



WERKSCHULHEIM FELBERTAL
ZUKUNFTS>CAMPUS

Werksschulheimstraße 11, 5323 Ebenau

Sind Daten ein Asset?

Die wirtschaftliche Bedeutung der Datensammlung im digitalen Zeitalter

Abschließende Arbeit

von

Simon Michael Wimmer

9b-Klasse

Betreuender Lehrer:

Robert Mutter

Ebenau, im Februar 2025

Abstract

Mit dem Advent des digitalen Handelns und Werbens entwickelten sich Nutzerdaten zur wertvollen Unternehmensressource, die mittlerweile über eigene Broker gehandelt werden und deren Verwendungszwecke inzwischen weit über die Marketingdimension hinausreichen.

Ziel dieser Arbeit ist es, die Prinzipien und Funktionsweisen dieses Marktes in wissenschaftlicher Manier, also sachlich und nüchtern, ohne voreilige Anschuldigungen, dem Lesenden verständlich näherzubringen.

Dazu wurden einige Kernforschungsfragen definiert: Zwischen welchen Datenarten wird grundlegend unterschieden und wie werden sie erhoben? Welche gesetzlichen Regulierungen gibt es bei deren Erhebung? Welche ökonomischen Maßstäbe erreicht der Markt des Datenhandels? Und welche Vor- respektive Nachteile birgt diese Art des Informationshandels?

Zur Beantwortung dieser Fragen wurde umfangreiche Fachliteratur herangezogen und kritisch ausgewertet. Die Fachliteratur zu diesem Thema wurde oftmals von reinen Betriebswirten oder reinen Datenschützern publiziert. Der Anspruch dieses Werkes ist es, Brücken zu schlagen und den Sachverhalt so neutral wie möglich darzulegen. Dazu wurden auch die Angebotsprospekte der Data Broker herangezogen und beleuchtet. Abgerundet wird dies durch Vernetzungen und eigene Thesen seitens des Autors.

Das Werk unterstreicht die zunehmende Bedeutung des Datenhandels in der modernen Wirtschaft, vergisst dabei aber nicht, auf etwaige Risiken hinzuweisen.

Vorwort

Die moderne Datenverarbeitung und die zwangsläufig dabei mitlaufende Datensammlung ist in unserer modernen Welt omnipräsent und unverzichtbar geworden. Schon lange, bevor ich meine technische Ausbildung begann, interessierten mich die fundamentalen Konzepte der Datenverarbeitung und ihre tiefgreifenden Auswirkungen auf unsere Welt.

Parallel dazu entwickelte ich ein starkes Interesse für die Wirtschaftslehre, was mich dazu veranlasste, die Datensammlung nicht nur aus technischer Sicht zu betrachten, sondern auch in ihrer makroökonomischen Dimension zu erforschen.

Bedanken möchte ich mich bei dieser Gelegenheit für die technisch sowie universalwissenschaftlich fundierte Ausbildung seitens des Werkschulheims Felbertal, für die fachlich hochspannenden Diskussionen, die ich mit meinem Ausbildungslehrer führen durfte und bei all jenen, die mir diesen Weg erst ermöglicht haben. Alle, die sich angesprochen fühlen, sind höchstwahrscheinlich gemeint.

Braunau am Inn, 16.02.2025

Simon Michael Wimmer

Inhalt

Abstract	I
Vorwort	II
Einleitung	1
0. Grundlagen der Datenwelt	3
0.1 Begriffsdefinition „Daten“	3
0.2 Kategorisierung von Daten	5
0.2.1 Informationstechnologische	5
0.2.2 Analytische	8
0.2.3 Betriebswirtschaftliche	9
0.2.4 Regulatorische	11
0.3 Personenbezogene Daten und ihre Korrelation mit privaten Daten	14
0.4 Datenqualität	17
1. Datensammlung	20
1.1 Methoden zur Sammlung von Nutzerdaten	20
1.1.1 Primär überwacht (Aktiv)	20
1.1.2 Primär unüberwacht (Passiv)	21
1.1.3 Sekundär	22
1.1.4 Überblick	24
1.2 Vergleich der Stärken und Schwächen der Verfahren	25
1.3 Gesetzliche Einschränkungen beim Sammeln von Nutzerdaten	27
1.3.1 USA	27
1.3.2 EU	28
2. Datenhandel	29
2.1 Daten werden zu Assets	29
2.2 Herkunft der gehandelten Daten	31
2.2.1 Erwerb bei externen Unternehmen	31
2.2.2 Öffentliche Quellen	31

2.2.3	Eigene Umfragen	31
2.2.4	Inferentielle Datengenerierung.....	32
2.3	Verwendung der gehandelten Daten.....	33
2.3.1	Wirtschaft und Finanzen	33
2.3.2	Marketing	33
2.4	Führende Data Broker und ihre Marktposition	36
2.5	Der Wert von Daten	38
2.6	Gesetzliche Einschränkungen beim Weiterverkauf von Daten.....	39
2.6.1	USA	39
2.6.2	EU	39
3.	Mehrwert und Risiken.....	41
3.1	Vorteile für Unternehmen	41
3.2	Vorteile für Konsumenten.....	42
3.3	Nachteile für Konsumenten.....	43
	Resümee	44
	Abbildungsverzeichnis.....	V
	Tabellenverzeichnis.....	V
	Literaturverzeichnis	VI

Einleitung

Kunden- und Nutzerdaten sind längst nicht mehr unscheinbares Nebenprodukt wirtschaftlichen Treibens, sondern essenzielle Wegführer in sämtlichen Prozessen. Von der Konzeption über die Entwicklung bis hin zur Vermarktung von Gütern und Dienstleistungen helfen sie maßgeblich dabei, die Probleme der Kundschaft zu identifizieren, bestehende Verwendungsmöglichkeiten zu optimieren oder das Produkt gezielt der richtigen Zielgruppe vorzustellen.

Datafizierung ist der Anglizismus für die zunehmende Beschreibung aller uns messbaren Phänomene durch digitale Daten. Seit der Erfindung der computergestützten Datenverarbeitung schreitet die Quantifizierung unserer Welt unaufhaltsam voran. Nicht verwunderlich, wenn man bedenkt, dass die Vermessung unserer Welt schon seit Auftauchen des Homo sapiens eine Rolle spielte.

Ultimatives Ziel der Datafizierung ist natürlich nicht die reine Aggregation von Petabytes an Informationen, sondern das Streben nach Erkenntnissen über den Menschen und seine Umwelt, die uns die Analyse solch großer Datenmengen liefern kann. Erkenntnisse über die Probleme, Hoffnungen und Träume der Menschen sind bedeutend für die Wirtschaft, ist es doch die Gründungsintention eines jeden Unternehmens, ein menschliches Bedürfnis zu befriedigen.

Wirtschaftstreibende rund um den Globus erkannten, dass ein einzelnes Unternehmen nur schwer in der Lage ist, die benötigte Menge an Daten zu sammeln, um solche Erkenntnisse gewinnen zu können. Ganz im Sinne der *Sharing Economy* wurde die Lösung im Zusammentragen von Daten aus verschiedensten Quellen gefunden. Gegen monetäre Gegenleistung stellen Unternehmen die von ihnen gesammelten Daten Dritten zur Verfügung. Diese akkumulieren Daten aus verschiedensten Quellen und können anschließend dem einzelnen Unternehmen Datensets von beeindruckender Granularität weiterverkaufen. So wurde das Geschäftskonzept von Data Brokern geboren.

Mittlerweile handelt es sich bei Data Brokern um eine Milliarden-Euro-schwere Industrie, die in allen geografischen Märkten Fuß gefasst hat. Sie werden von Megakonzerne gleichsam wie von mittelständischen Unternehmen angefragt, wenn diese unter einem Informationsdefizit leiden.

Doch trotz seiner offensichtlich so bedeutsamen Rolle in der Weltwirtschaft findet dieser Markt kaum öffentliche Beachtung. Ziel dieser Arbeit ist es deswegen, Licht in die dunklen Serverräume jener Datengiganten zu bringen.

Dazu widmen wir uns zuerst ganz grundlegenden Fragen wie: Was sind überhaupt Daten und welche Arten von ihnen gibt es? Nachdem dies geklärt wurde, werden die verschiedenen Methoden zum Sammeln von Daten erörtert. Sind die Daten erst einmal gesammelt, kann der Sprung zum eigentlichen Schwerpunkt dieses Werkes vollzogen werden: der Rolle von Daten als Assets, der Aufbau und die Funktionsweise des sie umgebenden Marktes, die dabei beteiligten Akteure und ihr ökonomisches Ausmaß. Natürlich dürfen bei alledem die rechtlichen Vorgaben in den verschiedenen Phasen nicht außer Acht gelassen werden. Dabei werden die wichtigsten Rahmenbedingungen in den zwei größten politischen Märkten (USA und EU) präsentiert. Zu guter Letzt muss eine wirtschaftliche Betrachtung natürlich mit einer Gegenüberstellung der größten Vor- und Nachteile aufseiten der Unternehmen wie auch Konsumenten angestellt werden.

Das Wissen zur Beantwortung der aufgeworfenen Fragen entstammt zahlreichen Werken der Fachliteratur und Gesetzestexten. Wichtig war dem Autor dabei, die teilweise eingengten Blickwinkel der Puristen aus Betriebswirtschaftslehre oder Datenschutz, die teilweise schon an eine *Déformation professionnelle* erinnern, zu überwinden, um dem Anspruch eines wissenschaftlich neutralen Werkes gerecht zu werden.

Gleichzeitig sei im Namen des Autors angemerkt, dass eine vollumfängliche Analyse aller Aspekte dieses Marktes den Rahmen einer einzelnen Arbeit vollends sprengen würde. Allein zu vermeintlichen Unterpunkten wie der juristischen Sachlage innerhalb der EU oder der Branchenanalyse ließe sich mehr als ein Werk verfassen. Intention dieses Werkes ist es, einen sinnstiftenden Überblick über den Markt als ganzheitliches Konstrukt zu verschaffen.

0. Grundlagen der Datenwelt

0.1 Begriffsdefinition „Daten“

Um die einleitende, essenzielle Fragestellung nach der Kategorisierung von Daten beantworten zu können, muss primär geklärt werden, welche Bedeutung sich hinter dem vielfach verwendeten und doch nichtssagenden Überbegriff „Daten“ überhaupt befindet.

Etymologisch betrachtet sind Daten die Pluralform von Datum. Das Lehnwort geht auf das lateinische Partizip Perfekt Passiv (PPP) von „dare“ zurück, was sich wiederum mit „geben“¹ übersetzen lässt. Übergebene Schriftstücke wurden mit dieser Einleitungsformel mit Orts- und Zeitangabe versehen.²

Die heutige Definition des Pluralnomens „Daten“ lautet wie folgt: „gesammelte Informationen; Angaben über Ort, Zeit und andere Fakten“³ Es handelt sich hierbei um eine interessante Form der Polysemie, bei der zumindest umgangssprachlich eine strikte Trennung der Bedeutung des Singulars und der des Plurals vorgenommen wird.⁴ Um die Unterscheidung der Ursprungsdefinition im zeitlichen und der neuartigen Definition im informationstechnischen Sinne zu erleichtern, werden für die Numeri der verschiedenen Bedeutungen Synonyme verwendet. So wird der Plural von „Datum“ als zeitliche Angabe oft als „Datumsangaben“ ausgedrückt und der Singular von „Daten“ als allgemeine Information oft als „Datenpunkt“.

So viel zur Bedeutung des Wortes im umgangssprachlichen Sinne. Im informationstechnischen Gebrauch ist jedoch eine genauere Aufschlüsselung notwendig, da hier stärker zwischen Daten, Informationen und Wissen unterschieden werden muss. Im Alltagsgebrauch werden diese Begrifflichkeiten, wie aus der Wörterbuchdefinition ersichtlich ist, fast schon synonymhaft verwendet. Dies ist auch logisch, wenn man bedenkt, dass wir größtenteils in ganzen Sätzen sprechen, in denen aus Datenpunkten Informationen gebildet werden und durch den Kontext eines Gesprächs Wissen geschaffen wird. Die Grenzen verschwimmen, wodurch auch die Differenzierbarkeit leidet.

¹ Lošek 2016, S. 198

² Kluge und Seebold 1999, S. 163

³ Pabst, Fussy und Steiner (Red.) 2016, S. 155

⁴ Pabst, Fussy und Steiner (Red.) 2016, S. 156

Bei der informationstechnologischen Speicherung wird Wissen jedoch zwangsläufig in seine atomaren Bestandteile zerlegt. Wer etwa ein Formular ausfüllt, gibt seine Daten weiter. Von einer Information kann jedoch erst gesprochen werden, wenn der eingegebene Datenpunkt mit einem weiteren Datenpunkt verknüpft wird, der ihn beschreibt. Die zweitgenannte Art von Daten wird als Metadaten bezeichnet und im nachfolgenden Kapitel genauer erörtert. Der Eintrag „Wimmer“ in einem Datenfeld ist nichtssagend und stellt noch keine Information dar. Es könnte sich hierbei genauso um einen Straßen- oder Produktnamen statt um einen Familiennamen handeln. Erst beim Hinzunehmen der Attributbeschreibung, in diesem Fall „Familiennamen“, kann von einer Information gesprochen werden. Die Vernetzung von Informationen kann schließlich als Wissen aufgefasst werden.

Der deutsche Wirtschaftsinformatiker Freimut Bodendorf hat für die Definition von Daten im informationstechnologischen Sinne ein vierstufiges Modell entwickelt, das die Begriffe Zeichen, Daten, Information und Wissen hierarchisch anordnet und beschreibt, wie durch Kontextualisierung, Semantik und Vernetzung aus einfachen Zeichen letztlich Wissen entsteht.

Laut ihm werden Daten „aus Zeichen eines Zeichenvorrats nach definierten Syntaxregeln gebildet [...]. Daten werden zu Informationen, wenn ihnen eine Bedeutung (Semantik) zugeordnet wird.“⁵

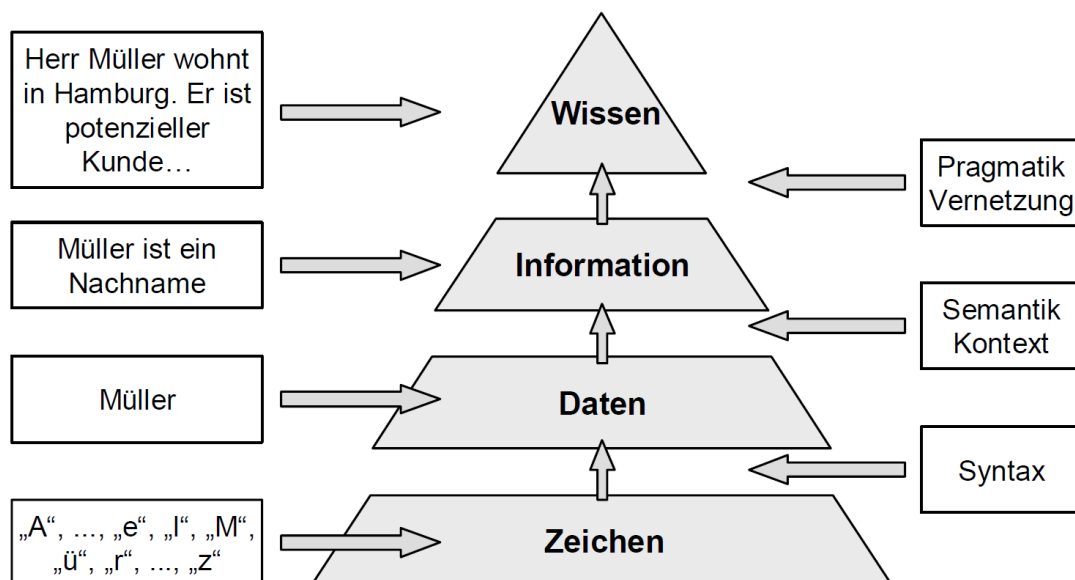


Abb. 1: Begriffshierarchie. Bodendorf 2006, S. 1

⁵ Bodendorf 2006, S. 1

0.2 Kategorisierung von Daten

Die Arten von Informationen und Fakten, die unter das Schlagwort Daten fallen, sind mannigfaltig. Deswegen ist es essenziell, Arten von Daten mit verschiedenen Eigenschaften voneinander zu differenzieren und sie dementsprechend zu kategorisieren. Vor allem für die spätere Auseinandersetzung bei der die betriebswirtschaftlichen Vorteile den potenziellen, individuellen Nachteilen gegenübergestellt werden, ist es unabdingbar eindeutige Namen für unvergleichbare Datenarten zu definieren, sodass beim Diskurs die gegenüberstehenden Parteien stets über die gleichen Konzepte reden.

Die Hauptproblematik in der Kategorisierung von Daten liegt jedoch in der Tatsache, dass mit dem Aufkommen der elektronischen Datenverarbeitung und dem Einzug ihrer in fast alle Arbeitsbereiche sich in den verschiedenen Arbeitsdomänen unterschiedlichste Systeme zur Kategorisierung etabliert haben, die teilweise Parallelen haben, aber fast niemals ident sind. Keines dieser Systeme ist objektiv als falsch oder richtig zu bewerten, ihre Essenz liegt vielmehr darin, die Art und Weise, wie Daten in der spezifischen Arbeitsdomäne gehandhabt werden und welche Zwecke sie darin erfüllen, zu reflektieren.

0.2.1 Informationstechnologische

In der informatischen Domäne, also in EDV-Abteilungen oder unter Datenbankadministratoren beispielsweise, spielt vordergründig die Speicherungs- sowie Erhebungsart eine Rolle. Bei der Speicherung wird konkret zwischen drei Strukturformen unterschieden: Strukturiert, semistrukturiert oder unstrukturiert.

Strukturierte Daten verfügen, wie der Name schon andeutet, über eine klar definierte Struktur. Es wird vorab definiert, in welcher Form die Daten abgelegt werden. Beispiele hierfür wären relationale Datenbanken oder CSV-Dokumente, bei denen die Daten jeweils in Tabellen eingetragen werden, wobei jedes Feld für einen spezifischen Datenpunkt und jede Zeile für einen Datensatz steht. Jeder Datenpunkt wird durch einen weiteren beschreibenden Datenpunkt kontextualisiert, wodurch er zur Information raffiniert wird. Im zuvor genannten Beispiel wäre dies der Titel der Tabellenspalte. Diese beschreibenden Datenpunkte werden auch als „Metadaten“ bezeichnet. Das aus dem

Altgriechischen entlehnte Präfix „Meta“ symbolisiert, dass sich das Element hierarchisch auf einer Ebene darüber befindet.⁶ Es handelt sich hierbei um Daten, die wiederum andere Daten charakterisieren, ordnen oder sie identifizierbar machen.⁷ Sie bilden eine weitere Unterkategorie von Daten, deren Existenz für alle strukturierten sowie semistrukturierten Daten notwendig ist. Strukturierte Daten können leicht maschinell weiterverarbeitet werden, weshalb ihnen in der Datenanalyse ein großer Stellenwert zugeschrieben wird.

Unstrukturierte Daten sind Anhäufungen von Daten ohne Metadaten. A priori können aus ihnen keine Informationen gewonnen werden. Stattdessen ist es notwendig, dass sie extrinsisch kontextualisiert werden. Beispiele für unstrukturierte Daten sind Grafiken oder dieser Text. Eine extrinsische Kraft kann ein Mensch sein, der seine linguistischen Vorkenntnisse in den Bereichen der Syntax und Semantik dazu verwendet, um diesen Text verstehen zu können. Maschinen können unstrukturierte Daten durch spezielle Algorithmen kontextualisieren.⁸

Semistrukturierte Daten sind eine Mischform der beiden zuvor genannten Strukturen. In einer Datei liegen sowohl Datenpunkte vor, die mittels Metadaten kontextualisiert wurden, als auch unstrukturierte Daten. Ein Beispiel hierfür wäre eine E-Mail, bei der es einerseits klar definierte Felder für Adressaten, Adressanten oder Zeitstempel gibt und andererseits den unstrukturierten Textinhalt. Weitere gängige Dateitypen für semistrukturierte Daten sind XML sowie JSON.⁹

Während Menschen in ihrer puren Daseinsform ausschließlich unstrukturierte Daten produzieren, werden sie bei der Interaktion mit informationstechnologischen Systemen fast immer zu semistrukturierten Daten transformiert. Die meisten Anwendungsprogramme kreieren eigenständig Metadaten zu den eingegebenen unstrukturierten Daten. In den meisten Fällen sind dies zumindest statistische Daten wie Zeitpunkt der Dateneingabe, eingebender Nutzer, Umfang der eingegebenen Daten und dergleichen.

Strukturierte Daten werden zumeist eigenständig von Maschinen generiert (z. B. Einnahme von Sensoren) oder wenn Maschinen den eingebenden Nutzer zur strukturierten Eingabe zwingen, wie dies bei Formularen der Fall ist.

⁶ Duden (Hrsg.) o.J.

⁷ Kranz 2024

⁸ Wuttke 2024

⁹ Seiter 2023, S. 24

Weiters kann bei der Speicherung von Daten nicht nur unterschieden werden, in welcher Struktur die Daten abgelegt wurden, sondern auch in welchen Datentypen.

Bei Datentypen im programmiertechnischen Sinne handelt es sich um Strukturen innerhalb einer Computer-Software, denen ein definierter Speicherplatz im Rechner und damit einhergehend ein möglicher Wertebereich zugeteilt wird.¹⁰ Des Weiteren werden den Datentypen Operationen (z. B. Addieren, Multiplizieren, Negieren etc.) zugeteilt, die mit ihnen durchgeführt werden können.¹¹

Die Datentypen können anschließend in Standardtypen und abgeleitete Typen aufgeschlüsselt werden.¹² Standardtypen sind von der Programmiersprache vorgegeben und haben sich über einzelne Sprachen hinweg im gesamten Informatikfachbereich etabliert.¹³ Es sei jedoch angemerkt, dass trotz der bekannten Trivialnamen und vergleichbarer Wertebereiche sowie Operationsmöglichkeiten die genauen Bereiche und Möglichkeiten zwischen verschiedenen Programmiersprachen variieren können. Abgeleitete Typen werden vom Programmierer eigenständig definiert. Der Speicherplatz sowie Wertebereich darf in den Grenzen der Sprache frei gewählt und es können sogar eigene Operationen erschlossen werden.

¹⁰ Microsoft (Hrsg.) 2023

¹¹ Lackes und Siepermann 2018

¹² Langer 1993, S. 21

¹³ Lange und Stegemann 1985, S. 31

Hier ein Überblick der gängigsten Datentypen und ihrer charakteristischen Wertebereiche:¹⁴

Datentyp	Beschreibung	Charakteristischer Speicherbedarf [Bit]	Charakteristischer Wertebereich
boolean	Wahrheitswert	1	True oder False
byte	Ganzzahl	8	-128 bis 127
date	Datum	8	-657.434 (1. Januar 100) bis 2.958.465 (31. Dezember 9999)
char	Zeichen	16	Unicode-Zeichen
int	Ganzzahl	32	-2^{31} bis $+2^{31} - 1$
float	Gleitkommazahl	32	$\pm 3,40282347 \times 10^{38}$

Tab. 1: Gängige Datentypen. Eigene Tabelle.

0.2.2 Analytische

Bei der analytischen Domäne handelt es sich um informationstechnologische Fachkräfte, die Daten gezielt dazu verwenden, um aus ihnen Erkenntnisse zu gewinnen. Die Nähe der analytischen zur informationstechnologischen Domäne spiegelt sich auch in der ähnlichen Kategorisierung von Daten wider.

Zur analytischen Erkenntnisfindung dienen stets rechnergestützte Algorithmen, für deren Ein- und Ausgangsgrößen die Programmiersprache spezifische Datentypen erwartet. Bei der Eingabe von Daten in einen Algorithmus muss deswegen stets sichergestellt sein, dass die einzugebenden Daten mit den von der Programmiersprache erwarteten Datentypen übereinstimmen.

Zusätzlich dazu hat sich im Fachbereich der computergestützten Analytik aber noch eine weitere Metakategorisierung etabliert: die Unterscheidung zwischen „nondependency-oriented data“ sowie „dependency-oriented data“. Nach Aggarwal beschäftigt sich diese Kategorisierung damit, ob zwischen den einzelnen Datenpunkten eine Be-

¹⁴ Fässler, Scheuner und Sichau 2024, S. 6

ziehung vorliegt. Ferner wird zwischen impliziten als auch expliziten Beziehungen differenziert. Eine implizite Beziehung liegt vor, wenn die Verbindung zwischen den Datenpunkten nicht direkt angegeben ist, jedoch durch die Natur der Daten erschlossen werden kann. Ein Beispiel hierfür sind Sensordaten, die in schneller Abfolge, beispielsweise einmal pro Sekunde, erfasst werden. Nachbarwerte innerhalb einer solchen Zeitreihe, wie zum Beispiel die Temperaturwerte einer Maschine, stehen implizit in Verbindung, da davon ausgegangen wird, dass sie keine abrupten Veränderungen zeigen. Ein starkes Abweichen deutet somit auf einen relevanten Fall hin. Eine explizite Beziehung liegt hingegen vor, wenn die Verbindung zwischen Datenpunkten direkt definiert ist, wie bei Netzwerkdaten. Ein typisches Beispiel hierfür sind Kontakte zwischen Mitgliedern auf Social-Media-Plattformen, bei denen die Verbindungen explizit durch Freundschaften oder Follower-Beziehungen dargestellt werden.¹⁵

Für Analytiker sind solche Beziehungen zwischen den Datenpunkten von großer Relevanz, da sie bei der Deutung der Analyseergebnisse berücksichtigt werden müssen, respektive die Wahl des am besten geeigneten Algorithmus stark vom Typus der Daten abhängt. So muss für eine aussagekräftige Ausreißeranalyse eine implizite Beziehung der Datenpunkte vorliegen.

0.2.3 Betriebswirtschaftliche

Speziell im Kontext der datengestützten Unternehmensführung hat sich das Bedürfnis ergeben, eine Datenkategorisierungsmethode zu entwickeln, die Aussage darüber trifft, welchen betriebswirtschaftlichen Zweck die Daten für die operativen Geschäftstätigkeiten eines Unternehmens aufweisen. In der populärwissenschaftlichen Managementliteratur, insbesondere im deutschsprachigen Raum, hat sich eine Kategorisierungsmethode mit sieben Datenarten etabliert. Je nach Quelle variieren die Bezeichnungen und Definitionen dieser Kategorien und teilweise kommt es zu Überschneidungen zwischen unterschiedlichen Kategorien. Angesichts dessen ist diese Methode weniger als wissenschaftlicher Ansatz zu betrachten, sondern vielmehr als Versuch von Unternehmensberatern, Führungskräften die Welt des Datenmanagements näherzubringen. Dieser Eindruck wird durch die Tatsache verstärkt, dass ein Großteil der zugänglichen Dokumentationen zu dieser Methode auf Publikationen von Unternehmensberatungen basiert.

¹⁵ Seiter 2023, S. 57

Im Nachfolgenden hat sich der Autor darum bemüht, die unterschiedlichen Auslegungen dieses Modells auf einen gemeinsamen Nenner zu bringen und dadurch einen Einblick basierend auf hoher Abstraktion in diese Methodik zu geben. Selbstverständlich können sich beim Nachschlagen in einzelnen Fachwerken Differenzen in einzelnen Spezifika ergeben. Am meisten Ähnlichkeit weist diese Gliederung zur Beschreibung durch Daniel Liebhardt auf.¹⁶

1. **Metadaten:** beschreiben wie zuvor bereits ausgeführt andere Daten. Klassisches Beispiel hierfür ist der Titel einer Tabellenspalte.
2. **Referenzdaten:** kategorisieren allgemeine Geschäftsentitäten. Es handelt sich hier ebenfalls wie bei den Metadaten um Daten, die andere Daten einordnen. Im Unterschied zu jenen liegen hier allerdings Werte vor, die sich selten ändern und meistens von externen Quellen bezogen werden. Dabei handelt es sich z. B. um Flughafenkürzel oder eine Liste der Bundesländer eines Staates. Man spricht hierbei des Öfteren auch von „stabilen Informationen“.
3. **Unternehmensweite Strukturdaten:** repräsentieren die Organisation und den Aufbau eines Unternehmens, einschließlich seiner Produkte, Dienstleistungen und Verantwortlichkeiten.
4. **Transaktionsstrukturdaten:** beschreiben den Aufbau und die Grundstruktur von Geschäftstransaktionen. Dazu gehören typischerweise die beteiligten Geschäftsentitäten wie Kunden und Produkte.
5. **Inventardaten:** beschreiben die Vermögenswerte eines Unternehmens und deren Quantität, wie z. B. Lagerbestände, Kontostände oder die Anzahl der Immobilien.
6. **Transaktionsdaten:** beschreiben die kommerziellen sowie rechtlichen Geschäftsaktivitäten eines Unternehmens. Es liegt dabei immer ein Austausch von geistigen oder physischen Gütern zugrunde. Beispiele hierfür wären Rechnungen (ein- und ausgangsseitig), Wertpapierkäufe, Rückgaben oder Gutschriften.

¹⁶ Liebhart 2010, S. 21

7. **Auditdaten:** protokollieren die einzelnen Schritte einer Transaktion und dienen zur späteren Überprüfung jener. Durch sie wird Transparenz bezüglich der Datengenerierung gewährleistet und eine Möglichkeit zur Überprüfung der Korrektheit geschaffen¹⁷

0.2.4 Regulatorische

Eine vierte Domäne, der in diesem Werk später noch eine besondere Bedeutung zukommt, ist die regulatorische. Gemeint ist damit die legislative Gewalt, die, im Zuge der zunehmenden Kommerzialisierung von Nutzerdaten von Privatpersonen, für sich erkannt hat, dass es im Sinne der Persönlichkeitsrechte zu Beschränkungen bei der Datensammlung kommen muss.

Bei dieser Kategorisierungsmethode handelt es sich wohl um die populärste, die auch abseits von Juristen oder Politikern im Allgemeinen von Privatpersonen am häufigsten verwendet wird. Für sie ist es von besonderem Interesse, welche persönlichen Informationen mittels Daten ausgedrückt werden können. Man geht hier also einen Schritt weg von der puren Form oder Erhebungsart, hin zu den Erkenntnissen, die daraus gezogen werden können.

Am wichtigsten ist in dieser Domäne die Definition der personenbezogenen Daten. Zu diesen gehören alle Datensätze, die Informationen zu natürlichen Personen enthalten und die durch Identifizierung eindeutig zu einzelnen Individuen zuzuordnen sind. Die Identifizierbarkeit ist dabei in der Datenschutzgrundverordnung äußerst breit definiert: *„als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind, identifiziert werden kann“*.¹⁸

Des Weiteren nennt die DSGVO bestimmte Unterkategorien von personenbezogenen Daten, die als besonders schützenswert einzustufen sind. Dazu gehören unter anderem:

¹⁷ dataspot. GmbH kein Datum

¹⁸ Artikel 4 Absatz 1 der Verordnung (EU) 679/2016

- sensible personenbezogene Daten, bei deren Verarbeitung „[...] erhebliche Risiken für die Grundrechte und Grundfreiheiten auftreten können. Diese personenbezogenen Daten sollten personenbezogene Daten umfassen, aus denen die rassische oder ethnische Herkunft hervorgeht [...]“¹⁹. Diese Daten dürfen nur in äußersten Ausnahmefällen verarbeitet werden.
- „‘genetische Daten‘ [...] zu den ererbten oder erworbenen genetischen Eigenschaften [...], die eindeutige Informationen über die Physiologie oder die Gesundheit dieser natürlichen Person liefern und insbesondere aus der Analyse einer biologischen [...] gewonnen wurden;“
- „‘biometrische Daten‘ [...] zu den physischen, physiologischen oder verhaltens-typischen Merkmalen einer natürlichen Person, die die eindeutige Identifizierung dieser natürlichen Person ermöglichen [...]“
- „‘Gesundheitsdaten‘ [...], die sich auf die körperliche oder geistige Gesundheit einer natürlichen Person, einschließlich der Erbringung von Gesundheitsdienstleistungen, beziehen und aus denen Informationen über deren Gesundheitszustand hervorgehen“²⁰

Die regulatorische Sichtweise wurde dabei stark von der Arbeit diverser Datenschützer geprägt und spiegelt dabei deren Denkweise wider: Die Gefahr wird nicht durch die Daten per se geschaffen, sondern durch die Konklusionen, die durch die Vernetzung und Analyse verschiedener Datenattribute gezogen werden können. Unzählige Datenschutzadvokaten haben eine Vielzahl von Kategorisierungsmethoden entworfen, deren genaue Examinierung den Rahmen dieses Werks sprengen würde. Im Nachfolgenden wird exemplarisch eine besonders simple, aber ebenso weitverbreitete Methode, der NSA-Chefsyndika April Falcon Doss rezipiert.

1. **Eigentumsdaten** – „was wir haben“: sind Daten über unsere digitalen sowie analogen Besitztümer. Die Kaufhistorie eines Nutzers, der Waren über einen Onlineversandhändler erwirbt, gehört zu dieser Kategorie, genauso wie eine Liste der Anwendungsapplikationen, die er auf seinem Mobiltelefon installiert hat oder das Modell des digitalen Endgeräts, mit dem er ins Internet einsteigt.
2. **Verhaltensdaten** – „was und wie wir handeln“: beinhalten Muster unseres alltäglichen Handelns, mithilfe dessen auch unser zukünftiges Verhalten prognostiziert werden kann. Hierbei handelt es sich meistens um von Bestandsdaten

¹⁹ Erwägungsgrund 51 der Verordnung (EU) 679/2016

²⁰ Artikel 4, Absatz 13 – 15 der Verordnung (EU) 679/2016

abgeleitete Daten. Einzelhändler können so etwa durch Loyalitätsprogramme das Kaufverhalten ihrer Kundschaft aufzeichnen. Aus Bewegungsdaten können Navigationsapplikationen unseren Tagesablauf rekonstruieren.

3. **Identifizierungsdaten** – „wer wir sind“: machen uns als natürliche Personen eindeutig identifizierbar. Es handelt sich einerseits um alphanumerische Kennzeichen, die (in ihrem Geltungsbereich; z. B. innerhalb eines Staates, einer Organisation etc.) pro Person nur einmal zugeordnet werden. Darunter fallen Sozialversicherungsnummern, Kontonummern, Kennzeichen oder auch Telefonnummern. Aber auch allgemeine Körpereigenschaften, die exakt einer Person zugeordnet werden können. Hierzu zählen insbesondere biometrische Daten (Fingerabdrücke, Iriserkennung etc.) genauso wie genetische Sequenzen.
4. **Psychografische Daten** – „was wir glauben, wissen und denken“: Die Abgrenzung zu Verhaltensdaten erfolgt auf der psychologischen Ebene. Während man unter Verhaltensdaten zumeist unterbewusste, regelmäßige Vorgänge versteht, die in der analogen Welt von Mitmenschen frei beobachtet werden können, versteht man unter psychografischen Daten, kognitive Entscheidungen oder persönliche Eigenschaften, die bewusst kommuniziert oder auch verheimlicht werden können. In der Psychologie spricht man vom Persönlichkeitsaspekt der Privatperson, also von Aspekten, deren Offenlegung im Entscheidungsbereich der Person selbst liegt.²¹ In diese Gruppe fallen politische Einstellungen, Sexualpräferenzen oder Wertevorstellungen.²²

Diese Gruppen sind keineswegs als exklusiv abgrenzbar einzustufen. Die Grenzen sind transitional und oftmals werden Informationen, die einer Gruppe zuordenbar sind, durch Aufbereitung von Daten einer anderen Kategorie gewonnen. Meistens handelt es sich bei den Rohdaten, die durch eine Datenerhebung entstehen, um Eigentumsdaten, die durch Analyse auf Verhaltensmuster oder psychografische Daten hindeuten. Außerdem ist es mit genügend Daten und entsprechenden Algorithmen möglich, eine natürliche Person mithilfe von Nicht-Identifizierungsdaten zu identifizieren. Die Intimität der Daten und potenzielle Risiken bei missbräuchlicher Nutzung steigen gleichsam der Nummerierung.

²¹ Lahmer 2023, S. 146

²² Doss 2020, S. 11 - 31

0.3 Personenbezogene Daten und ihre Korrelation mit privaten Daten

Auch wenn keine eindeutige Definition für den Terminus „private Daten“ besteht, werden nachfolgend damit Informationen bezeichnet, deren potenzielle Risiken bei missbräuchlicher Nutzung so umfangreich sind, dass der gemeine Nutzer Bedenken bezüglich ihrer Veröffentlichung hat und sie deswegen vorzugsweise privat halten möchte.

Erwähnenswert dabei ist, dass sich unser Verständnis von privaten Daten mit zunehmender Digitalisierung grundlegend verändert hat und viele Daten, die in der analogen Welt als öffentlich angesehen werden, in digitalen Umgebungen als privat klassifiziert werden.

Während allgemeine biometrische Daten wie Gesichtszüge oder Stimmprofile in der analogen Welt für jede passierende Person frei sicht- wie hörbar sind und somit den öffentlich zugänglichen Daten angehören, so werden sie in der DSGVO als besonders sensible als auch schützenswerte Daten gelistet.

Während es gemeinhin als vollkommen legitim gilt, wenn der heimische Bäcker basierend auf dem vorangegangenen Kaufverhalten seiner Kundschaft ihnen personalisierte Empfehlungen und Angebote gibt, so werden mediale Debatten über die Legitimität von personalisierten Anzeigen auf den Seiten von Onlineversandhändlern geführt.

Die Antwort auf die Frage, warum in den zwei vorgestellten Beispielen den identen Daten in unterschiedlichen Situationen (computergestützte oder menschliche Verarbeitung) unterschiedliche Risiken zugeschrieben werden, ist schnell zu finden: Während es wohl keinem Passanten möglich sein wird, durch Hören einer Stimme diese so zu imitieren, dass er sich vor einem Dritten ohne Zweifel als diese Person ausgeben kann, so ist das Generieren von sogenannten Deepfakes, die genau dies darstellen, mittlerweile auch für Laien möglich. Des Weiteren verfügen Unternehmen durch Cloud-Computing über die Möglichkeit, gewaltige Mengen an Nutzerdaten zu aggregieren und sie vereint zu analysieren. In Sekundenschnelle können damit Prognosen zum Kundenverhalten generiert werden, von deren Extensivität der heimische Bäcker nicht einmal träumen kann.

Es lässt sich somit feststellen, dass die potenziellen Risiken für Nutzer nicht durch die Daten per se ausgehen, sondern durch neuartige Analyse- und Prognosemodelle, sowie die technischen Plattformen, die deren Ausführung erst ermöglichen.²³

Erstaunlich ist jedoch auch, dass parallel dazu bei manchen Daten eine gegensätzliche Entwicklung stattgefunden hat. In der Welt der Datenschützer hat sich hierfür der Begriff Datenschutzparadoxon etabliert. Er beschreibt das Phänomen, dass Nutzer trotz ihrer (teilweise eklatanten) Bedenken bezüglich Privatsphäre im digitalen Raum gleichzeitig dazu bereit sind, persönliche Daten online zu teilen, die sie im analogen Leben nur äußerst ungern preisgeben würden.²⁴

Davon sind besonders Informationen über medizinische Zustände, Sexualpräferenzen oder politische Meinungen betroffen. Direkte Vergleichsstudien – die alle wissenschaftlichen Standards erfüllen – zwischen Informationsoffenheit in digitalen versus analogen Umgebungen sind nur schwer umzusetzen. Forschende der *Brookings Institution* haben jedoch durch Analyse der Google Suchstatistiken oder Verkaufsdaten von Onlineversandhändlern interessante Zusammenhänge in diesem Kontext aufgedeckt, die zumindest Raum für spekulative Theorien schaffen.²⁵

Diese Erkenntnis, die zumeist zunächst als kontraintuitiv erscheint, lässt sich auf den Fakt zurückführen, dass viele Menschen sich online anonym fühlen. Die Privatsphäre im digitalen Leben ist deutlich abstrakter als das Pendant im realen. Während wir uns im analogen Leben häufig vor der Fragestellung wiederfinden, ob es uns angenehm ist, wenn die Person, derer wir gerade in die Augen sehen, gewisse Informationen über uns erhält, so spielt es für uns oftmals keine Rolle, ob ein Webserver dieselbe Information über uns speichert. Auch wenn eine natürliche Person diese Informationen analysiert, so werden wir mit dieser höchstwahrscheinlich niemals in direkten Kontakt treten.²⁶

²³ Doss 2020, S. 35 - 36

²⁴ Gerber, Gerber und Volkamer 2018, S. 226

²⁵ Wittes und Liu 2015, S. 11 - 20

²⁶ Doss 2020, S. 83

Abschließend sei zum Thema private Daten angemerkt, dass auch Regulierungsbehörden auf die Widersprüchlichkeit, Daten, die im analogen Leben als öffentlich angesehen werden, im digitalen Leben zu reglementieren, aufmerksam wurden. Denn auch für den heimischen Bäcker, der manuelle Listen über das Kaufverhalten seiner Kundenschaft führt, gilt mittlerweile die DSGVO, genauso wie für den international tätigen Onlineversandhändler. Ausschließlich das menschliche Gedächtnis und die daraus entspringenden Gedanken bleiben selbstverständlich frei, wodurch das Recht auf Vergessen werden hier an seine Grenzen stößt.

0.4 Datenqualität

Für die betriebswirtschaftliche Verwendbarkeit von Daten kommt vor allem ihrer Qualität eine gewichtige Bedeutung zu. Neben dem Datenumfang ist dies der zweite große Faktor, der über den Wert eines Datensets, auf den später noch genauer eingegangen wird, entscheidet.

Klassischerweise wird Datenqualität in die drei Dimensionen Vollständigkeit, Korrektheit und Konsistenz gegliedert.²⁷

- **Vollständigkeit** meint bei den strukturierten Daten, dass so viele Attribute eines Datensatzes als nur möglich definiert sind. Am Beispiel einer Datentabelle bedeutet dies, dass möglichst alle angelegten Felder ausgefüllt sind. Weiters impliziert Vollständigkeit die Anlegung von ausreichenden Attributen, um der Komplexität der Analyseaufgabe entsprechend gerecht zu werden. Auch bei unstrukturierten Daten wird hierbei vorausgesetzt, dass sich in der Datensammlung über genügend Datenpunkte zur ausreichenden Lösung der Analyseaufgabe befinden. Ein bekanntes Praxisbeispiel sind Pflichtfelder bei Formularen, die dazu dienen, die Vollständigkeit zu erhöhen. Es gibt jedoch auch die Ausnahme, bei der fehlende Attribute ein Indiz für einen relevanten Fall im Sinne der Analyse sein können.
- **Korrektheit** ist die wohl die am einfachsten zu erklärende, aber am schwierigsten umsetzbare Qualitätsdimension. Um korrekte Analyseergebnisse ableiten zu können, sind korrekte Eingabedaten unerlässlich. Validierungsregeln wie die Bedingung, dass eine E-Mail-Adresse das @-Symbol sowie eine Top-Level-Domain enthalten muss, tragen Sorge dafür, dass offensichtlich falsche Informationen erst gar nicht eingetragen werden können. Werden Daten primär durch manuelle Eintragungen generiert, sind Qualitätsschulungen für die Eintragenden empfehlenswert.²⁸
- **Konsistenz** stellt die Verknüpfbarkeit mit anderen Datenquellen sowie die Verwendung von Analysealgorithmen sicher. Es handelt sich dabei um die Formkorrektheit, also um die Abspeicherung in den richtigen (vorab definierten) Datentypen.²⁹

²⁷ Seiter 2023, S. 25

²⁸ Schlageter und Stucky 1983, S. 287

²⁹ Kähler 1990, S. 144

In der Fachliteratur überschneiden sich oftmals die Definitionen von Korrektheit oder Konsistenz, teilweise werden diese beiden Aspekte auch zusammengelegt, was ebenfalls einen vollkommen legitimen Ansatz darstellt. Der Autor spricht sich jedoch für eine Trennung zwischen intrinsischer (bezüglich des Inhalts) und extrinsischer (bezüglich der Form) Korrektheit aus. Auch finden sich in der Fachliteratur verschiedene Synonyme für den Begriff Konsistenz, dazu gehören Einheitlichkeit oder Zugänglichkeit, um nur ein Paar zu nennen.

Ein weiteres Qualitätskriterium, welches des Öfteren unterschätzt und infolgedessen unbeachtet bleibt, ist die Aktualität. Die Bedeutung dieses Kriteriums hängt stark vom jeweiligen Anwendungszweck ab, steigt jedoch mit dem zunehmenden Alterungsprozess unseres Digitalzeitalters. Um diesen Aspekt einfließen zu lassen, wird teilweise das Prinzip des mathematisch-biologischen Verfallsprozesses implementiert. Dabei werden numerische Werte, deren Größe von Bedeutung ist, mit einer natürlichen Exponentialfunktion multipliziert.

$$Q(w, A) := e^{-\text{Verfall}(A) \times \text{Alter}(w, A)}$$

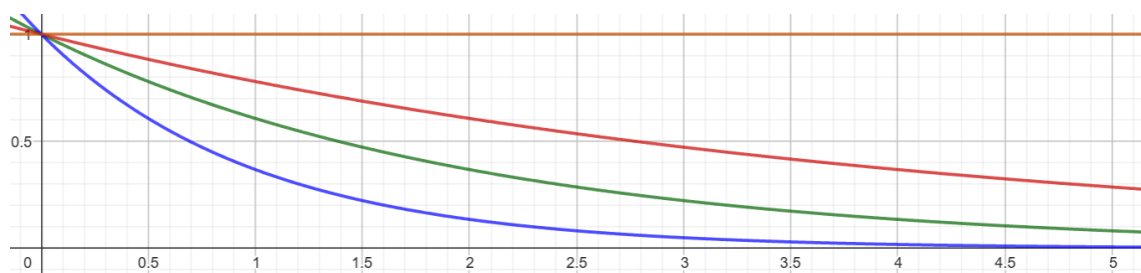
w ... Attributwert in einer Datenmatrix Q

A ... Attribut der Datenmatrix Q

Verfall(A) ... Verfallsrate des Attributs A

Alter(w,A) ... Alter des Attributwerts w

Bei Metriken von unbeschränkter Lebensdauer nimmt der Verfall den Wert 0 an, wodurch die Exponentialfunktion zur Konstante mit Wert 1 wird.³⁰



Verfall = 0 Verfall = 0,25 Verfall = 0,5 Verfall = 1

Abb. 2: Verfallsfunktionen. Eigene Grafik.

³⁰ Seiter 2023, S. 62 - 63

Um die Datenqualität zu verbessern, können drei grundlegende Strategien angegeben werden:

- **Nacherhebung:** zur Erhöhung der Vollständigkeit werden Forschungsanstrebungen gestartet, um zumindest für besonders relevante Attribute die korrekten Daten in Erfahrung zu bringen. Diese Strategie ist oftmals mit hohem Ressourcenaufwand verbunden, da erneute Feldforschung notwendig ist.
- **Datenausschluss:** Instanzen mit geringer Aussagekraft (durch viele fehlende Attribute) werden aus dem Datensatz entfernt. Hierbei handelt es sich um die einfachste Strategie, die jedoch eine Reduktion des Datenumfangs und damit auch der Aussagekraft darauf basierender Analysen mit sich zieht.
- **Imputation:** hierbei wird versucht, durch andere Attribute des Datensatzes eine wahrscheinliche Größe für den fehlenden Wert abzuleiten. Die Sinnhaftigkeit dieser Vorgehensweise hängt maßgeblich vom Kontext des Datensets ab. Bei homogenen Datenansammlungen kann dies durchaus eine valide Strategie darstellen, die noch dazu kostengünstig ist, werden Datensets jedoch künstlich homogenisiert, führt dies zu falschen Analyseergebnissen.³¹

³¹ Aggarwal 2015, S. 35

1. Datensammlung

1.1 Methoden zur Sammlung von Nutzerdaten

Am Anfang jeder Datenanalyse steht die Beschaffung der Rohdaten. Nachdem nun die grundlegenden Begrifflichkeiten in der Welt der Daten geklärt wurden, widmen wir uns jetzt diesem fundamentalen Schritt.

Grundlegend kann bei der computergestützten Beschaffung von Daten zwischen Primär und Sekundärerhebung unterschieden werden. Bei der primären Erhebung wird der Betreiber eines Online-Services selbst als wesentlicher Datenerfasser tätig. Bei der sekundären wird eine außenstehende Entität Erfasser der Daten, die mit oder ohne Wissen des Service-Betreibers Daten aus dessen Service extrahiert. Die sekundäre Erhebung setzt stets voraus, dass zuvor eine primäre Erfassung durch den Betreiber stattfand.

Bei der primären wird ferner zwischen überwachter und unüberwachter Erhebung differenziert. Bei der überwachten Erhebung ist dem Nutzer bewusst, dass all seine durchgeführten Aktionen der Datengenerierung dienen. Bei der unüberwachten ist die Datengenerierung ein Nebenprodukt von Aktionen mit anderen Grundintentionen. Die Begriffe überwacht und unüberwacht entspringen der Begebenheit, dass unüberwachte Erhebungen fast ausschließlich automatisiert stattfinden, während es einige überwachte Erhebungsmethoden gibt, bei denen der Datenerheber den Datengeber über den gesamten Sammelprozess begleitet und überwacht. Angemerkt sei auch, dass überwachte Erhebungen oftmals mit Methoden der unüberwachten Erhebung kombiniert werden, in diesen Fällen spricht man jedoch ausschließlich von überwachten Erhebungen, da die zu analysierende Person in jedem Fall weiß, dass sie analysiert wird.³²

1.1.1 Primär überwacht (Aktiv)

Fragebögen: Wie bei seinem analogen Pendant werden auch beim digitalen Fragebogen Menschen darum gebeten, eine Reihe von quantitativen oder qualitativen Fragen zu beantworten. Der Vorteil dieser Methode liegt in der Natur der direkten Fragen, wodurch hierbei auch äußerst spezifische Daten gewonnen werden können.

³² Höchstötter 2009, S. 176

Computerlaborexperimente: Hierbei geht es darum, das Verhalten von Nutzern im Umgang mit digitalen Endgeräten in Erfahrung zu bringen. Die Probanden werden in einem Computerlabor vor verschiedenste Aufgaben gestellt, die sie mit den ihnen verfügbaren Technologien abarbeiten sollen. Dabei stehen sie unter konstanter Beobachtung der Versuchsleiter.³³ Da dabei in der Regel keine personenbezogenen Daten erzeugt werden, liegt diese Methode nicht im Fokus dieses Werkes, als allgemeine Methode, um Nutzerverhalten verstehen zu können, soll sie jedoch dennoch erwähnt werden.

1.1.2 Primär unüberwacht (Passiv)

Logdateien: Webserver und andere Computerprogramme protokollieren für gewöhnlich die von ihnen ausgeführten Aktionen, sowie die von Clients mitgelieferten Informationen, in sogenannten Logdateien.³⁴ Für die Generierung solcher Logdateien wurden einheitliche Standards entwickelt. Zu den bekanntesten und am häufigsten verwendeten gehören das „*NCSA Common Logfile Format*“ sowie das „*W3C Extended Log Format*“. Auf diese Weise kann zum Beispiel erhoben werden, zu welchem Zeitpunkt ein Nutzer welche Unterseite ausgerufen hat, auf welchen Wegen er dorthin gelang (Weiterleitung; eng. „referral“) oder mit welchem Browser er darauf zugriff.³⁵ Die Daten werden meistens chronologisch in Textdateien gespeichert. Die Daten solcher Logdateien können für statistische Auswertungen über die Verwendung eines Online-Services herangezogen werden.

Nutzungsprofilaten: Während Logdateien sich standardmäßig strikt auf Serveraktionen beschränken und die Speicherung in standardisierten Formaten stattfindet, ist Nutzungsprofilaten ein Überbegriff für alle Daten, die bei Nutzung eines Online-Services mitprotokolliert und direkt einem spezifischen Nutzer zugeordnet werden. Die Speicherung findet zumeist in Datenbanken statt, bei denen die erhobenen Daten mit den Nutzerprofilen verknüpft werden. Typischerweise fallen darunter Informationen zu betrachteten oder gekauften Artikeln, abgeschlossene Transaktionen oder abgegebene Bewertungen.

³³ Höchstötter 2009, S. 177

³⁴ Hegewald 2017, S. 22

³⁵ Höchstötter 2009, S. 180

Interaktionsanalyse: Beschreibt alle Methoden, bei denen automatisiert getestet wird, auf welche Art und Weise ein Nutzer mit einem Online-Service interagiert. Eine der meistverbreiteten Taktiken ist dabei, mittels Zählpixeln die auf einer Website mit der Maus zurückgelegten Wege zu messen. Zählpixel sind kleine, transparente Grafiken, die einen Aufruf auslösen, sobald der Mauszeiger über sie bewegt wird. Einerseits werden Interaktionsanalysen ähnlich wie Computerlaborexperimente dazu genutzt, um herauszufinden, wie intuitiv digitale Produkte für Endverbraucher sind, andererseits ist es denkbar durch Kennzahlen wie Klickgeschwindigkeit und dergleichen Rückschlüsse auf demografische Daten (wie das Alter) des Nutzers zu ziehen.

1.1.3 Sekundär

API: API ist das englische Akronym für „*Application Programming Interface*“ und wird im Deutschen häufig als „*Programmierschnittstelle*“ übersetzt. Prägnant zusammengefasst handelt es sich bei APIs um standardisierte Protokolle, die den strukturierten Datenaustausch zwischen verschiedenen Softwareanwendungen ermöglichen. Ohne APIs wären die Vielfalt und Interoperabilität moderner IT-Infrastrukturen kaum denkbar. Größtenteils dienen sie der effizienten Kommunikation zwischen Programmen und ermöglichen Funktionen wie den Zugriff auf Cloud-Dienste oder die Integration externer Systeme. Im Kontext der sekundären Datenerhebung erlauben sie es jedoch auch, auf bereits existierende Datensätze von Drittanbietern zuzugreifen, indem sie automatisierte Anfragen an externe Plattformen oder Datenbanken ermöglichen. Dies erleichtert die Aggregation und Analyse von Informationen aus verschiedenen Quellen.³⁶ Anschaulich wird diese Form der sekundären Datenerhebung durch das Beispiel von *Meta* und *Cambridge Analytica*, welches mittlerweile auch in der Populärmedienwelt Bekanntheit erlangte. Durch Metas *Facebook Open API* war es dabei *Cambridge Analytica* bis Mai 2015 möglich, Informationen aus den Facebook-Profilen von Nutzern auszulesen.³⁷

Web Scraping: Umfasst alle Methoden und Technologien zur (automatischen) Extraktion von Inhalten eines Online-Services, mit Ausnahme von APIs. Während APIs von Online-Service-Betreibern Dritten zur Verfügung gestellt werden, um Daten zu erheben, so werden Technologien, die in diese Gruppierung fallen, oftmals ohne Wissen des Betreibers verwendet. Web Scraping eignet sich als kostengünstige und einfache

³⁶ Goodwin 2024

³⁷ Isaak und Hanna 2018, S. 56 - 59

Methode zum Erheben großer Datensätze (Big Data), deren Informationen bereits von anderen publik gemacht wurden. Bei Verwendung dieser Technologie ergeben sich des Öfteren jedoch auch legale als auch ethische Bedenken, vorwiegend im Kontext des Urheberrechts.³⁸

Open Data: Forschungsinstitutionen, gemeinnützige Organisationen oder Regierungskörper veröffentlichen immer häufiger Rohdaten für die freie Distribution. Diese als „Open Data“ bekannten Datensätze eignen sich meistens für statistische Analyse-zwecke. Im Gegensatz zu APIs oder Web Scraping sind die so erhältlichen Daten deutlich selektierter, da die Distribuenten die Daten im Regelfall mit einer klaren Inten-tion veröffentlichen.³⁹

Third-Party-Cookies: Cookies sind Informationen, die von einem Online-Service lokal im Browser des Clients abgespeichert werden. Ursprünglich wurden sie dafür konzi-piert, dass Informationen und Einstellungen, die ein Nutzer in einem Webservice ein-getragen hat, über die Dauer einer einzelnen Sitzung hinweg gespeichert werden kön-nen, ohne, dass eine Anmeldung des Nutzers notwendig wäre. So kann sich eine Website beispielsweise merken, welche Sprache der Nutzer ausgewählt hat, ob An-zeigeoptionen geändert oder welche Artikel in einen Warenkorb gelegt wurden. Die Funktionen, die hinter diesem Grundgedanken stehen, bezeichnet man mittlerweile als notwendig oder funktional und ihre Cookies als „*First-Party-Cookies*“. Wichtig zu er-wähnen ist auch, dass Cookies grundsätzlich nur von der Domain gelesen und bear-beitet werden können, von der sie gesetzt wurden. First-Party-Cookies werden dabei vom Betreiber der besuchten Website gesetzt.

Mit der Kommerzialisierung des Internets und der zunehmenden Verbreitung digitaler Werbung entstand ein neuer Anwendungsbereich für Cookies. Webseitenbetreiber stellen dabei gegen eine finanzielle Vergütung Werbeflächen Dritten zur Verfügung, die eigene Inhalte einbinden. Diese Drittanbieter sind in der Lage, über ihre eingebet-teten Inhalte sogenannte Third-Party-Cookies zu setzen. Im Gegensatz zu First-Party-Cookies, die von der besuchten Domain stammen, werden Third-Party-Cookies von externen Entitäten verwaltet. Sie ermöglichen es, Nutzeraktivitäten über verschiedene Webseiten hinweg zu erfassen, indem sie unter anderem die besuchten Domains oder

³⁸ Krotov, Johnson und Silva 2020, S. 555 - 557

³⁹ Murray-Rus 2008, S. 1 - 2

Interaktionen speichern. Dies setzt jedoch voraus, dass der Drittanbieter auf mehreren, voneinander unabhängigen Webseiten Werbeeinbindungen besitzt. Nur dann kann er seine Cookies auf verschiedenen Seiten setzen und Nutzer wiedererkennen. Dies mag sich nach einem großen Hindernis anhören, tatsächlich verfügen aber weniger als 1 % der Werbenetzwerke Zugriff auf etwa 75 % der Websites mit Werbeeinbindungen.⁴⁰ Durch die Zusammenführung dieser Daten entsteht ein detailliertes Nutzungsprofil, das insbesondere für gezielte Werbeausspielung genutzt wird. Einerseits können Werbenetzwerke das gesammelte Surfverhalten analysieren, um individualisierte Anzeigen in Echtzeit bereitzustellen, um damit die Effektivität ihrer Werbekampagnen zu optimieren, andererseits können die so gewonnenen Nutzerdaten auch weiterverkauft werden.⁴¹

1.1.4 Überblick

Nachfolgend eine schematische Abbildung der eben erwähnten Methoden als Vorschlag zur einfachen Zuweisung:

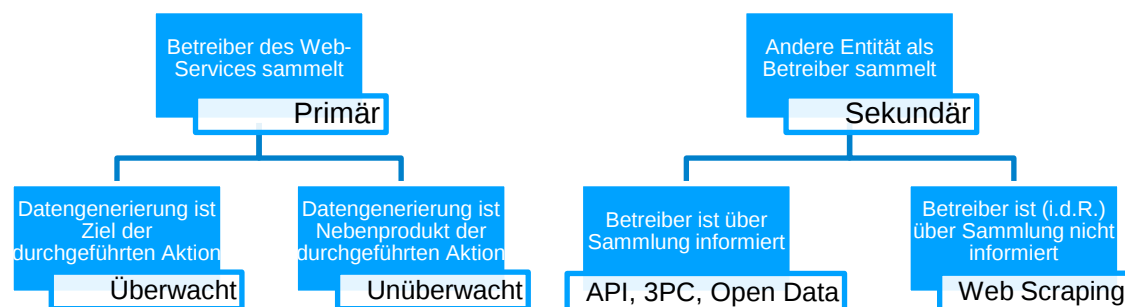


Abb. 3: Schematische Einteilung der Datenerhebungsmethoden. Eigene Grafik.

⁴⁰ Cahn, et al. 2016, S. 891

⁴¹ Moryl 2024

1.2 Vergleich der Stärken und Schwächen der Verfahren

Primär überwachte Erhebungen sind vorrangig für spezifische Fragestellungen geeignet, da durch die direkte Interaktion mit den Datengebern die benötigten Daten explizit erfragt werden können. Dem gegenüber stehen die hohen Kosten, da diese Erhebungen selten standardisiert sind und die dafür benötigten Formulare oder Experimente zuerst von den Forschenden konstruiert werden müssen.

Eine weitere Schwierigkeit stellt die Generierung größerer Datenmengen dar, da sie hier nicht wie bei den anderen Erhebungsmethoden als passiv hervorgebrachtes Nebenprodukt anderer Tätigkeiten anfallen, sondern aktiv durch die Nutzer generiert werden. Die Datengeber müssen deswegen zur Teilnahme incentiviert werden, was einen weiteren Kostenfaktor darstellt.

Zudem ist bei der Auswertung der Ergebnisse der psychologische Hawthorne-Effekt (engl. Observer effect) einzubeziehen. Jener besagt, dass Menschen, die wissen, dass sie beobachtet werden, respektive ihre Eingaben analysiert werden, ihr Verhalten ändern oder ihre Meinungen anders präsentieren, um stärker ihrem Selbstbildnis als auch der sozialen Gruppenerwartung gerecht zu werden.⁴² Zum Beispiel bei sozialen oder finanziellen Fragen besteht häufig das Bedürfnis, sich selbst in ein besseres Licht zu rücken, weswegen die Wahrheit verzehrt wird.

Des Weiteren erlaubt diese Erhebungsmethode bewusste Falschangabe von Informationen. So können Teilnehmer, die etwa ihre Adresse oder Telefonnummer nicht weitergeben wollen, schlichtweg eine falsche Information eingeben.

⁴² Anteby und Khurana o.J.

Dahingegen sind unüberwachte Methoden, durch ihre fast ausschließlich automatisiert stattfindende Erhebung, äußerst kostengünstig und zuverlässig. Während jeder Laienutzer in ein Formularfeld eine falsche Adresse einfügen kann, so ist es ihm deutlich schwieriger, seine IP-Adresse zu ändern, die ebenfalls Informationen zu seinem Standort liefert.

Durch die fehlende Möglichkeit, aus der Datensammlung in Logdateien oder Transaktionsdaten auszusteigen, können große Datenmengen zu virtuell jedem Nutzer gesammelt werden.

Der größte Nachteil ergibt sich aus dem limitierten Kontext. Während bei Umfragen die Analysesubjekte ihre Handlungen rational erklären können, steht hier der Forschende vor dem Problem, die Handlungen richtig deuten zu müssen.

Sekundäre Erhebungsmöglichkeiten erlauben das Erlangen großer Datenmengen in kürzester Zeit, mit geringerem Ressourcenaufwand. Die Zuverlässigkeit der Daten hängt jedoch stark von der Vertrauenswürdigkeit der Quelle ab und es ergeben sich zusätzlich legale Bedenken, wie etwa bezüglich des Urheberrechts.

1.3 Gesetzliche Einschränkungen beim Sammeln von Nutzerdaten

Nach etlichen Skandalen bezüglich Datenlecks und breiten öffentlichen Diskussionen zur ethischen Vertretbarkeit der Erhebung von sensiblen, privaten Daten, erkennen immer mehr Regulierungsorganisationen rund um den Globus die Notwendigkeit, strikte Auflagen zu definieren, die unser loses Verständnis von Privatsphäre schützen sollen.

Die Ansätze differieren teils massiv von Judikative zu Judikative. Es ist nicht die Absicht dieses Werkes, einen ausführlichen Leitfaden zum Umgang mit den verschiedenen Rahmenbedingungen zu geben. Vielmehr geht es darum, den Lesenden einen Einblick in die verschiedenen Betrachtungsweisen zu verschaffen, sodass er im Anschluss über ein grundlegendes Verständnis bezüglich der breit gefächerten Strategien verfügt.

1.3.1 USA

Als Ursprungsland vieler technischer Errungenschaften hat es sich diese Judikative verdient, als Erstes beleuchtet zu werden, vor allem da mit diesem datenschutztechnischen Blickwinkel viele der Produkte, die wir täglich nutzen, entwickelt werden.

Als Einstieg ist es erwähnenswert zu vermerken, dass das US-Recht in seiner Originalfassung das Konzept der Privatsphäre nicht kannte. Der Begriff tauchte in keinem zur Unabhängigkeitserklärung erschienenen Gesetz auf. Vielmehr wurde die Privatsphäre indirekt durch Eigentumsverhältnisse garantiert. Niemand durfte beobachten, was jemand in den eigenen vier Wänden treibt, da das Eindringen auf fremdes Grundstück verboten war. Kein unbeteiligter Dritter durfte den Inhalt eines Briefes von Fremden lesen, da er das in Form eines Briefes materialisierte Eigentum entwendet hätte. Die Haltung, dass die Privatsphäre auch ohne Übertreten physischer Grenzen beschnitten werden kann, wurde erst später in den Gesetzbüchern inkorporiert.⁴³

Diese stückweisen Ergänzungen wurden jedoch stets sehr eng ausformuliert. So verpflichtet der *Health Insurance Portability and Accountability Act (HIPPA)* nur spezifische Kategorien von Unternehmen oder Organisationen (wie Krankenhäuser oder Versicherungsgesellschaften) zum Schutz der ihnen mitgeteilten Gesundheitsinforma-

⁴³ Doss 2020, S. 42 - 49

tionen. Unternehmen, die nicht in eine der definierten Kategorien fallen, jedoch dennoch Gesundheitsinformationen speichern, wie der Anbieter einer Smartwatch beispielsweise, müssen sich nicht an die strengen Regulierungen halten.⁴⁴

Ein weiteres Spezifikum des US-amerikanischen Rechts ist der föderalistische Ansatz. Die Rechte auf Privatsphäre im Internet differieren von Bundesstaat zu Bundesstaat. So trat in Kalifornien, dem amerikanischen Zentrum der IT-Branche, 2020 der *California Consumer Privacy Act (CCPA)* in Kraft. Ein breit gefächertes Gesetz, welches sich augenscheinlich auch an den Prinzipien der europäischen DSGVO orientiert.⁴⁵

1.3.2 EU

Möchte man den europäischen Blickwinkel auf Privatsphäre verstehen, so muss man sich Europas dunkle Vergangenheit bezüglich Diskriminierung und Ausgrenzung in Erinnerung rufen. Vor allem die Datensammlung der Nationalsozialisten, bei denen Menschen aufgrund ihrer religiösen Zugehörigkeit, politischen Einstellung, sexuellen Orientierung oder ähnlichen persönlichen Charakteristiken verfolgt wurden, dient als finsternis leuchtendes Mahnmal des Missbrauchs personenbezogener Informationen. Im Zuge dessen ist es nicht verwunderlich, dass solche für Missbrauch sensible Daten nur in äußersten Ausnahmen erst verarbeitet werden dürfen. Außerdem steht es jeder in der EU ansässigen Person zu, eine Auskunft der über sie gespeicherten Daten zu erhalten und sie bei Bedarf ändern oder löschen zu lassen.

Außerdem erwähnenswert ist Europas ganzheitlicher („*One size fits all*“) Ansatz. Daten müssen entsprechend ihrer Sensibilität von jeder Entität gleichsam geschützt werden. Kritiker werfen diesem Ansatz jedoch vor, es handle sich um eine bürokratische Überregulierung, die Europas Position bei technischen Entwicklungen gefährde. Auch das heiß diskutierte *Recht auf Vergessenwerden* wirft ethische Fragestellungen in Bezug auf Meinungs- und Pressefreiheit auf. Diese Diskussionen liegen jedoch außerhalb der Grenzen dieses Werkes.⁴⁶

⁴⁴ Doss 2020, S. 24 - 25

⁴⁵ Doss 2020, S. 62 - 63

⁴⁶ Doss 2020, S. 235 - 242

2. Datenhandel

2.1 Daten werden zu Assets

Für die meisten Unternehmen beginnt die Erhebung von Nutzerdaten mit der Intention, die angebotenen Dienstleistungen und Artikel durch datengestützte Analysen zu verbessern oder besser zu vermarkten. Der Return on Investment (Abk. ROI; dt. Kapitalrendite), also die Amortisation des für die Datensammlung eingesetzten Kapitals⁴⁷, wird demnach durch Umsatzsteigerung, aufgrund höheren Absatzes als Folge von kompetitiveren Produkten oder Erschließung neuer Kundensegmente, geschaffen.

Mit dem Durchbruch des Online-Marketings, also der Werbung auf digitalen Plattformen, die vor allem auf Personalisierung der angezeigten Werbebotschaften setzt, entstand jedoch ein Markt für Daten per se. Während die Marketingbranche früher Werbung auf Makro-Ebene entwickelte, indem sie repräsentative Marktforschung betrieb, um einen möglichst großen Anteil der Rezipienten anzusprechen, setzt sie heute auf Modifikationen auf Mikro-Ebene, um den einzelnen Rezipienten genau die Art von Werbung zu zeigen, auf die er mit größter Wahrscheinlichkeit reagiert.

In diesem Kontext reicht es nicht aus, über repräsentative Stichdaten zu verfügen, die den größten gemeinsamen Nenner der Zielgruppe widerspiegeln, sondern es müssen persönlich generierte Daten von jedem potenziellen Konsumenten verfügbar sein. Essenziell dabei ist die Granularität der erhobenen Daten, also deren Detailtiefe und Genauigkeit in Bezug auf individuelle Nutzerpräferenzen und Verhaltensmuster. Je feiner die Daten segmentiert sind, desto gezielter können Werbebotschaften personalisiert und in Echtzeit angepasst werden.⁴⁸

Für ein einzelnes Unternehmen wäre es eine Mammutaufgabe, eine solch granulare Datenerhebung neben der eigentlichen betriebswirtschaftlichen Tätigkeit vorzunehmen. Aus dieser Marktlücke sind Unternehmen hervorgetreten, die es sich zur dezierten Aufgabe gemacht haben, Daten aus verschiedensten Quellen zusammenzutragen und an Kunden weiterzuverkaufen. Diese Unternehmen werden als „Data Brokers“ bezeichnet.

⁴⁷ SRH Fernhochschule GmbH o. J.

⁴⁸ Doss 2020, S. 82 - 86

Der Broker ist in der Finanzwelt der angloamerikanische Begriff für Handelsunternehmen, die als Geschäftsvermittlung zwischen Käufern und Verkäufern agieren. Dort handelt es sich meistens um Assets wie Wertpapiere oder Rohstoffe.⁴⁹

Mit dem Stichwort „Asset“ sind wir auch schon bei der eher rhetorischen Leitfrage dieser Arbeit angelangt: Im Allgemeinen ist ein Asset ein beliebiger Vermögensgegenstand, der an diversen Märkten veräußert werden kann. Im Spezifischeren handelt es sich auch um einen Fachterminus des Rechnungswesens, der einen Posten der Aktivseite mit bestimmten Merkmalen beschreibt. Zu diesen Merkmalen zählt, dass diese Ressource *„durch eine frühere Unternehmensaktivität geschaffen wurde, in der Verfügungsgewalt des Unternehmens steht und von der erwartet wird, dass sie einen zukünftigen Nutzen für das Unternehmen erbringen wird.“* Außerdem wird im Vergleich zum teilweise verwandten Begriff des Vermögensgegenstandes keine Einzelveräußerbarkeit vorausgesetzt.⁵⁰

Diese Charakteristiken treffen eindeutig auf die Ressource(n) „Daten“, wie sie im nullten Kapitel dieses Werkes definiert und kategorisiert wurden, zu. Durch die in 1.1 angeführten Unternehmensaktivitäten werden sie generiert, es handelt sich um immaterielle Vermögenswerte, die sich in der Gewalt des besitzenden Unternehmens befinden und durch die mittels Analysen geschaffenen Erkenntnisse ergibt sich ein Wettbewerbsvorteil, der sich monetär auswirken kann. Des Weiteren ist eine Veräußerbarkeit, wenn auch nicht im Maßstab eines einzelnen Datenpunktes, über Data Broker möglich. Die Klärung der Einzelheiten dieser Möglichkeit ist ein Schwerpunkt der Arbeit und wird nachfolgend behandelt.

⁴⁹ Heldt 2018

⁵⁰ Hinz und Weiß 2020

2.2 Herkunft der gehandelten Daten

Daten werden immer mithilfe der in Kapitel 1.1 ausgeführten Methoden erhoben. Die Daten, die auf den Portalen von Data Brokern zum Verkauf angeboten werden, sind dementsprechend ursprünglich auf selbem Wege erhoben worden. Die Besonderheit von Data Brokern ist, dass sie hauptsächlich sekundäre Erhebungsformen verwenden. Sie greifen auf eine Vielzahl von Quellen zurück, tragen die dort erhaltenen Informationen zusammen und bieten sie anschließend gesammelt zum Weiterverkauf an oder erfüllen damit spezifische Aufträge. Zusammengefasst fungieren sie als zentrale Informationsauskunftei für Unternehmen mit einem Informationsdefizit.

Konkret können bei den Quellen drei gängige Kategorien unterschieden werden:

2.2.1 Erwerb bei externen Unternehmen

Wie in der Einleitung dieses Kapitels kurz angeschnitten, sammelt ohnehin eine beträchtliche Anzahl der Unternehmen, vorwiegend die, die im reinen Online-Geschäft tätig sind, mittlerweile Kundendaten zur Verbesserung der eigenen Produkt- oder Servicepalette. Der Verkauf von allgemeinen Daten, deren Veräußerung keinen Wettbewerbsnachteil für das eigene Unternehmen mit sich zieht, stellt ein lukratives Nebengeschäft dar. In der Fachwelt spricht man hierbei des Öfteren von einer Datenkommerzialisierung.

2.2.2 Öffentliche Quellen

Vor allem staatliche Institutionen publizieren im Sinne des demokratischen Transparenzgedankens eine Menge von personenbezogenen Informationen. Hierunter fallen unter anderem die Firmenbücher und Gewerberegister, das Grundbuch oder das Insolvenzregister. Aber auch private Körperschaften, wie die Herold Business Data GmbH, publizieren mit ihrem öffentlich zugänglichen Telefonbuch personenbezogene Daten.⁵¹

2.2.3 Eigene Umfragen

Die hier einzige erwähnte primäre Erhebungsmethode ist wohl auch die am geringsten vorkommende, die nur den größeren Data Brokern mit entsprechenden Ressourcen vorbehalten ist. Für jene Data Broker gewinnt diese Quelle jedoch zunehmend an Bedeutung und so führte etwa der Data Broker Epsilon mit 20 Millionen teilnehmenden Haushalten die größte kundenspezifische Umfrage in Nordamerika durch.⁵²

⁵¹ Neally 2019, S. 30 - 46

⁵² Epsilon Data Management LLC 2025

Dies sind Methoden, wie Daten grundsätzlich von Nutzern erhoben werden können. Sie stellen aber nicht die einzige Möglichkeit dar, wie Daten über Nutzer generiert werden können.

2.2.4 Inferentielle Datengenerierung

Eine Inferenz ist eine aus einem logischen Regelsystem gebildete Schlussfolgerung. Bei der inferentiellen Datengenerierung wird versucht, durch statistische Analyse aus bekannten, erhobenen Daten weitere Attribute eines Betrachtungsobjekts abzuleiten. So könnte beispielsweise bei einem Konsumenten, der häufig Herrenpflegeprodukte erwirbt, abgeleitet werden, es handle sich um ein männliches Individuum. Bei einem Käufer eines bestimmten Videospieltitels könnte das Alter auf unter 25 Jahre geschätzt werden.⁵³

Da es sich hierbei nicht um eine tatsächliche Erhebung von Nutzerdaten, sondern schon um das Ergebnis einer Analyse dieser handelt, wurde diese Möglichkeit nicht in Kapitel 1.1 erwähnt. Dennoch ist sie im Kontext der Angebotsbetrachtung von Data Brokern von immenser Bedeutung, da das Kerngeschäft vieler führender Broker mittlerweile die Bereitstellung solcher Erkenntnisse über Menschen ist.

Diese Datenart stellt für viele Betriebswirte sicherlich die interessanteste der verfügbaren dar. Dennoch sollte bei der Verwendung dieser mit Vorsicht umgegangen werden. Inferentielle Daten versuchen oftmals Informationen über Alter, Geschlecht, Gesundheit, sexuelle Orientierung, politische Ansichten, Hobbys oder dergleichen wiederzugeben. De facto gibt es aber keinen Algorithmus, der solche Zuordnungen treffsicher vollziehen kann, und die Verwendung gewisser Zuordnungen wirft ebenso ethische Bedenken auf.⁵⁴ Vor allem, wenn den Algorithmen gängige Vorurteile zugrunde liegen. Nachprüfungen der durch prädiktive Analysen generierten demografischen Nutzerdaten eines marktführenden Data Brokers haben ergeben, dass diese gerade mal eine Genauigkeit von 50 % aufweisen.⁵⁵ Es kann jedoch davon ausgegangen werden, dass die Genauigkeit mit zunehmender Datafizierung und längeren Historien zunimmt.

⁵³ Doss 2020, S. 68 - 69

⁵⁴ Venkatadri, et al. 2019, S. 1920 - 1930

⁵⁵ Ruschemeier 2023, S. 30

2.3 Verwendung der gehandelten Daten

Dem Kauf von Daten, respektive den daraus resultierenden Informationen, geht immer ein Informationsdefizit des Käufers voraus, welches durch den Verkäufer gemindert oder gelöst werden kann. Die Arten der Informationsdefizite sind vielfältig, klassischerweise beschränken sich die Anbieter jedoch auf die zwei Hauptkategorien Wirtschaft/Finanzen und Marketing.

2.3.1 Wirtschaft und Finanzen

Anbieter, die in diesem Bereich agieren, sammeln klassischerweise Informationen zur Identität von Personen (Vorname, Adresse, Geburtsdatum, Geschlecht, Sozialversicherungsnummer etc.) zum Kaufverhalten (welche Produkte in welcher Stückzahl auf welchen Plattformen erworben werden) und zum Vermögen.

Einerseits werden darauf basierend Identifikationstools zum Verifizieren der Identität von Kunden und Überprüfen der angegebenen Daten zum Schützen vor Betrug angeboten. Andererseits werden daraus Prognosen zum Bewerten der Bonität abgeleitet.⁵⁶

2.3.2 Marketing

Fügt man zu den zuvor genannten Daten zum Kaufverhalten beispielsweise noch weitere persönliche Daten wie die aus Social-Media-Profilen hinzu, können Rückschlüsse auf Interessen und Hobbys gezogen werden. Darauf basierend können Werbe-Profile generiert werden, mithilfe dessen den einzelnen Personen individuelle Werbung zugeschaltet werden kann. Gewöhnlicherweise werden diese Werbe-Profile in aussagekräftige Segmente unterteilt und der Kunde kann sich ein Segment aussuchen, das seiner Zielgruppe am stärksten entspricht.

⁵⁶ Beckmann 2024

Hierbei muss jedoch noch weiter differenziert werden, ob die gesammelten Nutzerdaten tatsächlich in Rohform verkauft werden und der Käufer damit die Werbung selbst schaltet, oder ob der Käufer den Data Broker dazu beauftragt, Werbung in diesem Kundensegment zu schalten. Mittlerweile treten die größten Data Broker häufig nicht mehr als reine Datenhändler auf, sondern haben sich selbst zur Werbeagentur transformiert, die die gesamte Wertschöpfungskette einer Werbekampagne bedient.

Im Falle eines tatsächlichen Datenaustausches, wie es dem ursprünglichen definitionsgemäßen Sinn eines Data Brokers entspricht, händigt der Data Broker dem Käufer eine extensive Datenliste aus, in der sich in der Regel neben eindeutigen Identifikatoren (Vorname, Adresse etc.) werberelevante Daten wie persönliche Interessen, Bildungsgrad oder Vermögenswerte befinden.⁵⁷

Der amerikanische Data Broker „Epsilon Data Management LLC“ gibt folgendes Beispiel für ein Nutzerprofil an:

Demografie & Lifestyle	34-jährig, verheiratet, weiblich, zwei Kinder
Gesundheit & Wellness	Aktiv in Yoga, Pilates und Laufen. Probiert häufig verschiedene Bio-Diätpläne aus. Leidet unter Allergien und hohem Blutdruck und besitzt keine Zahnzusatzversicherung.
Finanzen & Vermögen	Hauseigentümer, Haushaltseinkommen zwischen 125k – 150k \$, Nettovermögen des Haushalts zwischen 250 – 500k \$, aktiver Investor, zwei Fahrzeuge im Haushaltsbesitz: Audi (2015) und Chrysler (2016)
Kaufdaten auf Händler-ebene	Ausgaben über 5.000 \$ bei United Airlines, konsumfreudiger Käufer bei Brooks Brothers und Marshalls, kauft häufig Vitamine von Tide, Crest und Nature's Way

⁵⁷ Doss 2020, 85 - 88

Einkaufsdaten von Abacus Cooperative [Anm. hierbei handelt es sich um eine von Epsilon initiierte Transaktionsdatenbank, die von über 3.000 partizipierenden Unternehmen betrieben wird] ⁵⁸	Vielkäufer mittelpreisiger Damenbekleidung, im vergangenen Jahr mehr als 1.000 \$ verteilt auf 15 Transaktionen ausgegeben
Neigungen & Absichten	auf der Suche nach einem neuen Luxusauto, möchte wahrscheinlich Bio-Lebensmittel kaufen, den Telefonanbieter wechseln, eine Mehrfachversicherung abschließen, einen Familienurlaub machen und für Kinderprojekte spenden.
Einstellungen & Vorlieben	Tech Early Adopter, kauft häufig bei Amazon und anderen Online-Händlern ein

Tab. 2: Beispiel für ein Data Broker Nutzerprofil. Eigene Tabelle mit Informationen entnommen aus Epsilon Data Management LLC 2025

Laut eigenen Angaben verfügt die Epsilon Data Management LLC über umfangreiche Datenlisten zu 250 Millionen US-Amerikanern, was nahezu die gesamte konsumierende Bevölkerung der USA abdeckt. Jeder dieser Datensätze soll mit einem verifizierten Echtnamen verknüpft sein.⁵⁹

⁵⁸ Epsilon Data Management LLC 2025

⁵⁹ Epsilon Data Management LLC 2025

2.4 Führende Data Broker und ihre Marktposition

Laut einem Bericht von Maximize Market Research sind Acxiom (USA), Experian (Irland), Equifax (USA), CoreLogic (USA) und TransUnion (USA) die Data Broker mit dem größten Marktanteil. Es wird global von bis zu 5.000 aktiven Data Brokern ausgegangen. Insgesamt wurde das Marktvolumen der Branche im Jahr 2024 auf 270,4 Mrd. USD geschätzt und eine durchschnittliche jährliche Wachstumsquote (CAGR) bis 2032 von 7,25 % prognostiziert, wodurch die Branche in diesem Jahr ein Marktvolumen von 473,35 Mrd. USD erreichen soll.

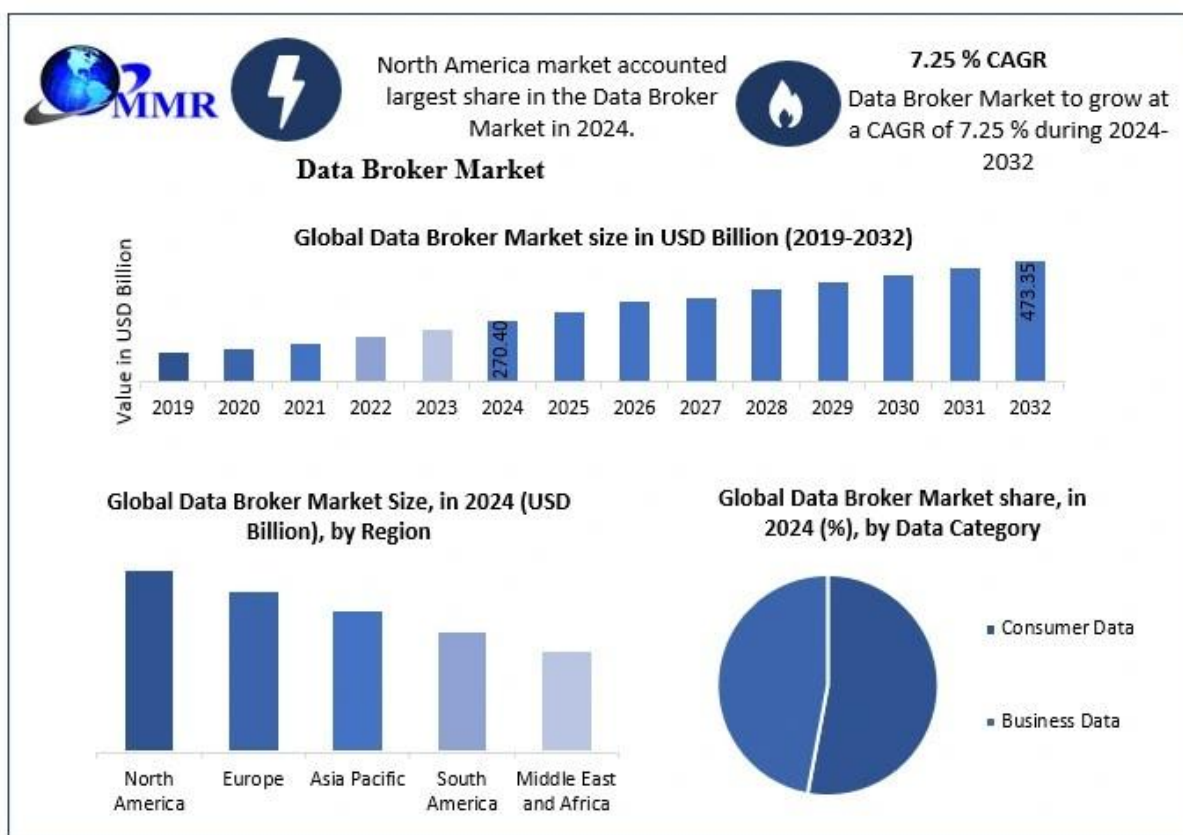


Abb. 4: Wirtschaftliche Kennzahlen der Data Broker Branche. Maximize Market Research 2025, abgerufen von <https://www.maximizemarketresearch.com/wp-content/uploads/2020/04/Data-Broker-Market-1.webp>

Zum Vergleich: Das Marktvolumen des Mobiltelefonsektors, dem primären Endgerät zur Generierung von Nutzerdaten, wurde für 2024 auf 649,13 Milliarden USD geschätzt. Prognosen zufolge soll es mit einer durchschnittlichen jährlichen Wachstumsrate von 6,8 % bis 2032 auf rund 1,1 Billionen USD anwachsen.⁶⁰ Damit übersteigt das

⁶⁰ SkyQuest Technology Group 2025

Volumen des Mobiltelefonmarktes das des Datenhandels mehr als doppelt, weist jedoch eine um 0,45 Prozentpunkte geringere jährliche Wachstumsrate bis 2032 auf. Die Telekommunikationsbranche weist ein Volumen von 1,99 Billionen USD und ein prognostiziertes CAGR von 6,5 % bis 2030 auf.⁶¹

Mit einem Marktanteil von 34,68 % ist Nordamerika die am stärksten vertretene geografische Region, dicht gefolgt von Europa und Asien. Vor allem in diesen Regionen hat die Branche aber mit legislativem Gegenwind zu kämpfen. In Europa vorwiegend mit der allseits bekannten *Datenschutzgrundverordnung* (DSGVO). In den Vereinigten Staaten geht die Regulierung der Branche durch den föderalistischen Ansatz schlepender voran, einzelne Bundesstaaten wie Kalifornien haben mit dem *California Consumer Privacy Act* (CCPA) beispielsweise bereits Vorstöße gewagt.

Demgegenüber steht das steigende Interesse der Wirtschaftstreibenden, die jährlich wachsenden Datenmengen auch abseits von Produktverbesserungen monetär geltend zu machen und beim Wechsel vom analogen zum digitalen Marktplatz das Wissen über individuelle Kunden nicht zu verlieren.⁶²

Ein anderer Bericht von *Research and Markets* spricht sogar von einem geschätzten Marktvolumen von 389,765 Mrd. \$ und sieht bis zum Jahr 2029 eine jährliche Wachstumschance von 7,58 %.⁶³

⁶¹ Grand View Research Inc. 2024

⁶² Maximize Market Research PVT. LTD. 2025

⁶³ Research and Markets 2024

2.5 Der Wert von Daten

Der Wert der gehandelten Daten ist nur äußerst schwierig zu bemessen, da vor allem die Big Player der Branche verstärkt auf ein Dienstleistungsangebot im Abo-Modell oder Marketingkomplettlösungen setzen. Beispielsweise hängt der Abonnementpreis für die Dienstleistung der Identitätskontrolle von der Größe des Kunden, der durchschnittlichen Anfrageanzahl pro Abrechnungszeitraum und der Komplexität der Abfragen ab. In besonders schwierigen Fällen bieten die führenden Data Broker ebenfalls an, im Falle, dass die angefragte Person sich nicht in den bereits bestehenden Datensätzen befindet, kostenpflichtige Anfragen bei Regierungsinstitutionen oder eigene Recherchen durchzuführen. Dieses Vorgehen führt zu einem äußerst variablen Preis, der in den meisten Fällen individuell mit dem Kunden verhandelt wird. Zum Zeitpunkt der Recherche für dieses Werk gibt keiner der marktführenden Data Broker Pauschalpreise für seine Dienstleistungen an.

Data Broker, die vollständige Datensätze in Rohform weiterverkaufen, machen in der Regel keine Angaben zu ihren Preisen. Dies dürfte unter anderem darauf zurückzuführen sein, dass sie in den vergangenen Jahren vermehrt im Fokus investigativer Recherchen standen. So berichtete das Forbes Magazine, dass der auf Gesundheitsdaten spezialisierte US-amerikanische Data Broker *MEDbase 200* Listen mit sensiblen Gesundheitsinformationen für 79 USD pro 1.000 Datensätze anbietet, also weniger als 8 Cent pro Person.⁶⁴

Eine weitere wichtige Kennzahl ist der Gegenwert solcher Information. Auch hier gilt wieder der Grundsatz, dass genaue Werte aufgrund der hohen Abstraktion nur schwierig anzugeben sind. Der Data Broker Experian schätzte jedoch 2012 den Wert einer E-Mail-Adresse auf £84,50. Die Angabe basiert auf dem durch Mail-Werbung genierten Umsatz über die Lebensdauer der Adresse.⁶⁵

⁶⁴ Hill 2013

⁶⁵ Jenkins 2012

2.6 Gesetzliche Einschränkungen beim Weiterverkauf von Daten

2.6.1 USA

Vor allem im Gründungsland der meisten marktführenden Broker läuft der Handel in rechtlich sicheren Gewässern ab. Auch wenn die Rechtsprechung zunehmend restriktiver wird, wie in Kapitel 1.3 ausgeführt, und sich einige Anbieter in *Senate Hearings*⁶⁶ zu ihren Geschäftstätigkeiten rechtfertigen mussten, so gilt im US-amerikanischen Recht weiterhin die Devise die Datensammlung im Allgemeinen nicht zu minimieren, sondern Nutzer besser über diese aufzuklären oder ihnen das Recht zu geben, sich individuell der Datensammlung zu entziehen.⁶⁷

So werden etwa im *California Consumer Privacy Act (CCPA)*, einem Vorreitergesetz im amerikanischen Datenschutz, dem Nutzer die Rechte eingeräumt, Auskunft über die gesammelten Daten zu erhalten, den Weiterverkauf an Dritte abzulehnen oder die gesammelten Daten löschen zu lassen.⁶⁸ Die allgemeine Forderung nach *Zweckbindung* oder *Datenminimierung*⁶⁹, zwei Konzepte, die in der Europäischen Datenschutzgrundverordnung klar definiert wurden, gibt es hier jedoch nicht. Die Anwendbarkeit des Gesetzes auf Data Broker wird durch die kalifornische Generalstaatsanwaltschaft explizit erwähnt, man verweist jedoch hierbei lediglich, dass Nutzer bei ihnen dieselben Rechte geltend machen können und eine Liste aller registrierten Broker auf der Website der Generalstaatsanwaltschaft zu finden ist.⁷⁰

Das Mantra lautet hierbei oft, dass eine Balance zwischen Datenschutz und Wirtschaftlichkeit gefunden werden muss. Beides zu haben sei unrealistisch, nur eins davon nicht erstrebenswert. Gestärkt wird diese Haltung durch die Größe der Technologiebranche in den USA und ihrer Relevanz für den Wirtschaftsstandort.

2.6.2 EU

Anders sieht das Bild innerhalb der Europäischen Union aus, wo Datenschützer spätestens seit Herausgabe der DSGVO eine Rechtswidrigkeit in der Geschäftspraxis von Data Brokern verorten.

Konkret werden dabei oftmals die Unterpunkte b und c der Grundsätze für die Verarbeitung personenbezogener Daten zitiert. In ihnen heißt es, dass personenbezogene

⁶⁶ U.S. Senate Committee on Commerce, Science, and Transportation 2015

⁶⁷ Doss 2020, S. 277 - 283

⁶⁸ §§ 1798.105, 1798.110, 1798.120 California Consumer Privacy Act (CCPA), Verordnung (Kalifornien) 2018

⁶⁹ Artikel 5, Absatz 1 der Verordnung (EU) 679/2016

⁷⁰ State of California Department of Justice Office of the Attorney General 2024

Daten nur „für festgelegte, eindeutige und legitime Zwecke erhoben werden und [...] nicht in einer mit diesen Zwecken nicht zu vereinbarenden Weise weiterverarbeitet werden [...]“ dürfen sowie „dem Zweck angemessen und erheblich sowie auf das für die Zwecke der Verarbeitung notwendige Maß beschränkt [...]“⁷¹ sein müssen. Kritiker werfen den Brokern vor, dass eine massenhafte Datenaggregation, was Kernkompetenz eines jeden Data Brokers ist, nicht mit diesem Grundsatz in Einklang gebracht werden kann. Die Gegenseite bringt vor, dass die gesammelten Datenmengen für Werbemaßnahmen oder etwa zur finanziellen Beurteilung notwendig sowie legitim sind und die Verwendung in den Allgemeinen Geschäftsbedingungen der Partner eindeutig festgelegt ist.

Problematisch ist auch, dass Data Broker in der DSGVO mit keinem Wort erwähnt werden, was für beide Seiten Interpretationsspielräume offenlässt. Es gibt keine spezifischen Auflagen zum Handel, da der Geltungsbereich laut Artikel 2 sich schlichtweg auf jede Verarbeitung bezieht. Dies schließt jedoch auch den Handel ein, da unter Verarbeitung „das Erheben, das Erfassen, die Organisation, das Ordnen, die Speicherung, die Anpassung oder Veränderung, das Auslesen, das Abfragen, die Verwendung, die Offenlegung durch Übermittlung, Verbreitung oder eine andere Form der Bereitstellung, den Abgleich oder die Verknüpfung, die Einschränkung, das Löschen oder die Vernichtung“⁷² fällt.

Der erste und bis dato größte strafrechtliche Vorgang in der EU gegen einen Data Broker selbst wurde 2019 in Polen geführt. Der Broker sammelte unter anderem Adress- und Kontaktdaten von polnischen Einwohnern aus öffentlichen Quellen, informierte aber von dieser Sammlung nicht die Personen, von denen keine E-Mail-Adresse hinterlegt war.⁷³ Aus eigenen Angaben entschied man sich dagegen, weil eine postalische Informationsmitteilung zu kostspielig gewesen wäre. Für dieses Vorgehen wurde der Broker mit einer Strafe von umgerechnet ca. 220.000 € belegt. Brisanz erlangte der Fall jedoch auch, weil die Daten mittels Web Scraping erhoben wurden und ein Großteil der Daten aus öffentlichen, von staatlichen Organisationen geführten Registern stammte.⁷⁴

⁷¹ Artikel 5, Absatz 1 der Verordnung (EU) 679/2016

⁷² Artikel 4, Absatz 2 der Verordnung (EU) 679/2016

⁷³ Urząd Ochrony Danych Osobowych (UODO) 2023

⁷⁴ Doss 2020, S. 271

3. Mehrwert und Risiken

3.1 Vorteile für Unternehmen

Im Hinblick auf Marketing ist der größte Vorteil für Wirtschaftstreibende wohl ohne Zweifel, dass potenzielle Kunden mit hoher Kaufwahrscheinlichkeit direkt angesprochen werden können und sich dadurch das Kapital, welches in nicht gewinnbringende Werbung gesteckt wird, maßgeblich reduziert wird.

Wie jedoch in Kapitel 2.3 angesprochen, gehen die Dienstleistungsangebote von Data Brokern und damit auch die Vorteile für Wirtschaftstreibende weit über die Marketingdimensionen hinaus.

Mithilfe von ausführlichen Bonitätseinschätzungen zu virtuell jedem Kunden kann das unternehmerische Risiko drastisch gesenkt werden. Dienstleistungen wie die Identitätsverifikation ermöglichen erst die sichere Durchführung von kapitalreichen Geschäftsfällen ohne direkten Kundenkontakt. Und durch das ständige Kontrollieren der eigens hinterlegten Nutzerinformationen zur Vergleichsdatenbank des Data Brokers können fehlerhafte Einträge und Inkonsistenzen entdeckt sowie bereinigt werden.

Nicht zuletzt eröffnen sich durch den Zugriff auf aggregierte Marktdaten strategische Wettbewerbsvorteile. Unternehmen können Kaufverhalten, demografische Trends oder regionale Besonderheiten analysieren, um ihre Produkte und Dienstleistungen gezielt an die Bedürfnisse des Marktes anzupassen. Diese datengetriebene Entscheidungsfindung ermöglicht es, neue Märkte effizient zu erschließen, Preismodelle zu optimieren oder potenzielle Wachstumschancen frühzeitig zu identifizieren.

3.2 Vorteile für Konsumenten

Durch maßgeschneiderte Werbung wird dem Nutzer im besten Fall nur Werbung angezeigt, die für seinen Lebensstil relevant ist. Geht man in unserer kapitalistischen Wirtschaftswelt von der Prämisse aus, dass ein (gesunder) Konsum essenziell für das Bestehen und den Erfolg des Systems ist, so kann man dies auch als Vorteil für den Endkunden verbuchen.

Aber auch die angesprochenen anderen Data Broker Dienstleistungsangebote bringen nicht nur für den Unternehmer einen Mehrwert mit sich. Wie in 3.1 ausgeführt wird es erst durch Identitätsverifikation und Betrugsversuchserkennung möglich, dass Geschäfte auf sichere Art und Weise digital stattfinden können. Der digitalisierte Mensch, der im Allgemeinen die Freiheit und Flexibilität genießt, die ihm durch die Möglichkeit Geschäftsfälle über das Mobiltelefon abzuschließen gegeben wird, schätzt es, dass immer mehr Geschäftsfälle digital abgeschlossen werden können.

3.3 Nachteile für Konsumenten

Die von manchen Data Brokern prognostizierten und hinterlegten Informationen können tiefgreifend sein und Inhalte enthalten, die Menschen aus berechtigtem Interesse nicht vollständig publik machen wollen. Darunter fallen insbesondere Informationen zu politischen Ansichten, Religionsangehörigkeit oder sexueller Orientierung.⁷⁵

Des Weiteren können fehlerhafte Informationen oder Zuordnungen zu folgenschweren Entscheidungen seitens der Unternehmen gegenüber den Kunden führen. Vor allem bei kritischen Dienstleistungen wie dem Kreditscoring gibt es zu wenige gesetzliche Regulierungen, die eine qualitätsvolle Analyse gewährleisten. Durch die fehlende Standardisierung wird durch Gewichtung einzelner (teils nicht nachvollziehbarer Variablen) eine Subjektivität in die als objektiv angepriesenen Modelle gebracht. Hinzu kommt, dass es keine Auflagen bezüglich Aktualität der herangezogenen Daten gibt und die Transparenz gegenüber dem Kunden fehlt.⁷⁶

Auch die Möglichkeit, dass Privatpersonen solche Datensets kaufen, um an Informationen zu Personen in ihrem näheren Umfeld zu gelangen, sorgt für Beunruhigung. Zahlreiche Datenschützer warnen vor einer neuen Ära des digitalen Stalkings.⁷⁷ Besonders der Fakt, dass der in dieser Arbeit bereits erwähnte Data Broker *MEDbase 200* de-anonymisierte Listen zu HIV-infizierten Personen oder Vergewaltigungsopfern zum Verkauf anbot, ist in diesem Kontext weit mehr als schlichtweg bedenklich.⁷⁸ Die Anonymisierung von veräußerten Datensets stellt nach neuesten Forschungsergebnissen jedoch auch keine befriedigende Lösung dar. So konnte in einem amerikanischen Versuch bestätigt werden, dass die Re-Identifikation von anonymisierten Individuen durch lediglich 15 demografische Attribute mit einer Erfolgswahrscheinlichkeit von 99,98 % möglich ist.⁷⁹

⁷⁵ Beckmann 2024

⁷⁶ Rothmann, Sterbik-Lamina und Peissl 2014, S. 57

⁷⁷ Rostow 2017, S. 667

⁷⁸ Hill 2013

⁷⁹ Rocher, Hendrickx und Montjoye 2019, S. 1

Resümee

Daten sind ein Asset, und der Branche, die sie handelt, wird ein dreistelliger Milliardenwert zugeschrieben. Die positive Entwicklung mit Wachstumschancen um die 7 % pro Jahr scheint gesichert zu sein, solange der scheinbar unaufhaltsame Trend weg vom analogen zum digitalen Marktplatz anhält.

Mit dieser Transformation werden die Dienstleistungsangebote der Data Broker für Wirtschaftstreibende stetig relevanter. Nicht nur zu Zwecken der zielsicheren Werbung, sondern auch für im analogen Leben selbstverständliche Zwecke, wie Identitätsverifikation oder Geschäftstüchtigkeitseinschätzungen.

Gleichzeitig dürfen legale wie ethische Schwierigkeiten nicht außer Acht gelassen werden. In den USA scheint die Technologieoffenheit auch in der Gesetzgebung weiter Bestand zu halten. Anders sieht die Lage innerhalb der Europäischen Union aus, wo die Geschäftspraktiken vieler Data Broker bereits in die Grauzonen des geltenden Rechts (DSGVO) fallen und Datenschutzinitiativen immer mehr Fahrt aufnehmen.

Wenn die Datensammlung außer Kontrolle gerät und weder die Akteure noch ihre Zwecke mehr durchschaubar sind, kann die Transformation zum gläsernen Menschen mit all ihren zivilen wie politischen Folgen nicht mehr gestoppt werden.

Deshalb ist es unabdingbar, einen offenen Dialog über die Funktionsweise, Größe und Ziele dieses gesellschaftlich wenig bekannten Marktes zu fördern. Nur wer informiert ist, kann fundierte Entscheidungen treffen. Dies gilt gleichsam für die Unternehmen, die auf die Dienstleistungen von Data Brokern angewiesen sind, als auch für die Menschen, deren Daten gehandelt werden.

Abbildungsverzeichnis

Abb. 1: Begriffshierarchie. Bodendorf 2006, S. 1	4
Abb. 2: Verfallsfunktionen. Eigene Grafik.....	18
Abb. 3: Schematische Einteilung der Datenerhebungsmethoden. Eigene Grafik.....	24
Abb. 4: Wirtschaftliche Kennzahlen der Data Broker Branche. Maximize Market Research 2025, abgerufen von https://www.maximizemarketresearch.com/wp-content/uploads/2020/04/Data-Broker-Market-1.webp	36

Tabellenverzeichnis

Tab. 1: Gängige Datentypen. Eigene Tabelle.....	8
Tab. 2: Beispiel für ein Data Broker Nutzerprofil. Eigene Tabelle mit Informationen entnommen aus Epsilon Data Management LLC 2025	35

Literaturverzeichnis

Aggarwal, Charu. Data Mining. The Textbook. Heidelberg: Springer, 2015.

Anteby, Michel, und Rakesh Khurana. Harvard Business School. o.J.

<https://www.library.hbs.edu/hc/hawthorne/09.html#nine> Zugriff am 04. Februar 2025.

Beckmann, Werner. nordvpn s.a. 09. Februar 2024.

<https://nordvpn.com/de/blog/was-ist-datenbroker/> Zugriff am 13. Februar 2025.

Bodendorf, Freimut. Daten- und Wissensmanagement. (2. Aufl.). Berlin: Springer, 2006.

Cahn, Aaron, Scott Alfeld, Paul Barford, und Shan Muthukrishnan. „An Empirical Study of Web Cookies.“ WWW '16: Proceedings of the 25th International Conference on World Wide Web. Montréal: ACM, 2016. 891 - 901.

dataspot. GmbH. dataspot.at. kein Datum.

<https://www.dataspot.at/glossary/datenarten/> Zugriff am 13. Januar 2025.

Doss, April Falcon. Cyber Privacy. Who Has Your Data and Why You Should Care. Dallas, 2020.

Duden (Hrsg.). duden.de. o.J. https://www.duden.de/rechtschreibung/meta_ Zugriff am 16. Dezember 2024.

Epsilon Data Management LLC. Epsilon. 2025. <https://www.epsilon.com/us/products-and-services/data>, <https://www.epsilon.com/abacus> Zugriff am 13. Februar 2025.

Fässler, Lukas, Barbara Scheuner, und David Sichau. „ETH Zürich.“ 29. November 2024. https://doc.et.ethz.ch/latest/p_java1_TH.pdf Zugriff am 17. Dezember 2024.

Gerber, Nina, Paul Gerber, und Melanie Volkamer. „Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior.“ Computers & Security, August 2018: 226-261.

Goodwin, Michael. IBM. 09. April 2024. <https://www.ibm.com/think/topics/api> Zugriff am 03. Februar 2025.

Grand View Research Inc. Dezember 2024.

<https://www.grandviewresearch.com/industry-analysis/global-telecom-services-market> Zugriff am 13. Februar 2025.

Hegewald, Tony. „Im Internet weiß niemand, dass du ein Hund bist“ –

Personalisierung von Onlinewerbung in Deutschland. Mittweida: Hochschule Mittweida, 2017.

Heldt, Cordula. Gabler Wirtschaftslexikon. 19. Februar 2018.

<https://wirtschaftslexikon.gabler.de/definition/broker-27861/version-251503>
Zugriff am 13. Februar 2025.

Hill, Kashmir. 19. Dezember 2013.

<https://www.forbes.com/sites/kashmirhill/2013/12/19/data-broker-was-selling-lists-of-rape-alcoholism-and-erectile-dysfunction-sufferers/> Zugriff am 13. Februar 2025.

Hinz, Michael, und Gregor Weiß. Gabler Banklexikon. 08. April 2020.

<https://www.gabler-banklexikon.de/definition/asset-55743/version-377253>
Zugriff am 13. Februar 2025.

Höchstötter, Nadine. „Methoden der Erhebung von Nutzerdaten und ihre Anwendung in der Suchmaschinenforschung.“ In Handbuch Internet-Suchmaschinen, von Dirk Lewandowski, 175 - 203. Heidelberg: AKA Verlag, 2009.

Isaak, Jim, und Mina Hanna. „User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection.“ Computer, 14. August 2018: 56 - 59.

Jenkins, Richard. The Drum. 04. April 2012.

<https://www.thedrum.com/opinion/2012/04/04/how-much-your-email-address-worth> Zugriff am 13. Februar 2025.

Kähler, Wolf-Michael. SQL. Bearbeitung relationaler Datenbanken. Wiesbaden: Vieweg+Teubner, 1990.

Kluge, Friedrich, und Elmar Seebold. Etymologisches Wörterbuch der deutschen Sprache (23. Auflage). Berlin: de Gruyter, 1999.

Kranz, Garry. ComputerWeekly.de. September 2024.

<https://www.computerweekly.com/de/definition/Metadaten> Zugriff am 16. Dezember 2024.

- Krotov, Vlad, Leigh Johnson, und Leiser Silva. „Tutorial: Legality and Ethics of Web Scraping.“ Communications of the Association for Information Systems, 15. Dezember 2020: 555 - 581.
- Lackes, Richard, und Markus Siepermann. Gabler Wirtschaftslexikon. 19. Februar 2018. <https://wirtschaftslexikon.gabler.de/definition/datentyp-29319/version-252929> Zugriff am 17. Dezember 2024.
- Lahmer, Karl. Kernbereiche Psychologie & Philosophie. Wien: westermann, 2023.
- Lange, Otto, und Gerhard Stegemann. Datenstrukturen und Speicherarten. Braunschweig: Vieweg+Teubner, 1985.
- Langer, Erasmus. Programmieren in Fortran. Wien: Springer, 1993.
- Liebhart, Daniel. „Die glorreichen sieben Datenarten.“ netzwoche, 2010: 21.
- Lošek, Firtz (Hrsg.). Stowasser. Lateinisch-deutsches Schulwörterbuch. (1. Aufl. 4. Druck). München: Oldenbourg Schulbuchverlag, 2016.
- Maximize Market Research PVT. LTD. Januar 2025.
<https://www.maximizemarketresearch.com/market-report/global-data-broker-market/55670/> Zugriff am 13. Februar 2025.
- Microsoft (Hrsg.). Microsoft Corporation. 03. August 2023.
<https://learn.microsoft.com/de-de/office/vba/language/reference/user-interface-help/data-type-summary> Zugriff am 17. Dezember 2024.
- Moryl, Beata. Piwik PRO. 10. Oktober 2024. <https://piwikpro.de/blog/was-ist-der-unterschied-zwischen-first-party-cookies-und-third-party-cookies/> Zugriff am 04. Februar 2025.
- Murray-Rus, Peter. „nature.com.“ 18. Januar 2008.
<https://www.nature.com/articles/npre.2008.1526.1> Zugriff am 03. Februar 2025.
- Neally, Daniel. „Data Brokers and Privacy: An Analysis of the Industry and How It's Regulated.“ Adelphia Law Journal, 2019: 30 - 46.
- Pabst, Christiane M., Herbert Fussy, und Ulrike Steiner (Red.). Österreichisches Wörterbuch (43. Aufl.). Wien: Österreichischer Bundesverlag Schulbuch GmbH, 2016.

Research and Markets. Dezember 2024.

<https://www.researchandmarkets.com/reports/5987173/data-broker-market-forecasts> Zugriff am 13. Februar 2025.

Rocher, Luc, Julien Hendrickx, und Yves-Alexandre Montjoye. „Estimating the success of re-identifications in incomplete datasets using generative models.“ *nature Communications*, 2019.

Rostow, Theodore. „What Happens When an Acquaintance Buys Your Data?: A New Privacy Harm in the Age of Data Brokers.“ *Yale Journal on Regulation*, 2017: 667 - 707.

Rothmann, Robert, Jaro Sterbik-Lamina, und Walter Peissl. *Credit Scoring in Österreich. Studie*, Wien: AK Wien, 2014.

Ruscheimer, Hannah. „Data Brokers and European Digital Legislation.“ *European Data Protection Law Review (EDPL)*, 2023: 27 - 38.

Schlageter, Gunter, und Wolffried Stucky. *Datenbanksysteme: Konzepte und Modelle*. Wiesbaden: Vieweg+Teubner, 1983.

Seiter, Mischa. *Business Analytics. Wie Sie Daten für die Steuerung von Unternehmen nutzen*. (3., überarbeitete Auflage). München: Franz Vahlen, 2023.

SkyQuest Technology Group. *SKYQUEST*. Januar 2025.

<https://www.skyquestt.com/report/smartphones-market> Zugriff am 13. Februar 2025.

SRH Fernhochschule GmbH. *SRH Fernhochschule - The Mobile University*. o. J.

<https://www.mobile-university.at/studium/return-on-investment/> Zugriff am 13. Februar 2025.

State of California Department of Justice Office of the Attorney General. 13. März 2024. <https://oag.ca.gov/privacy/ccpa> Zugriff am 15. Februar 2025.

U.S. Senate Committee on Commerce, Science, and Transportation. *What Information Do Data Brokers Have on Consumers, and How Do They Use It?* Senatsanhörung, Washington D.C.: U.S. Government Publishing Office, 2015.

Urząd Ochrony Danych Osobowych (UODO). 20. September 2023.

<https://uodo.gov.pl/en/553/1572> Zugriff am 15. Februar 2025.

Venkatadri, Giridhari, Piotr Sapiezynski, Elissa Redmiles, Alan Mislove, Oana Goga, und Michelle, Gummadi, Krishna Mazurek. „Auditing Offline Data Brokers via Facebook's Advertising Platform.“ WWW '19. San Francisco: Association for Computing Machinery, 2019. 1920 - 1930.

Wittes, Benjamin, und Jodie Liu. The privacy paradox: The privacy benefits of privacy threats. Washington: The Brookings Institution, 2015.

Wuttke, Lorenz. Datasolut. 27. Juni 2024. <https://datasolut.com/wiki/daten-definition/>
Zugriff am 16. Dezember 2024.

Eidesstattliche Erklärung

Ich, Simon Michael Wimmer, erkläre, dass ich diese abschließende Arbeit eigenständig angefertigt und nur die in den Verzeichnissen angeführten Quellen, Materialien und Hilfsmittel benutzt habe.

(Braunau am Inn, 16.02.2025)