattens erlangen können, geschweige denn eine aktive Kontrolle desselben ausüben. Dies ist als ein zentrales Projektergebnis zu betrachten, das sich erst durch das teilweise Scheitern der ursprünglichen Projektplanung vermittelt.

5.4 Umfang des Datenschattens und Einflussgrößen

Die in Kapitel 5.3 vorgestellten Ergebnisse der Selbstauskünfte geben einen teilweisen Einblick in mögliche Strukturen und den Umfang des Datenschattens von Konsumenten bei einzelnen Unternehmen. Anders als geplant führen sie jedoch nicht zu einer systematischen und verlässlichen Bestandsaufnahme, aus der sich ein vollständiges Bild des Datenschattens ableiten ließe. Wie dargestellt, sind dafür die mangelnde Vollständigkeit der bereitgestellten Auskünfte, aber auch deren vermutete Qualität und Begrenztheit verantwortlich. Zahlreiche Unternehmen haben sich der Auskunftspflicht ganz entzogen, andere haben in einer Form geantwortet, die nahelegt, dass nur ein kleiner Teil der vorliegenden Daten oder sogar trotz einer formal erfolgten Auskunft gar kein Teil benannt wurde, obwohl für zahlreiche Probanden nach deren Angaben Kontakte bestehen (wie zum Beispiel durch Google, dem in der Anfrage unterstützend sogar Google-Accounts der Teilnehmer bekannt gegeben wurden).

Da die im Projektrahmen gegebenen Durchsetzungsmittel wie erläutert ausgeschöpft wurden und weitere entsprechende Maßnahmen nach Einschätzung der Fachbehörden (hier dem LDI in NRW) von ungewissem Erfolg wären³², kann diese Datenquelle nicht für die gewünschte systematische Bestimmung des Datenschattens herangezogen werden. Sie vermittelt aufschlussreiche Eindrücke, erlaubt aber wie schon erläutert weder valide noch reliable Aussagen zum tatsächlichen Umfang des Datenschattens. Dies ergibt sich schon aus der Tatsache, dass alle Unternehmen aus dem Bereich Social Media zu den Nicht-Antwortern gehören. Damit entzieht sich ein vermutlich wesentlicher Bestandteil des Datenschattens der im Projekt intendierten Betrachtung auf objektiver Grundlage.

Um sich dem Datenschatten ersatzweise auch ohne eine gemessene Datengrundlage auf Unternehmensseite empirisch anzunähern und im möglichen

³² Derartige Maßnahmen würden voraussichtlich zumindest zeitliche Verzögerungen im Bereich mehrerer Monate bis zu einem Erfolg in Anspruch nehmen.

Rahmen Schlussfolgerungen ziehen zu können, sollen abschließend die Projektergebnisse aus der Befragung und die Beobachtung über die App Tracy miteinander kombiniert werden. Ausgangspunkt dafür sind die in Kapitel 5.1 zum Modell des Datenschattens vorgestellten theoretischen Überlegungen. Sie zeigen, welche Quellen zu berücksichtigen sind. Soweit diese entweder in der Befragung oder aber den technischen Messungen der App abgebildet wurden, können sie zu einem „Schätzer des Datenschattens“ zusammengeführt werden. Das gewählte Vorgehen stellt dabei lediglich einen Vorschlag dar und soll die weitere Auseinandersetzung anregen. Die im Folgenden erläuterte empirisch ausgerichtete Operationalisierung des Datenschattens war in der ursprünglichen Projektplanung in dieser Form nicht vorgesehen, ihr Ergebnis liefert aber auch ohne den Einbezug der Unternehmensangaben eine Reihe von Anregungen.

Ersatzweise für eine systematische Auswertung der Unternehmensangaben wurde ein mehrdimensionaler *Datenschattenindex* gebildet. Er nutzt die erhobenen Daten aus Befragung und App, führt folgende darin enthaltenen Merkmale zusammen und gewichtet diese nach einem multivariaten statistischen Verfahren:

Tabelle 8 Bildung Datenschattenindex³³

eingehende Dimensionen	Gewichtungsfaktor	Mittelwert auf einer Skala 0-100
aus der Begleitbefragung		
- Anzahl Online-Konten (Anzahl der online erstellen Kundenkonten wie z. B. Amazon oder Ebay sowie der Mitgliedschaften bei Social Media-Anbietern wie z. B. Facebook oder Instagram)	1,0	43
- zeitliche Intensität der Unternehmenskontakte (von Befragten angegebene Kontaktintensität mit den 21 einbezogenen Unternehmen, bei denen um Auskünfte gegeben wurden)	1,0	38
- Nutzung neuer „smarter“ Geräte (z. B. Haushaltsgeräte, Fitnesstracker, Sprachassistenten)	0,9	27
- persönlich getroffene Sicherheitsmaßnahmen (z. B. aktiv getroffene Einstellungen zum Identitätsschutz in genutzten Browsern)	-0,7	14
aus der App Tracy bzw. dem dortigen regelmäßigen Recall-Kontakt		
- genutzte Websites (Anzahl der im Recall angegeben genutzten Websites)	0,4	14
- auf dem Smartphone installierte App-Prozesse (technische Messung aller Prozesse durch Tracy)	0,4	27
- Smartphone-Nutzungsdauer (durch Tracy gemessene mittlere tägliche Nutzungsdauer)	0,6	23
- GPS-Trackingmöglichkeit (durch Tracy gemessene durchschnittliche tägliche räumliche Trackingmöglichkeit)	0,7	26
Datenschattenindex insgesamt (0-350 theoretisch mögliche Punkte)		159

Abhängig von den berichteten bzw. gemessenen Aktivitäten jedes einbezogenen Befragten konnte so für jede Person ein individueller Schätzwert für das Ausmaß des Datenschattens ermittelt werden. Die durch den Index abgebildete theoretische Spannweite stellt den „heutigen Datenschatten“ ge-

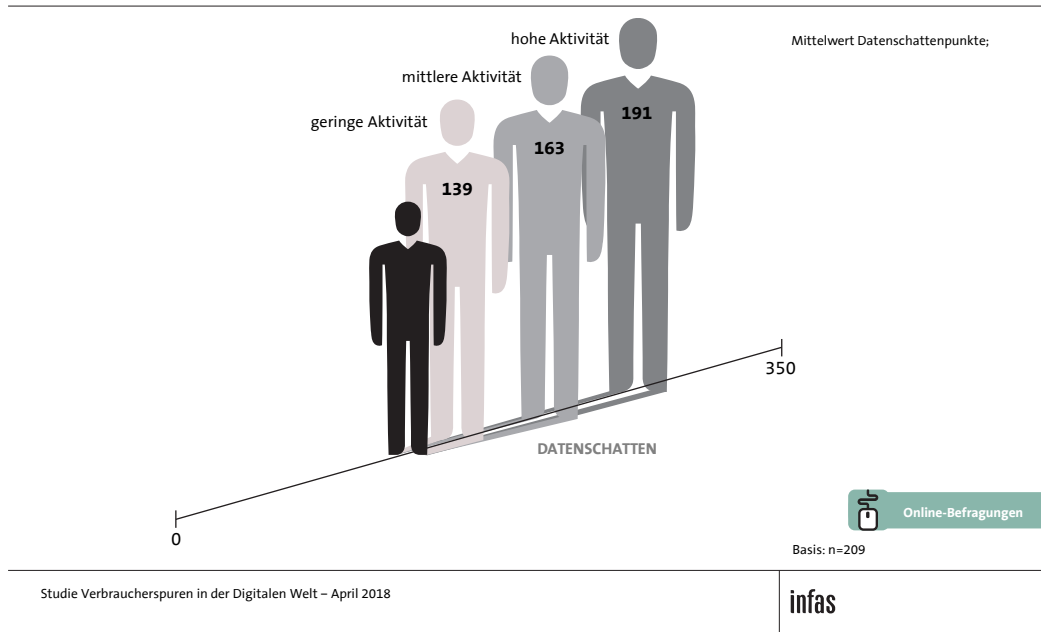
³³ Für jede der genannten Dimensionen wurde ein Summenindex mit dem jeweiligen Wert zwischen 0 und 100 berechnet. Alle Dimensionen wurden im Rahmen einer Faktorenanalyse betrachtet. Dieses Verfahren liefert auf statistischer Basis inhaltlich möglichst homogene Faktoren. Zudem erkennt es Redundanzen durch möglicherweise mehrfach einbezogene sehr ähnliche Dimensionen, die so ausgeschlossen werden konnten. Zusätzlich ermittelt das Verfahren ein Maß für die Bedeutung jeder einzelnen Dimension innerhalb eines Faktors. Dieses Maß wurde jeweils als Gewichtungsfaktor verwendet. Durch die identische Skalierung von jeweils 0-100 konnte verbunden mit dem Gewichtungsfaktor jeder Dimension ein einfacher summativer Gesamtindex gebildet werden – der hier so genannte *Datenschattenindex*. Er liegt bedingt durch die Anzahl der eingehenden Dimensionen und ihrer Gewichtung in einem Wertebereich zwischen 0 und 350.

mäß unseren Vorüberlegungen dar (siehe Kapitel 5.1.1). Den dabei rechnerisch möglichen Maximalwert hat allerdings keiner der Probanden aus der Stichprobe erreicht.

Insgesamt ergibt sich auf der Gesamtskala zwischen 0 und 350 „Datenschattenpunkten“ ein relativ hoher Mittelwert von 159 Punkten. Der Median, also der Wert, unter bzw. über dem jeweils die Hälfte der Befragten liegt, ist mit 156 Punkten nahe am Mittelwert. Nur etwa 10 Prozent der einbezogenen Studienteilnehmer haben einen Datenschattenindex von weniger als 100 Punkten. Der geringste Wert liegt bei knapp 20 Punkten. Personen ohne Datenschattenpunkt sind gar nicht vertreten. Etwa 10 Prozent weisen aber einen Punktwert von 225 oder mehr auf.

Noch aufschlussreicher als diese Verteilungskennzeichen sind die Ergebnisse unterschieden nach den drei Nutzergruppen von „gering“ bis „hoch“, die bezogen auf die Online-Aktivitäten gebildet und oben bereits vorgestellt wurden. Zu erwarten wären deutliche Abstände zumindest proportional zu der gemessenen Aktivität – also ein sehr großer Datenschatten bei Befragten mit hoher und ein entsprechend kleinerer bei Befragten mit geringer Aktivität. Das in der folgenden Abbildung dargestellte Ergebnis zeigt, dass es sich anders verhält. Die drei Datenschattenwerte liegen mit Punkten von 139, 163 und 191 relativ eng beieinander. Also weisen auch Personen mit der Einstufung einer geringen Aktivität einen bereits recht hohen Datenschattenindex auf. Dies korrespondiert mit der Ableitung aus der Summe der Unternehmenskontakte, die am Ende von Kapitel 5.4 vorgestellt wurde.

Abbildung 28 Schätzwert des Datenschattenindex nach Online-Aktivität

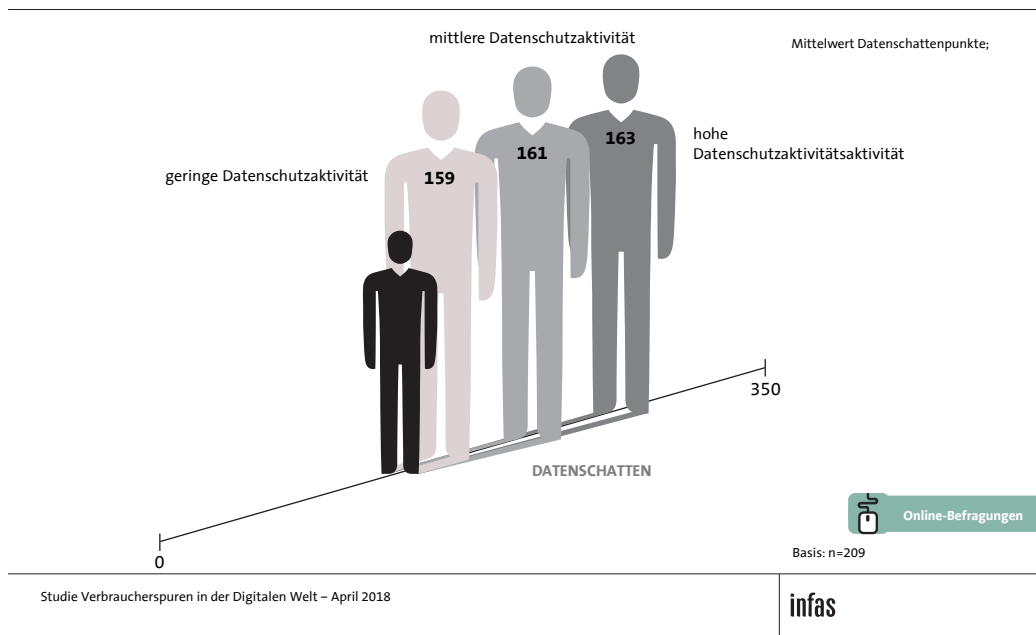


Inhaltlich ist dafür vor allem die Beobachtung verantwortlich, dass die Mehrzahl der Befragten im Alltag inzwischen gewisse Online-Basisangebote nutzt. Ebenso steht sie wie berichtet zu einer größeren Zahl der 21 einbezogenen Unternehmen zumindest gelegentlich in Kontakt. Es genügen also bereits wenige Alltagsaktivitäten, um eine größere Anzahl bzw. eine Art „Sockel“ von Datenspurten zu hinterlassen. Dieser wächst mit intensiverer Tätigkeit weiter an, jedoch nicht unbedingt proportional zum Aktivitätsumfang.

Die Ergebnisse zeigen erneut, dass individuell getroffene Schutzmaßnahmen wie etwa der restriktive Umgang mit Cookies, die Aktivierung von identitätsschützenden anderen Browser-Einstellungen oder der sparsame Umgang mit persönlichen Angaben nur bedingt geeignet sind, um zu einem kleinen Datenschatten zu gelangen. Dieser kann derartigen Restriktionen etwas verringert werden, ist aber in erster Linie eher abhängig vom eigenen Aktivitätsausmaß. Wird die oben eingeführte Segmentation nach persönlich getroffenen Datenschutzmaßnahmen auch hier zur Unterscheidung herangezogen, zeigen sich im Datenschattenindex zwischen geringen und hohen Sicherheitsbemühungen kaum Unterschiede. Die entsprechenden Werte

liegen bei 159, 161 und 163 Datenschattenpunkten. Sie sind also so gut wie unabhängig von den berichteten Datenschutzaktivitäten.

Abbildung 29 Schätzwert des Datenschattens nach Datenschutzaktivität

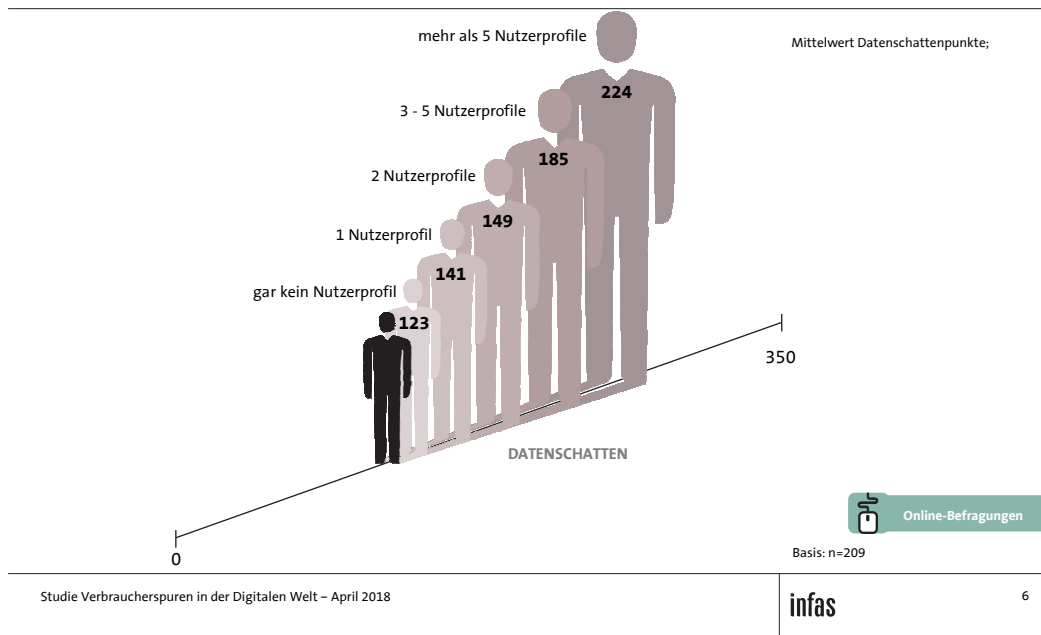


Ein Blick auf die Hintergründe zeigt die bereits an anderer Stelle betrachtete Ursache dafür. Befragte mit hohen Sicherheitsanstrengungen sind im Durchschnitt online sehr viel aktiver als Personen, die sich darüber kaum Gedanken machen. Ursächlich ist also eine Art „Selbstsensibilisierung“. Eine hohe Aktivität führt zu mehr individuellen Sicherheitsaktivitäten. Diese können aber das schon erreichte Niveau des Datenschattens offenbar nicht mehr entscheidend regulieren.

Noch aufschlussreicher ist eine dritte Differenzierung des Datenschattenindex nach einer Nutzerunterscheidung abhängig von der Anzahl der persönlich genutzten sozialen Netzwerke. Hier zeigen sich sehr deutliche Unterschiede je nach Aktivitätsgrad. Personen, die über 5 oder mehr Nutzerprofile verfügen, verzeichnen einen nahezu doppelt so hohen Datenschattenindex wie Personen, die ganz auf diese Art der individuellen Online-Präsenz verzichten. Ein umfassender Datenschatten wird also maßgeblich durch dieses Aktivitätsfeld geprägt. Umgekehrt zeigt der Datenschattenwert

für Personen außerhalb der Social Media-Aktivitäten von immerhin noch 123 Punkten, dass eine diesbezügliche Abstinenz den Datenschatten offenbar verringert, nicht aber auf geringem Niveau halten kann.

Abbildung 30 Schätzung des Datenschattenindex nach Zahl der Social Media-Profile



Die dargestellten Ergebnisse erheben nicht den Anspruch auf eine objektive Darstellung des individuellen Datenschattens. Sie zeigen jedoch, dass der individuelle Datenschatten bereits anhand weniger Aktivitäten entsteht und ein relativ großes Ausmaß annimmt. Zwar stellen die Befragten in der Studie keinen aktuellen repräsentativen Bevölkerungsquerschnitt dar und sind im Schnitt vermutlich mit einem größeren Datenschatten konfrontiert als der Bevölkerungsquerschnitt, doch sind sie – wie bereits erläutert – nach unserer Überzeugung eine gute Annäherung für eine Nachfragesituation in nur wenigen Jahren, wenn die Online-Durchdringung noch weiter fortgeschritten sein wird.

Ein umfassender Datenschatten ist so nach den vorliegenden Erkenntnissen bereits heute Realität, auch für Personen, die sich außerhalb der Online-Welt vermuten. Er nimmt zwar insbesondere durch Social Media-Aktivitäten große Ausmaße an, doch entzieht sich der Umfang der indivi-

duellen Datenspurten der subjektiven Einschätzung – und damit auch der Sensibilität und mitunter sogar einer Chance für eine notwendige Kontrolle. Vor allem Personen mit geringen Aktivitäten neigen vermutlich dazu, ihren eigenen Datenschatten zu unterschätzen. Zudem stellt oft nicht die reine „Größe“, sondern der inhaltliche Gehalt des Datenschattens eine potentielle „Gefahrenquelle“ dar.

Dass gerade Offliner und Personen mit geringer Online-Aktivität zu Fehleinschätzungen gelangen können, zeigt schließlich sogar die Größe des Datenschattens – operationalisiert anhand des gebildeten Index – abhängig von der Antwort auf die Frage nach der geschätzten Größe des eigenen Datenschattens. Personen, die in der Befragung ihren Datenschatten als „sehr viel oder etwas kleiner als im Durchschnitt“ einschätzen, weisen immerhin einen Indexwert von bereits etwa 140 Punkten auf. Sie liegen damit nur knapp unter dem Mittelwert aller Studienteilnehmer. Umgekehrt ist der Datenschatten der Personen, die ihre Datenspurten als „etwas oder sehr viel größer als im Durchschnitt“ einstufen, mit etwa 200 Indexpunkten offensichtlich nicht ganz so umfangreich, wie es die hier vorgestellte näherungsweise Berechnung mit sich ergebenden tatsächlichen Maximalwerten von etwa 250 Punkten nahelegt.

Dass nach diesem Ergebnis Personen, die ihren Datenschatten für eher klein halten, bereits hohe Werte verzeichnen, wirft wiederholt ein kritisches Licht auf die unterstellte Kontrollierbarkeit. Diese würde ein erhebliches Wissen und eine häufige Auseinandersetzung mit dem Thema voraussetzen und ist selbst dann aufgrund von Komplexität und Dynamik vermutlich nicht gegeben. Es kann also nicht genügen, das Management und die Kontrolle des Datenschattens den einzelnen Nutzern zu überlassen. Dies kann nur *ein* Baustein eines ganzen Systems einer ausgebauten Reglementierung sein.

5.5 Ausblick

Die vorgestellten Analysen beschreiben die Entstehungsgeschichte und die Auswirkungen eines persönlichen Datenschattens als überwiegend frei von gezielten Eingriffen durch den „Schattenwerfer“. Demgegenüber steht die populäre Meinung, das Datenschattenwachstum ließe sich durch die individuellen Nutzer bremsen oder gar unterbinden. Einzelne Bürger könnten entweder die Entstehung ihres Datenschattens vermeiden oder sie könnten diesen im Nachhinein steuern und verkleinern. In beiderlei Hinsicht, so konnte gezeigt werden, sind die Einflussmöglichkeiten einzelner Nutzer jedoch deutlich geringer als landläufig angenommen und der Umfang des Datenschattens ist mittelfristig weitgehend verhaltensunabhängig.

Auch die individuellen Kontrollmechanismen sind selbst bei einer funktionierenden Selbstauskunft und besseren Sanktionsmechanismen der neuen DSGVO nur ein Element unter mehreren. Dies ist der Entwicklungsdynamik und zahlreichen Interaktionseffekten geschuldet, denen der einzelne Verbraucher und möglicherweise sogar einige Unternehmen nicht folgen können.

Was also tun?

Ein Verzicht oder eine Reduktion des Gebrauchs digitaler Geräte und Medien zur Verhinderung der Entstehung eines Datenschattens ist wenig realitätsnah. Wie die vorliegenden Ergebnisse anhand ausgewählter Personen beispielhaft zeigen, ist das Alltagsleben bereits heute zu einem hohen Grad digitalisiert – und der Umfang wird zunehmen. Ein Rückzug aus digitalen Angeboten entspräche einem umfangreichen Verzicht auf persönliche Entfaltung und ökonomisches Fortkommen. Sollten sie es dennoch versuchen, fehlen Konsumenten in der Regel die für eine aufgeklärte Entscheidung notwendigen Informationen. Wie bereits ausgeführt, ist es darüber hinaus inzwischen bereits Unternehmenspraxis, über das Verhalten Dritter einen signifikanten Datenschatten aufzubauen. Damit greift selbst eine vollkom-

mene persönliche Datenaskese zu kurz. Das Konzept einer Individualisierung der Problemlage erweist sich so als wenig belastbar.

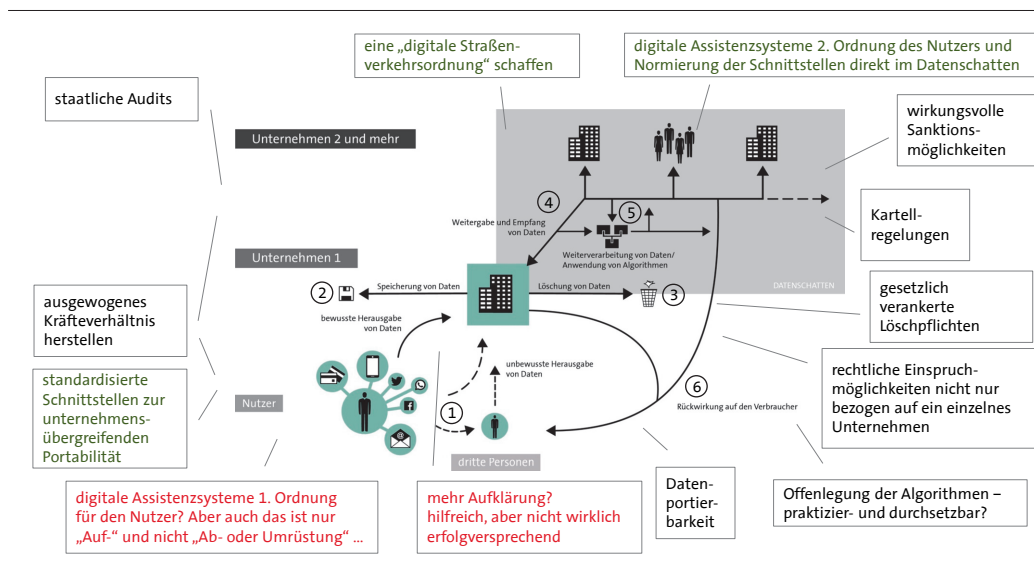
Wenn ein Datenschatten bereits entstanden ist – und davon ist nach den vorgestellten Ergebnissen in so gut wie allen Fällen auszugehen – bleibt einem Bürger noch die theoretische Möglichkeit, diesen individuell zu steuern oder gar zu verkleinern. Auch hier hat der individuelle Nutzer jedoch deutlich weniger Einflussmöglichkeiten als oft angenommen. Mit Blick auf die Transparenz des Datenschattens reichen klassische, an üblichen Mechanismen beim Kauf eines Produkts oder der Inanspruchnahme einer Dienstleistung ausgebildete Verbraucherschutzmechanismen nicht aus.

Ausschlaggebend ist, dass es sich bei der Nutzung digitaler Geräte nur vordergründig um eine individuell zu treffende, rationale Kaufentscheidung handelt. Und bei weiteren einen Datenschatten induzierenden Vorgängen wie etwa einer Bonitätsbewertung ist möglicherweise gar keine übliche Kauf- oder Verbraucherentscheidung auslösend. In Geschäftsmodellen großer Internet-Firmen, doch bereits in den Kundenbindungsprogrammen klassischer Anbieter, tauscht ein Nutzer Daten gegen Leistung. Er tauscht jedoch nicht wirklich bloße Daten, sondern die Möglichkeit zukünftiger automatisierter Entscheidungen über seine Person. Diese sind zum Zeitpunkt des Einwilligens in einen „Kaufvertrag“ nicht absehbar. Dem Nutzer ist also vielfach nicht klar, welche persönlichen Nachteile er möglicherweise gegen den Nutzen eines Dienstes eintauscht. Er müsste erst wissen, welche Daten in welchem Kontext für welche zukünftige Entscheidung mit welchen Bewertungsschemata herangezogen werden. Da diese Transparenz bei den gegenwärtigen Regulierungsmechanismen ex ante gar nicht möglich ist, kann er seine Entscheidung nicht rational fällen. Dies wird verschärft durch die mittlerweile zu konstatierende Tatsache, dass persönliche Daten zu persönlichem ökonomischen Kapital werden, mit deren kostenloser Preisgabe nach gegenwärtiger Praxis der Zugang zu bestimmten Dienst-

leistungen oder Waren überhaupt erst gestattet wird – etwa zu einem Facebook-Konto oder bei der Freischaltung eines bereits erworbenen iPhones.

Im Gegensatz zu einer klassischen Kaufentscheidung „erwirbt“ der Nutzer kein isoliertes Produkt, sondern den langfristigen Gebrauch eines Dienstes. Sowohl Nutzen als auch die Kosten – inklusive automatisierter Entscheidungen zu seinen Ungunsten – unterliegen dabei kontinuierlichen Veränderungen, Interaktionseffekten und technischen Weiterentwicklungen der Datennutzung. Zum Zeitpunkt des Einwilligens in beispielsweise eine Datenschutzerklärung sind diese Entwicklungen für einen individuellen Verbraucher nicht absehbar. In Summe greift daher die einmalige Kaufentscheidung als Modell für digitale Services wesentlich zu kurz.

Abbildung 31 Ableitungen und Empfehlungen zu möglichen Regelungsansätzen



Studie Verbraucherspur in der Digitalen Welt – April 2018

infas

Mit Blick auf das Management des Datenschattens benötigen Verbraucher also deutlich komplexere und langfristig angelegte „Verträge“ und Steuerungsmöglichkeiten. Diese auszuarbeiten und zu bewerten ist nicht Aufgabe des dargestellten Projekts. Aber die erarbeiteten Befunde und selbst das nicht belastbare Ergebnis im Rahmen der eingeholten Selbstauskünfte

föhren zu möglichen Ansatzpunkten, die in der vorangehenden Abbildung kurz angedeutet und in der Fachdiskussion bereits Gegenstand sind.

In der Summe wird es auf eine Regulierung ankommen, die wie in anderen Lebensbereichen auch sowohl dem Verbraucher Möglichkeiten für selbstbestimmte Entscheidungen an die Hand gibt als auch ihn an einer „höheren“ Stelle schützt, wenn er in seinem individuellen Handeln vermutlich überfordert ist – wie es die Studienergebnisse nahelegen – oder seine Rechte nicht beachtet werden. Wesentlich erscheint dabei, dass das vielschichtige Entstehen des Datenschattens nur durch ein ebenso komplexes und vorausschauendes generalisiertes System regulativer Eingriffe langfristig sozial verträglich sein wird und so seine vor allem im Dienstleistungssektor auch vorhandenen Vorteile bewahren kann.

6 Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse

Die Projektergebnisse zeigen, dass der Datenschatten zu einem nicht mehr wegzudenkenden Alltagsbestandteil geworden ist. Obwohl für online besonders aktive Probanden ein etwas umfassenderer Datenschatten abgeleitet werden konnte als für „analogere“ Probanden, entsteht er bereits heute in seiner Qualität weitgehend unabhängig von individuellem Handeln. Dies wird in den nächsten Jahren aller Voraussicht nach mit hoher Dynamik zunehmen, so dass sich individuelle Datenschattenvolumen immer mehr angleichen werden.

Gleichzeitig zeigen die Resultate, dass eine individuelle Steuerbarkeit aufgrund der hohen Komplexität und dem Verwobensein mit zunehmend mehr Alltagshandlungen die Verbraucher bereits jetzt überfordert. Der Datenschatten ist kein überschaubares, quasi statistisches digitales Abbild einer analogen Person, sondern weist eine sehr viel höhere Dynamik auf.

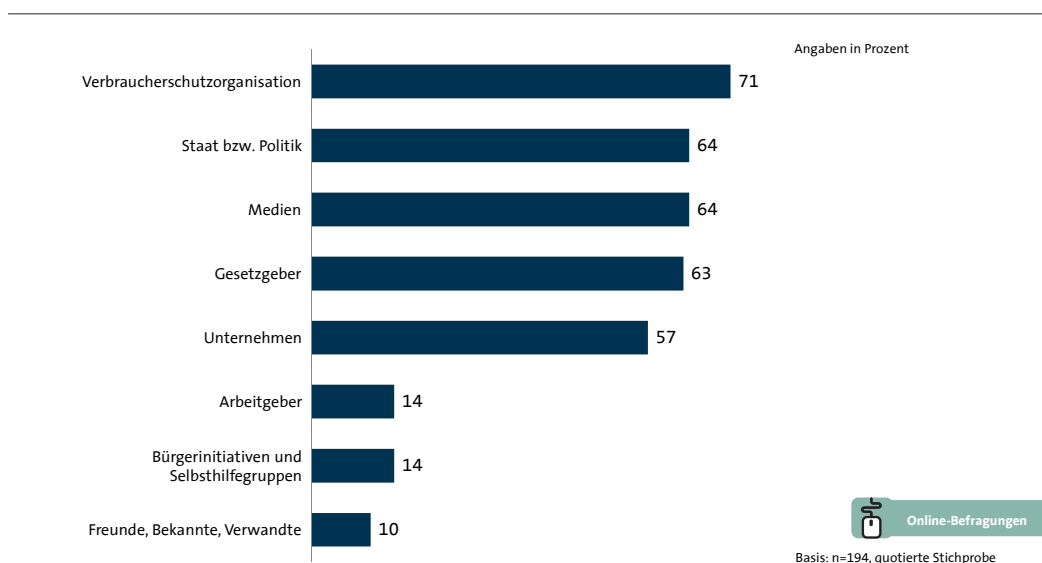
Vor diesem Hintergrund darf auch konstatiert werden, dass der beschriebene Ausfall der Auskünfte auf Basis von § 34 BDSG ein kleinerer Nachteil ist als es auf den ersten Blick erscheint. Selbst vollständige Auskünfte hätten sich vermutlich weitgehend in der Angabe zur Speicherung relativ statischer Daten erschöpft. Weiterverarbeitungen und Informationen zu Ableitungen enthalten selbst die erteilten Auskünfte in nur sehr eingeschränktem und wahrscheinlich zumeist nicht vollständigem Umfang. Ihre Nutzung würde daher die Komplexität des tatsächlich bereits bestehenden Datenschattens deutlich unterschätzen.

Darüber hinaus legen die Ergebnisse nahe, dass das wichtige Instrument der informationellen Selbstbestimmung auch in einer verbesserten Form nur einen kleinen Teil einer Datenschattenkontrolle bilden kann, so dass weitere Instrumente erforderlich sind. Ausschlaggebend dafür sind die Komplexität und die technische Dynamik der Datenschattenentstehung sowohl auf

der Seite des einzelnen Verbrauchers wie auch auf Seite der datenverarbeitenden Unternehmen.

Daher liefern die ersatzweise angestellten theoretischen Überlegungen und der zur Diskussion gestellte Versuch, über den Datenschattenindex zu einer Schätzung des Datenschattenumfangs zu gelangen, wichtige Einsichten. Die vorliegenden empirischen Ergebnisse über das Verhalten und Erleben einer an Datenthemen besonders interessierten Verbraucherstichprobe gibt Anstöße für die künftige Verbraucherpolitik in diesem Feld und regt dazu an, eine Regulierung des Datenschattens in erster Linie nicht von der Entstehungsseite und der bisherigen Entwicklung aus anzugehen, sondern ihn stattdessen von seinen zukünftigen Potenzialen aus zu betrachten und verbraucherschutzpolitisch wie gesetzgeberisch zu begleiten. Dies gilt umso mehr, wenn in das Bewusstsein rückt, dass individuelle Datenspurten nicht nur Spuren sind, sondern bereits gegenwärtig und zukünftig zunehmend zu individuellem ökonomischen Kapital werden, das einem Verbraucher ähnlich wie der Inhalt seiner Geldbörse den Zugang zu bestimmten Dienstleistungen ermöglicht.

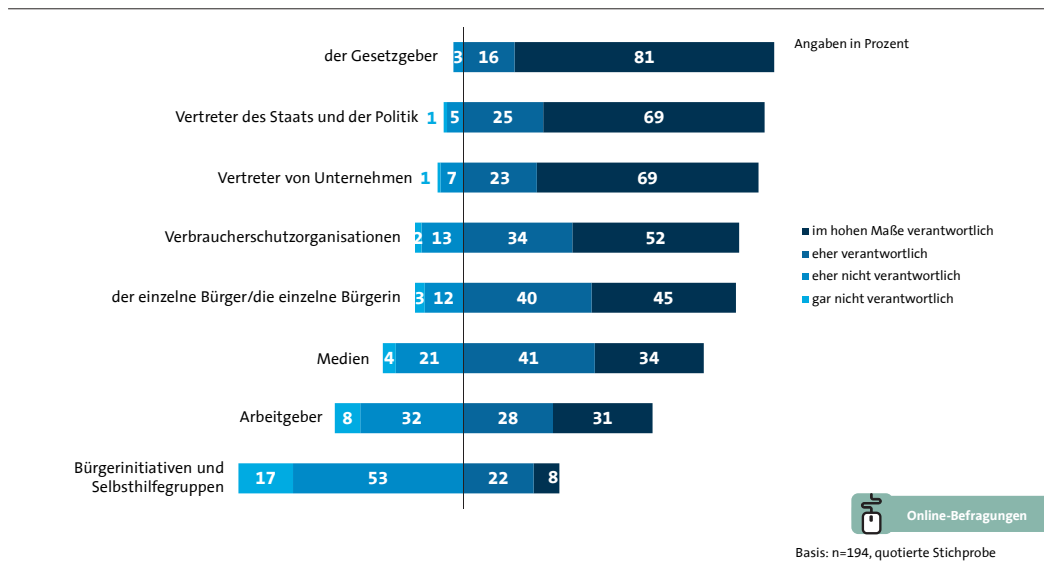
Abbildung 32 Präferierte Informationsquellen zum Thema Datenschatten



Die Erwartung eines komplexen und verantwortungsvollen Handlungsfelds teilen auch die befragten Probanden – sowohl mit Blick auf ihre Informationsquellen und Sensibilisierung als auch mit Bezug auf die verantwortlichen Institutionen (siehe Abbildungen 32 und 33). Hinsichtlich einer verbesserten Verbraucherinformation würde ein breites Feld der Akteure begrüßt, in dem allen voran (nicht-staatliche) Verbraucherschutzorganisationen, aber auch staatlichen Instanzen der Verbraucherpolitik, Medien und Unternehmen wichtige Rollen eingeräumt werden.

Stärker noch als im Bereich der Informationsvermittlung wird dem Gesetzgeber sowie den Vertretern von Staat und Politik im Schutz von Verbraucherdaten eine maßgebliche Verantwortung zugeschrieben. Auch die möglichen „Profiteure“ eines Datenschattens auf Unternehmensseite werden hier in hohem Maße in der Pflicht gesehen.

Abbildung 33 Zugeschriebene Verantwortung für den Schutz von Verbraucherdaten



7 Zusammenfassung

Im Auftrag des Bundesministeriums der Justiz und für Verbraucherschutz (BMJV) führte infas von Januar 2017 bis April 2018 eine Studie zum „Datenschatten“ durch. Ziel des Projekts war eine Bestandsaufnahme, welche persönlichen Daten Verbraucher digital im Alltag hinterlassen und wie diese Daten von Unternehmen genutzt werden.

Dazu wurde das digitale Verhalten bei über 200 Studienteilnehmern in einer dreimonatigen Feldphase untersucht. Dabei wurden sowohl Online-Befragungen als auch eine Smartphone-App eingesetzt, um das Verhalten und die Aktivitäten der Teilnehmer zu messen.

Zusätzlich wurde bei 30 auf dem deutschen Markt tätigen, marktrelevanten Unternehmen geprüft, welche Daten ihnen von den Studienteilnehmern vorliegen und zu welchen Zwecken diese verwendet werden. Dafür stellte infas in Vollmacht der Projektteilnehmer Anträge auf Datenselbstauskunft nach § 34 BDSG. Da jedoch lediglich 13 der Unternehmen Daten in unterschiedlicher Vollständigkeit zurücklieferten, konnte die empirische Bestimmung zur individuellen Größe des Datenschattens nicht wie geplant anhand dieses Projektmoduls erfolgen.

Dennoch lässt sich aus den empirischen Ergebnissen ableiten, dass bereits wenige Alltagsaktivitäten genügen, um eine größere Anzahl von Datenspurten zu hinterlassen. Der Datenschatten wächst mit intensiverer Tätigkeit weiter an, jedoch nicht proportional zum Aktivitätsumfang. Die Einflussmöglichkeiten der Verbraucher auf die Vermeidung von Datenspurten stellen sich als relativ begrenzt heraus. Ein Rückzug aus digitalen Angeboten ist wenig realistisch und Verbraucher können die Nutzung ihrer Daten in der Zukunft nicht voraussehen. Daher wird es auf eine Regulierung ankommen, die dem Verbraucher verbesserte Möglichkeiten für selbstbestimmte Entscheidungen an die Hand gibt und ihn gleichzeitig an anderer Stelle schützt.

8 Gegenüberstellung der ursprünglich geplanten zu den tatsächlich erreichten Zielen

Das Hauptziel der Studie war die Analyse, welche persönlichen Daten Verbraucher bei alltäglichen Handlungen digital hinterlassen und wie diese Daten von Unternehmen genutzt werden. Dies wurde durch eine Bestandsaufnahme der Smartphone-Nutzung ergänzt durch weitere Aspekte operationalisiert. Es sollte untersucht werden, ob Verbraucher selbst durch ihr alltägliches Verhalten zu ihrem Datenschatten beitragen und wie groß ihr diesbezüglicher Einfluss auf diesen ist. Zudem sollte betrachtet werden, welche Datentypen von den Unternehmen erfasst und verarbeitet werden.

Alle zur Zielsetzung vorgesehenen Projektmodule wurden entsprechend des inhaltlichen und zeitlichen Projektauftrags eingesetzt und durchgeführt. In der Studie kamen mit den Online-Befragungen und der App Tracy umfassende Module der empirischen Datenerhebung zum Einsatz. Diese beschreiben das Verbraucherverhalten und liefern erklärende Merkmale für die quantitative und qualitative Ausprägung des Datenschattens. Zwar beansprucht die Stichprobe der Projektteilnehmer auftragsgemäß keine Repräsentativität in Bezug auf die Gesamtbevölkerung. Die Befragungsergebnisse liefern jedoch bereits für sich genommen interessante Einblicke in das Verhalten und Erleben von Verbrauchern in der digitalen Welt.

Das eingesetzte Mittel der Selbstauskunft im Rahmen des Rechts auf informationelle Selbstbestimmung hat, anders als ausgehend von der gesetzlich verankerten Auskunftspflicht zu erwarten, nicht zu einer systematischen und verlässlichen Bestandsaufnahme geführt, aus der sich ein exemplarisches Bild des tatsächlichen Datenschattens ableiten ließe. Dafür ist im Wesentlichen die mangelnde Auskunftsbereitschaft und in Einzelfällen auch erkennbare Unfähigkeit zu einer zuverlässigen fristgerechten Zusammenstellung der Auskünfte durch die einbezogenen Unternehmen verantwortlich. Die zentrale Zielsetzung der Studie, den individuellen Datenschatten freiwilliger Probanden empirisch abzuleiten und Einflussgrößen auf

diesen zu benennen, hat sich mit dem gewählten Mittel der Selbstauskunft im Rahmen eines in seinen Kapazitäten und dem verfügbaren Zeitrahmen begrenzten Forschungsprojekt trotz erschöpfender Bemühungen nicht umsetzen lassen.

Die vor diesem Hintergrund ersatzweise vorgenommenen theoretischen Überlegungen und Schätzungen zum empirischen Datenschatten anhand eines gebildeten Datenschattenindex verstehen sich dennoch als Anregung für weitere Überlegungen. Sie führen zu Projektergebnissen, die über die ursprüngliche Planung hinausgehen. Dabei konnte gezeigt werden, dass weniger die physikalische Größe oder die Abhängigkeit des Datenschattens vom individuellen Handeln relevant sind, sondern ein Datenschatten zu einem Alltagsbegleiter nahezu jedes Bürgers geworden ist. Diese Datenspurten verlangen weniger nach reduzierenden Eingriffen alleine durch den Nutzer oder den Gesetzgeber, sondern vielmehr nach einer generellen regulatorischen und gesellschaftlichen Gestaltung. Dabei stehen alle Akteure in der Pflicht, um die mit vielen digitalisierten Prozessen verbundenen Vorteile und Innovationen nicht durch sozial nonkonformes Handeln zu gefährden.

9 Verwendete Literatur

Bitkom e.V. (2016): Was muss ich wissen zur EU-Datenschutzgrundverordnung? FAQ. Online unter: <https://www.bitkom.org/Presse/Anhaenge-an-PIs/2016/160909-EU-DSGVO-FAQ-03.pdf>, abgerufen am 26.08.2017.

Hermstrüwer, Yoan (2016): Informationelle Selbstgefährdung. Tübingen: Mohr Siebeck

Spitz, Malte und Biermann, Brigitte (2014): Was macht ihr mit meinen Daten? Hamburg: Hoffmann und Campe Verlag

SVRV Sachverständigenrat für Verbraucherfragen Hrsg. (2017): Der Wert persönlicher Daten. Ist Datenhandel der bessere Datenschutz? Online unter: http://www.svr-verbraucherfragen.de/wp-content/uploads/Open_Knowledge_Foundation_Studie.pdf, abgerufen am: 17.04.2018

Weigend, Andreas (2017): Data for the people. Wie wir die Macht über unsere Daten zurückerobern. Hamburg: Murmann Publishers GmbH

Anhang

Anschreiben an die Teilnehmer

Projektflyer

Datenschutzerklärung

Fragebögen (Auftaktbefragung, Recall-Befragung, Abschlussbefragung)

Unternehmensliste für die Selbstauskunft

Vollmachten an die Teilnehmer: intern, extern

Anschreiben der Unternehmen für die Selbstauskunft

Erinnerungsanschriften an die Unternehmen