

Universität Siegen, SoSe 2019

Fakultät I: Philosophische Fakultät

Lehrstuhl für Betriebswirtschaftslehre, insb. Medienmanagement

Strategien zur Fehlerprävention und zum Fehlermanagement bei Datenanalysen im Bereich des E-Commerce.

„Masterarbeit zur Erlangung des Grades Master of Arts an der Philosophischen Fakultät Siegen“

Betreuerin: Julia Müller

Erstprüfer: Univ.-Prof. Dr. Joachim Eigler

Zweitprüfer: Univ.-Prof. Dr. Volkmar Pipek

Abgabetermin: 05.09.2019

Svenja Wacker

M.A. Medien und Gesellschaft

(Medienmanagement/Unternehmensführung und Medienwissenschaft)

Inhaltsverzeichnis

Abkürzungsverzeichnis	III
Abbildungsverzeichnis	IV
Tabellenverzeichnis.....	V
1. Problemstellung	1
2. Theoretisch konzeptionelle Grundlagen zum Fehlermanagement und zur Fehlerprävention bei Datenanalysen im Bereich des E-Commerce	3
2.1. Zum Begriff und zur Einordnung des Fehlers und des Fehlermanagement	3
2.1.1. Definition des Fehlerbegriffs	3
2.1.1.1. Definition eines menschlichen Fehlers	3
2.1.1.2. Fehler im Rahmen des Datenmanagements	5
2.1.2. Theoretische Erklärungsansätze von Fehlerklassifikationen	8
2.1.2.1. Theoretischer Erklärungsansatz nach Auftretenshäufigkeit	8
2.1.2.2. Theoretischer Erklärungsansatz bezogen auf Arbeitssysteme	11
2.1.2.3. Theoretischer Erklärungsansatz bezogen auf Big Data Risiken	13
2.1.3. Zum Fehlermanagement und dessen organisatorischer Verortung in Bezug auf Datenanalysen im Bereich des E-Commerce.....	16
2.2. Begriff, Einordnung und Ziele von Big Data im Zusammenhang mit Datenanalysen im Bereich des E-Commerce	20
2.2.1. Begriffssystematisierung von Informationen und Daten	20
2.2.2. Zur Begriffsdefinition und Einordnung von Big Data	22
2.2.2.1. Begriffsdefinition von Big Data.....	22
2.2.2.2. Einordnung von Big Data im aktuellen Kontext.....	24
2.2.3. Charakteristika von Big Data	27
2.2.4. Ziele des strategischen Einsatzes von Big Data im Wertschöpfungsprozess	30

2.2.4.1.	Strategisches Management in Bezug auf Datenanalysen im Bereich des E-Commerce	30
2.2.4.1.1.	Big Data aus Sicht des Market based View.....	30
2.2.4.1.2.	Big Data aus Sicht des Resourced based View	33
2.2.4.2.	Risikomanagement in Bezug auf Datenanalysen im Bereich des E-Commerce.....	36
2.2.5.	Möglichkeiten und Notwendigkeit eines gezielten Datenmanagements im E-Commerce	38
2.3.	Grundlagen und Kennzahlen von Geschäftsmodellen im E-Commerce	41
2.3.1.	Grundlagen und Kennzahlen von Geschäftsmodellen des E-Commerce.	41
2.3.2.	Grundlagen und Kennzahlen datenzentrischer Geschäftsmodelle	44
2.4.	Grundlagen, Kennzahlen und Einordnung des E-Commerce in Bezug auf Datenanalysen.....	47
3.	Analyse von Strategien zur Fehlerprävention und zum Fehlermanagement bezogen auf Datenanalysen im Bereich des E-Commerce	50
3.1.	Einsatzbereiche von Big Data entlang des elektronischen Wertschöpfungsprozesses.....	50
3.1.1.	Informationssammlung als erste Stufe des Wertschöpfungsprozesses	50
3.1.2.	Informationsverarbeitung als zweite Stufe des Wertschöpfungsprozesses	53
3.1.3.	Informationsübertragung als dritte Stufe des Wertschöpfungsprozesses..	55
3.2.	Stufen des elektronischen Wertschöpfungsprozesses mit Blick auf mögliche Fehlerquellen von Datenanalysen im Bereich des E-Commerce	58
3.2.1.	Fehlerquellen in der Informationssammlung	58
3.2.1.1.	Fehlerquellen in der Prä-Informationssammlung	58
3.2.1.2.	Fehlerquellen in der Informationssammlung	61
3.2.1.3.	Fehlerquellen in der Post-Informationssammlung.....	63
3.2.2.	Fehlerquellen in der Informationsverarbeitung.....	65

3.2.2.1.	Fehlerquellen in der Prä-Informationsverarbeitung	65
3.2.2.2.	Fehlerquellen in der Informationsverarbeitung	67
3.2.2.3.	Fehlerquellen in der Post-Informationsverarbeitung	69
3.2.3.	Fehlerquellen in der Informationsübertragung.....	71
3.2.3.1.	Fehlerquellen in der Prä-Informationsübertragung.....	71
3.2.3.2.	Fehlerquellen in der Informationsübertragung	73
3.2.3.3.	Fehlerquellen in der Post-Informationsübertragung	74
3.2.4.	Zusammenfassung der analytischen Erkenntnisse zu den Fehlerquellen im elektronischen Wertschöpfungsprozess	76
3.3.	Maßnahmen einer Fehlerprävention	79
3.3.1.	Maßnahmen während der Informationssammlung.....	79
3.3.2.	Maßnahmen während der Informationsverarbeitung	81
3.3.3.	Maßnahmen während der Informationsübertragung	84
3.4.	Kommunikation in betrieblichen Krisensituationen als Teil des Fehlermanagements	86
4.	Fazit und Ausblick.....	88
	Literaturverzeichnis.....	91
	Anhang	98

Abkürzungsverzeichnis

24/7 24 Stunden an sieben Tagen die Woche

AI Artificial Intelligence

AMQP Advanced Message Queuing Protocol

BITKOM Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V.

bzw. beziehungsweise

DL Deep Learning

DS-GVO Datenschutz-Grundverordnung

HDFS Hadoop Distributed File System

IT Informationstechnik

JMS Java Message Service

ML Machine Learning

OASIS Organization for the Advancement of Structured Information Standards

TCP/IP Transmission Control Protocol/Internet Protocol

z.B. zum Beispiel

Abbildungsverzeichnis

Abbildung 1:	Fehlerklassifikation nach Auftretenshäufigkeit in Anlehnung an Chapanis (1951) (Hofinger 2012, S. 49)	9
Abbildung 2:	Risikobereiche von Big Data (BITKOM 2014, S. 136)	15
Abbildung 3:	Linke Abbildung: Informationspyramide nach Voß und Gutenschwager (2001, S. 13), Rechte Abbildung: Begriffshierarchie nach Krcmar (2015, S. 4)	20
Abbildung 4:	Basisgeschäftsmodelle des 4C-Net-Business-Model in der Internetökonomie (Wirtz und Becker 2002)	42

Tabellenverzeichnis

Tabelle 1:	Zusammenfassung der analytischen Ergebnisse von Fehlerquellen im elektronischen Wertschöpfungsprozess in Anlehnung an Kollmann (2019)	77-78
------------	---	-------

1. Problemstellung

Der Begriff der Digitalisierung ist in unserer heutigen Gesellschaft allgegenwärtig. Fast täglich werden neue Artikel in Online- und Printmedien zur Bedeutung der Digitalisierung in Deutschland veröffentlicht (Becker et al. 2019, S. 16). Für diese Masterarbeit sind im Zuge der Digitalisierung die damit einhergehenden unternehmerischen Möglichkeiten mit Blick auf Big Data Maßnahmen von Interesse. Big Data Analysen stehen, in Auswertungen der Boston Consulting Group, an dritter Stelle bezüglich der Bereiche, die in den nächsten drei bis fünf Jahren den meisten Einfluss auf die Industrie haben werden und die aktiv bearbeitet werden sollten (Ringel et al. 2018, S. 4). Die gewinnbringende Integration von Big Data Analysen für eine zukunftsfähige Positionierung im Wettbewerb ist dabei nicht ohne Herausforderungen möglich (Verhoef et al. 2016, S. 3). Unternehmen haben im Kontext von Big Data hohe Erwartungen an die neuen Möglichkeiten, die Big Data Analysen in der Theorie schaffen können (Verhoef et al. 2016, S. 3). Die relevanten Informationen aus den gesammelten Datenmengen herauszufiltern, stellt für viele allerdings ein Problem dar (Verhoef et al. 2016, S. 3). Marr (2015, S. 1) weist darauf hin, dass die Hälfte aller Projekte gerade durch mangelnde Kenntnisse und Fähigkeiten, die auch Verhoef et al. (2016, S.3) ansprechen, nicht nach den Zielvorstellungen der Unternehmen verlaufen und dadurch ein hohes Fehlerpotential bergen. Zunächst müssen Unternehmensziele genau definiert und fähiges Personal vorhanden sein, damit Big Data Analysen dem Unternehmen einen Mehrwert bieten und Unternehmensziele erreicht werden können (Marr 2015, S. 1). Für eine erfolgreiche Anwendung von Big Data erläutern auch McAfee und Brynjolfsson (2012, S. 66-67), dass ein Unternehmen klar definierte Ziele, Technologieplattformen und talentierte Mitarbeiter, die in der Lage sind, große Datenmengen aufzubereiten und zu visualisieren, zur Datenverarbeitung benötigt. Außerdem sollten die gewonnenen Informationen nach McAfee und Brynjolfsson (2012, S. 66-67) mit Unternehmensentscheidungen verknüpft und bei der Verarbeitung der Daten immer Rücksicht auf die durch das Unternehmen vertretenen Normen und Werte genommen werden.

Eine solche Integration enthält nicht nur theoretische Herausforderungen, sie kann auch in der Praxis zu Problemen führen. Das kann man an mehreren praktischen Beispielen

erkennen. Die US-Supermarktkette Target schickte beispielsweise im Jahr 2012 einer Minderjährigen per Post Werbung für Schwangere, da per Datenanalyse von Target ihre Schwangerschaft festgestellt werden konnte, noch bevor ihr eigener Vater davon wusste (Beuth 2014, S. 1). Auch beim Internetriesen Amazon schleichen sich im Kontext von Big Data und der Masse an zu verarbeitenden Daten fehlerhafte Handlungen ein: Ein Kunde verlangte gemäß der neuen DS-GVO vom Mai 2018 (Bundesministerium für Wirtschaft und Energie o.D.), dass er Auskunft über alle Daten erhält, die Amazon von ihm gesammelt hat (Fuest 2018, S. 1). Allerdings wurden diesem Kunden daraufhin nicht nur seine Daten geschickt, sondern auch 1700 Alexa Sprachaufnahmen eines anderen Kunden (Fuest 2018, S. 1). Der E-Commerce Händler Zalando hat ebenfalls Schwierigkeiten bei Big Data Maßnahmen (Jansen 2018, S. 1). Die Datenverarbeitung wird von Zalando zum Beispiel genutzt, um durch das sogenannte Targeting eine effektivere Zielgruppenansprache durch personalisierte Werbung zu ermöglichen (Jansen 2018, S. 1). Fehlerhaft ist allerdings, wenn dem Kunden dann immer noch Werbung zu einem Produkt angezeigt wird, welches dieser kurz zuvor schon gekauft hat (Jansen 2018, S. 1). Diese drei Beispiele zeigen bereits, dass Probleme bei einer Durchführung von Big Data Maßnahmen keine Seltenheit sind und auch den großen Playern in der Wirtschaft unterlaufen. Gerade deshalb ist es wichtig, in den Unternehmen ein Fehlerbewusstsein zu bilden oder zu vertiefen, um Fehlerquellen erkennen, diesen präventiv entgegenwirken zu können und bei geschehenen Fehlern ein solides Fehlermanagement als Rückhalt zu haben. Hier setzt diese Masterarbeit an und bietet ihren Mehrwert durch ihre Analyse und das damit zusammenhängende Aufzeigen möglicher Fehlerquellen, die von Unternehmen bei der Verwendung von Datenanalysen Beachtung finden sollten.

In dieser Masterarbeit soll zunächst insbesondere der Begriff des Fehlers und im Anschluss der Begriff des Fehlermanagements betrachtet werden. In dem Zusammenhang wird dargestellt, was einen Fehler mit Blick auf die Thematik auszeichnet. Wie die genannten Beispiele zeigen, können auch richtige Ergebnisse von Datenanalysen, wie bei dem Target-Beispiel, als negativ und somit als fehlerhaft aufgefasst werden. In einem weiteren Schritt erfolgt eine genaue Auseinandersetzung mit dem Schlagwort Big Data und den Zielen und der Notwendigkeit eines gezielten Datenmanagements. Nachdem anschließend der Blick auf Charakteristika und Geschäftsmodelle des E-Commerce ge-

richtet wurde, folgt der analytische Teil der Arbeit. Hierbei wird der elektronische Wertschöpfungsprozess nach Kollmann (2019, S. 61) in den Mittelpunkt gestellt, um in den drei Prozessstufen, Informationssammlung, -verarbeitung und -übertragung, den Einfluss von Big Data herauszustellen. Durch diesen Schritt können in den jeweiligen Stufen der Wertschöpfung die möglichen Fehlerquellen aufgezeigt und genauer untersucht werden. Für jede Prozessstufe wird zur ausführlicheren Betrachtung ein vorheriger und ein nachfolgender zusätzlicher Prozessschritt hinzugefügt. Anschließend wird auf präventive Maßnahmen für die Vermeidung der herausgestellten Fehlerquellen vor dem Hintergrund des Wertschöpfungsprozesses eingegangen, bevor die Kommunikation im Falle einer betrieblichen Krisensituation als Teil des Fehlermanagements dargestellt wird. Den Abschluss der Masterarbeit bilden das Fazit und die Evaluation der Ergebnisse.

2. Theoretisch konzeptionelle Grundlagen zum Fehlermanagement und zur Fehlerprävention bei Datenanalysen im Bereich des E-Commerce

2.1. Zum Begriff und zur Einordnung des Fehlers und des Fehlermanagement

2.1.1. Definition des Fehlerbegriffs

2.1.1.1. Definition eines menschlichen Fehlers

Fehler können die Ursache für Unfälle mit Folgen für Mensch, Umwelt oder Kapital sein, weshalb in unserer Gesellschaft nach einer Fehlerlosigkeit gestrebt wird (Hofinger 2012, S. 40). Es gibt in der Wissenschaft laut Hofinger (2012, S. 40) viele Definitionen zu Fehlern (van Dyck et al. 2005; Reason 1990; Rasmussen 1982), aber bis heute keine eindeutige, da diese aus vielen unterschiedlichen Forschungsrichtungen kommen und ihnen verschiedene Sichtweisen zugrunde liegen. Hofinger (2012, S. 40) hat sich für ihre Ausführungen den gemeinsamen Konsens aller Definitionen angeschaut und zusammengefasst: „Fehler sind eine Abweichung von einem als richtig angesehenen Verhalten oder von einem gewünschten Handlungsziel, das der Handelnde eigentlich

hätte ausführen bzw. erreichen können.“ Fehler schließen laut Hofinger (2012, S. 40) hierbei menschliches Handeln mit ein, da eine Maschine laut ihr keine Fehler machen, sondern nur falsch verwendet, programmiert oder defekt sein kann. Rasmussen (1982, S. 3) spricht in seinen Ausführungen beispielsweise davon, dass menschliche Fehler als „instances of man-machine or man-task misfits“ zu betrachten sind. Zudem passieren Fehler laut Hofinger (2012, S. 40) dann, wenn eine Absicht vorlag, die nicht wie geplant durchgeführt werden konnte. Das Target-Beispiel aus Kapitel eins hatte den Zweck, die Kundin beim Einkauf zu unterstützen. Dieses Ziel konnte allerdings nicht erreicht werden, da Target die Situation für das minderjährige schwangere Mädchen eher unangenehmer machte und sie dadurch verschlimmerte. Einen Fehler als solchen zu klassifizieren setzt demnach laut Hofinger (2012, S. 40) zunächst die Bewertung einer Handlung beziehungsweise eines Handlungsergebnisses als fehlerhaft voraus. Außerdem führt sie in Anlehnung an Weimer (1925) eine Abgrenzung eines Fehlers von einem Irrtum an, da für einen Irrtum das Wissen für das Erreichen des Ziels einer Handlungssequenz bereits bei Beginn fehlte (Hofinger 2012, S. 41). Um menschliche Fehler erklären zu können, erläutert Hofinger (2012, S. 41), dass verstanden werden muss, was bezogen auf das alltägliche menschliche Handeln und Denken als „Normal“ angesehen wird und was demnach als eine Abweichung von diesem gilt. Zum Teil können fehlerhafte Handlungen laut Frese und Keith (2015, S. 678) aber auch Anstöße für Lern- und Innovationsprozesse schaffen. Ist einem Mitarbeiter ein Fehler unterlaufen, wird er versuchen, diesen nicht zu wiederholen (Frese und Keith 2015, S. 677). In Anlehnung an Klahr und Simon (1999) erklären sie zudem, dass fehlerhafte Wege manchmal auch zu neuen Erkenntnissen führen und dadurch Innovation vorantreiben können (Frese und Keith 2015, S. 678). Allerdings muss an dieser Stelle festgehalten werden, dass diese Aussagen vor dem Hintergrund getätigt wurden, dass Menschen nicht fehlerfrei sind. Damit wird versucht, die positiven Aspekte von Fehlern hervorzuheben. Nichtsdestotrotz ist es allein aufgrund des Zeit- und Kostenfaktors erstrebenswert, dass in einem Unternehmen so wenig Fehler wie möglich gemacht werden.

Reason (1990, S. 9) legt seinen Ausführungen über Fehler mehrere Arbeitsdefinitionen zugrunde. Im Englischen kann, anders als in der deutschen Sprache, zwischen drei Begriffen für einen Fehler unterschieden werden: Error, Slip/Lapse und Mistake (Reason

1990, S. 9). Ein Error ist für ihn aus psychologischer Sicht eine Situation, in der eine geplante mentale oder physische Handlung nicht das Handlungsziel erreicht und das Nichterreichen auch nicht dem Zufall zugeschrieben werden kann (Reason 1990, S. 9). Slips und Lapses definiert er als Fehler, die während der Ausführung oder Speicherung einer Handlungssequenz auftreten und unabhängig davon sind, ob die Vorgehensweise sinnvoll war (Reason 1990, S. 9). Während ein Slip meist auch von Dritten wahrgenommen wird, wie zum Beispiel ein Versprecher, ist ein Lapse durch Erinnerungslücken zum Teil auch nur wahrnehmbar für den Ausführenden (Reason 1990, S. 9). Als Mistake bezeichnet Reason (1990, S. 9) Fehler, die in der Zielsetzung oder in der Festlegung von Hilfsmitteln zum Erreichen eines Ziels beispielsweise durch schlechtes Urteilsvermögen gemacht werden. Dieser Fehler ist unabhängig davon, wie die Handlungssequenz verläuft und birgt eine größere Gefahr als Slips oder Lapses, da diese zum Teil lange Zeit unbemerkt bleiben und schwerer herauszufiltern sind (Reason 1990, S. 9).

Für diese Masterarbeit ergibt sich demnach folgende Ansicht eines menschlichen Fehlers: In der Literaturrecherche hat sich gezeigt, dass gerade in den früheren Quellen die Auffassung herrscht, jegliche Fehler müssten vermieden werden, da fehlerhafte Handlungen ein Unternehmen Zeit und Geld kosten. Allerdings zeigt die Quelle von Frese und Keith (2015), dass menschliche fehlerhafte Handlungen bis zu einem gewissen Maß als unvermeidbar angesehen werden. Dadurch schlagen sie vor, Fehler nicht mehr nur als etwas Negatives aufzufassen, sondern auch positive Fehlerkonsequenzen hervorzuheben (Frese und Keith 2015, S. 678). Fehler menschlichen Ursprungs sind hierbei Handlungen, die durch eine Bewertung als fehlerhaft klassifiziert werden und deren Auftreten zukünftig verhindert werden soll. Trotzdem gilt es, die Fehlerrate bei jeglichen Geschäftsaktivitäten unter anderem durch Fehlerpräventionsmaßnahmen auf ein Minimum zu beschränken. Der Umgang mit beziehungsweise die Bedeutung der Prävention von Fehlern, wird in Kapitel 2.1.3 genauer erläutert.

2.1.1.2. Fehler im Rahmen des Datenmanagements

Da sich diese Masterarbeit mit Datenanalysen im Bereich des E-Commerce beschäftigt, ist eine zusätzliche Betrachtung von Fehlern im Zusammenhang mit Big Data

Technologien zur Datenanalyse notwendig. Grottke und Trivedi (2007, S. 107) bezeichnen ein komplettes Einstürzen eines Systems als „system failure“, welches durch einen „software fault, or bug“ ausgelöst wird (Grottke und Trivedi 2007, S. 107). Können die Bedingungen, die zum Fehler führten, während eines Softwaretests nachgestellt werden, kann dieser identifiziert werden. Solche Fehler werden von Grottke und Trivedi (2007, S. 107) in der Praxis als Bohrbugs bezeichnet. Allerdings gibt es auch Fehler, die nicht reproduzierbar sind und die chaotisches und nichtdeterministisches Verhalten hervorbringen und daher hoch komplex und unvorhersehbar sind. Diese Fehler heißen laut Grottke und Trivedi (2007, S. 107-108) Mandelbugs. Hierbei ist das Problem, dass ein Mandelbug während eines Softwaretests vermutlich nicht vom Programmierer erkannt wird und daher lange Zeit unbemerkt bleiben kann. Diese Bugs führen nur dann zu einem Fehler, wenn sie auf die komplizierten systeminternen Programmumgebungen stoßen (Grottke und Trivedi 2007, S. 108). Da eine Situation, die einen Mandelbug auslöst, nicht reproduziert werden kann, wird es nahezu unmöglich, diesen Bug vor dem Veröffentlichen der Software zu entfernen (Grottke und Trivedi 2007, S. 108). Grottke und Trivedi (2007, S. 107) erläutern, dass Ingenieure nichtsdestotrotz viel Zeit durch dynamische Softwaretests und statischen Techniken, wie einem Software-Walkthrough, für ein frühes Erkennen von Bugs aufwenden. Das Ziel hiervon ist es, Bugs während der Softwareentwicklung bereits gefunden und eliminiert zu haben und dadurch Kosten zu sparen (Grottke und Trivedi 2007, S. 107). Wie am Beispiel der Bugs gesehen werden kann, ist es zum Teil sehr aufwendig einen Fehler aufzuspüren, zu definieren und dessen Ursachen zu beseitigen. Für Unternehmen bedeutet das – egal, ob es um einen menschlichen Fehler oder einen Bug geht – zeitlichen und finanziellen Aufwand sowohl bei der Nacharbeit von aufgetretenen Fehlern als auch bezogen auf das frühzeitige Erkennen von Bugs.

Neben „statischer Software“ werden im Rahmen dieser Masterarbeit auch die immer populärer werdenden selbstlernenden Systeme betrachtet. McAfee und Brynjolfsson (2017, S. 4) machen deutlich, dass artificial intelligence (AI), und in diesem Zusammenhang vor allem das machine learning (ML), durch den aktuellen Stand der Technik an Bedeutung gewinnt. ML impliziert, dass die Technologie in der Lage ist, ihre Performance ohne menschliche Erläuterungen selbst stetig zu verbessern (McAfee

und Brynjolfsson 2017, S. 4). Braganza (2018, S. 1) erläutert, dass AI ein breites Spektrum von Fähigkeiten abbildet, die es einer Maschine ermöglicht, Schlüsse zu ziehen, Lösungen vorzuschlagen, Maßnahmen zu ergreifen oder Ergebnisse zu erzielen. Der Programmablauf ist daher „based on rules and outcomes rather than pre-specified commands and lines of computer code that established the actions to be taken at every step.“ (Braganza 2018, S. 1) Im Zusammenhang mit ML ist es daher wichtig zu verstehen, dass diese Systeme nicht für ein spezifisches Endergebnis programmiert werden, sondern anhand von Regelwerken lernen (McAfee und Brynjolfsson 2017, S. 6). Durch AI eröffnen sich Unternehmen vielfältige neue Möglichkeiten bei der Automatisierung von Unternehmensprozessen, bei der Datenanalyse oder der Kommunikation zwischen Kunden und Mitarbeitern beispielsweise in Form eines 24/7 Kundenservices (Davenport und Ronank 2018, S. 110-112). Die Fehlerrate von Spracherkennungssystemen, wie Siri, Alexa und Google Assistant, ist seit dem Sommer 2016 von 8,5 Prozent auf 4,9 Prozent gesunken (McAfee und Brynjolfsson 2017, S. 5). Nicht nur dieses Beispiel zeigt, dass beim ML die Fehlerrate durch ständige Verbesserungen der Technologie immer weiter sinkt. Bei der Bilderkennung sind die Systeme mittlerweile so trainiert, dass sie zum Beispiel die gewünschten Personen oder Tiere bei einer gefallenen Fehlerrate von 30 Prozent in 2010 auf 4 Prozent in 2016 erkennen (McAfee und Brynjolfsson 2017, S. 5). Außerdem weisen diese Technologien auch große Fortschritte hinsichtlich Kognition und Problemlösung auf und schlagen bei Online-Pokerspielen selbst menschliche Profis (McAfee und Brynjolfsson 2017, S. 5). McAfee und Brynjolfsson (2017, S. 6) machen deutlich, dass ML-Systeme nicht nur ältere Algorithmen bereits ersetzen, sondern auch Menschen, da sie die Aufgaben besser erledigen können. Obwohl die Systeme mit einer Fehlerrate von ungefähr fünf Prozent nicht fehlerfrei sind, erläutern McAfee und Brynjolfsson (2017, S. 6): „Although the systems are far from perfect, their error rate – about 5% – on the ImageNet database is at or better than human-level performance.“ Durch die neuen Möglichkeiten des technologischen Fortschritts befindet sich die Wirtschaft im Umschwung: „In our view, artificial intelligence, especially machine learning, is the most important general-purpose technology of our era.“ (McAfee und Brynjolfsson 2017, S. 11) Wie McAfee und Brynjolfsson (2017, S. 11) aber auch deutlich machen, werden Manager später nicht durch AI ersetzt. Vielmehr ersetzen Manager, die AI nutzen, diejenigen, die die Verwendung ablehnen und diesen Umschwung nicht wahrnehmen (McAfee und

Brynjolfsson 2017, S. 11). McAfee und Brynjolfsson (2017, S. 4) zeigen aber auch, dass die großen Möglichkeiten von AI noch nicht durch Unternehmen genutzt werden und bisher viele unrealistische Erwartungen geweckt wurden.

Hofinger (2012, S. 40) zeigt in ihren Ausführungen, dass eine Maschine keine Fehler machen kann. Diese Aussage kann durch die Tatsache, dass auch Bohr- oder Mandelbugs nur durch ein menschliches Einwirken – in diesem Fall durch eine falsche Programmierung der Software – entstehen, unterstrichen werden (Grottke und Trivedi 2007,

S. 107-108). Hierbei muss allerdings angemerkt werden, dass Grottke und Trivedi (2007, S. 107-108) von statischer Software ausgehen. In diesem Zusammenhang stellt sich die Frage, ob bei dynamischer Software, beispielsweise ML, fehlerhafte Handlungen auftreten. Folgt man den Ausführungen von McAfee und Brynjolfsson (2017, S. 5), dann ist das der Fall. Gerade zu Beginn der Einführung solcher Technologien, zum Beispiel beim Start von Spracherkennungssystemen (McAfee und Brynjolfsson 2017, S. 5), ist die Fehlerrate relativ hoch. Mit der Zeit wird die Technologie aber immer weiter verbessert, sodass die Fehlerrate der Systeme stetig und mit enormer Geschwindigkeit sinkt (McAfee und Brynjolfsson 2017, S. 5). Auch Davenport und Ronank (2018, S. 116) erläutern, dass ein Einsatz von AI nicht ohne Schwierigkeiten und Herausforderungen ablaufen wird, aber dass „with the right planning and development, cognitive technology could usher in a golden age of productivity, work satisfaction, and prosperity.“

2.1.2. Theoretische Erklärungsansätze von Fehlerklassifikationen

2.1.2.1. Theoretischer Erklärungsansatz nach Auftretenshäufigkeit

Nachdem in Kapitel 2.1.1 definiert wurde, was unter einem Fehler verstanden wird, dienen die folgenden Fehlerklassifikationen zum Aufzeigen von Einordnungsmöglichkeiten für fehlerhafte Handlungen. Sowohl Reason (1990, S. 3-5) als auch Hofinger (2012, S. 48-49) greifen die Fehlerklassifikation nach Auftretenshäufigkeit von Chapanis (1951, S. 1181-1182) wieder auf. Sie besagt, dass Fehler dahingehend klassifiziert werden können, wie oft sie in einem System auftreten (Reason 1990). Dies ermöglicht eine bessere Vorhersage von Fehlern bezogen auf „man-machine systems“

(Chapanis 1951, S. 1181). Sie basiert auf der Idee, dass eine Analyse der Fehlerverteilung dabei helfen kann, Systemprobleme zu entdecken und dadurch Gegenmaßnahmen ergreifen zu können (Hofinger 2012, S. 48). Durch das Wissen, welche Fehler mit welcher Häufigkeit auftreten, kann die Notwendigkeit von Gegenmaßnahmen abgeleitet werden (Hofinger 2012, S. 48). Bei dieser Klassifikation steht nicht das Handeln des Menschen im Vordergrund, „sondern das Auftreten von unerwünschten Ereignissen in einem Arbeitssystem, unabhängig von ihrer Verursachung“ (Hofinger 2012, S. 48-49). In diesem Kontext kann auf die beschriebenen Bugs in Kapitel 2.1.1.2 verwiesen werden, da es sich hierbei ebenfalls um „unerwünschte Ereignisse in einem Arbeitssystem“ handelt. Auch im Falle der Bugs ist nicht die handelnde Person, wie zum Beispiel der Nutzer einer App, die Fehlerquelle, sondern die fehlerhaften Programmierungen eines Softwareentwicklers. Trotz des Alters dieser Klassifikation ist sie für eine Einordnung und Vorhersage von Fehlerursachen, die unabhängig von den handelnden Personen sind, immer noch aktuell (Hofinger 2012, S. 49).

Da Chapanis (1951, S. 1181-1182) in seiner Darstellung zur Fehlerklassifikation die zufälligen Fehler nicht mit einbezieht, sondern diese lediglich im Text in Anlehnung an Schlink (1919) benennt, zeigt Abbildung eins Hofingers Darstellung (2012) zu den drei von Chapanis erwähnten Fehlerarten:

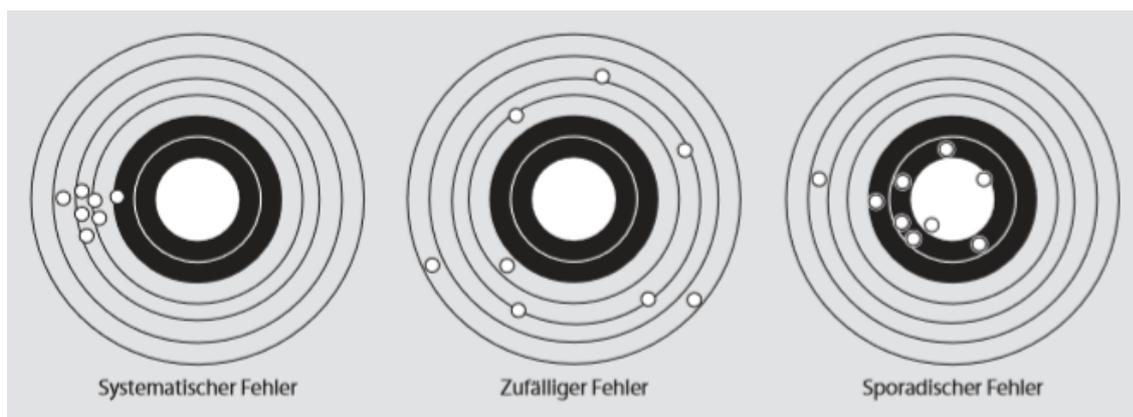


Abbildung 1: Fehlerklassifikation nach Auftretenshäufigkeit in Anlehnung an Chapanis (1951) (Hofinger 2012, S. 49)

Systematische Fehler oder laut Chapanis (1951, S. 1181) „Constant Errors“ sind Fehler, welche stetig während eines spezifischen Arbeitsprozessschrittes in immer dergleichen Art und Weise passieren (Hofinger 2012, S. 49). Chapanis (1951, S. 1181) definiert diese Art des Fehlers wie folgt: „A constant error is the difference between the average

of a large series of measurements and the true, or expected, value.” In diesem Fall ist der angestrebte Wert die Mitte der Scheibe zu treffen, wodurch die Distanz der Treffer vom Mittelpunkt den Fehler definiert (Reason 1990, S. 4). Da der Fehler immer wieder während eines bestimmten Schritts im Arbeitsprozess passiert, ist der Fehler vorhersagbar (Reason 1990, S. 4). Hofinger (2012, S. 49) merkt an, dass diesen Fehlern durch Veränderungen der Prozesse oder der verwendeten Geräte sowie durch eine veränderte Personalauswahl oder ein Personaltraining entgegengewirkt werden kann. Ein Praxisbeispiel ist, dass bei personalisierten Werbeanzeigen im Internet den Kunden immer wieder Produkte angezeigt werden, die diese kurz zuvor bereits gekauft haben (Jansen 2018, S. 1). Eine Veränderung der Prozesse kann dazu beitragen, diese Fehler auszuschließen. Zufällige Fehler haben laut Hofinger (2012, S. 49) unterschiedliche Ursachen und werden in Abbildung eins durch das zweite Trefferbild gezeigt. Sie sind nicht absehbar und können in ihrem Auftreten nur geschätzt werden (Chapanis 1951, S. 1182). Auch Hofinger (2012, S. 49) sagt, dass diese Fehler nicht zu verhindern sind, aber beispielsweise durch zusätzliche Sicherheitssysteme, wie „Fail-Safe-Systeme“ eingegrenzt werden können. Bei sporadischen Fehlern, oder nach Chapanis (1951) „Variable Errors“, handelt es sich laut Hofinger (2012, S. 49) um Fehler, die selten und ohne erkennbares Muster auftreten. Sie können durch eine statistische Größe gemessen werden, welche die Streuung oder Verteilung der einzelnen Messungen definiert (Chapanis 1951, S. 1181). Sie sind schwieriger vorherzusagen als die systematischen Fehler, was auch an dem dritten Trefferbild in Abbildung eins gesehen werden kann. Hofinger (2012, S. 49) merkt an, dass diese Fehler aufgrund ihrer Seltenheit und Unterschiedlichkeit eine Analyse und Entdeckung dieser Fehler verkomplizieren und Kosten für zusätzliche Sicherheitsanstrengungen verursachen.

Es bleibt anzumerken, dass die Klassifikation von Chapanis (1951) durch ihre allgemeine Auslegung auf viele Arbeitssysteme angewendet werden kann. Für den Themenschwerpunkt dieser Masterarbeit sind in Anlehnung an Chapanis‘ Klassifikation (1951) mehrere Überlegungen sinnvoll: Zum einen ist es interessant zu hinterfragen, wer über die Daten, die für den Arbeitsprozess genutzt werden, verfügen kann und wer welche Berechtigungen für den Umgang mit diesen hat. Der Entwickler einer Datenverarbeitungssoftware sollte beispielsweise über jegliche Verfügungsrechte für den Umgang und über Einsicht in alle Schritte des Arbeitssystems verfügen. Dadurch kann

eine Verknüpfung der Daten transparent und in allen Schritten nachvollziehbar gemacht werden. Dies bietet durch die Erhöhung der Transparenz im besten Fall die Möglichkeit, Fehler, wie beispielsweise die systematischen Fehler, während der Entwicklungsphase schnell aufzudecken, zu reduzieren oder bereits ganz zu vermeiden.

2.1.2.2. Theoretischer Erklärungsansatz bezogen auf Arbeitssysteme

Hofinger (2012, S. 49) bezieht in ihren Ausführungen auch die Klassifikation von Meister (1971, S. 26) mit ein: Meister (1971, S. 29-30) systematisiert Fehlerarten anhand der technischen Teilstufe in der Systementwicklung, in der diese aufgetreten sind, beziehungsweise in der diese durch eine menschliche Fehlhandlung initiiert wurden. Dabei basiert die Systementwicklung auf vier Stufen: Einer Designstufe, einer Produktionsstufe, einer Teststufe und einer Ausführungsstufe, in der das System letztlich nach Design, Produktion und Test ausgeführt wird (Meister 1971, S. 29). Hier ist anzumerken, dass in diesem Modell der Systementwicklung die Komplexität der einzelnen Stufen nicht beachtet wird und die Grenzen zwischen den Stufen gegebenenfalls nicht immer trennscharf sind (Meister 1971, S. 29). Es geht laut Meister (1971, S. 29) hierbei um Folgendes: „It is sufficient to say that equipment is designed in the design stage, turned into hardware during production, tested to ensure that it meets specification during test, and operated by its users during operations.“ Innerhalb der vier von Meister (1971, S. 29) benannten Systementwicklungsstufen, können seiner Aussage nach folgende Fehlerarten gefunden werden: „Design errors“, „production/workmanship errors“ oder auch „fabrication errors“ und „inspection errors“, „installation/maintenance errors“ und „operating errors“. Hofinger (2012, S. 49) benennt die „production/workmanship errors“ als „manufacturing error“.

Ein „design error“ tritt laut Meister (1971, S. 26) beispielsweise auf, wenn unangemessene Funktionen zugeteilt wurden, die Anforderungen unzureichend oder fehlerhaft umgesetzt wurden, oder das technische Design durch den Menschen nicht gut entworfen wurde. Das Resultat aus diesem „design error“ ist eine ungeeignete und ungenügende Designgrundlage für die weiteren Schritte der Systementwicklung (Meister 1971, S. 26).

Ein „production error“ ist laut Meister (1971, S. 38) eigentlich eher als „workmanship error“ zu bezeichnen, da es sich hierbei meist um Fehler von individuellen Arbeitern handelt, die dafür sorgen, dass die Anforderungen an den zu produzierenden Artikel nicht erreicht werden. In einer zusammenfassenden Tabelle der Fehlerklassifikation ordnet Meister (1971, S. 26) der Produktionsstufe auch den „fabrication error“ und den „inspection error“ zu. Die Fehlerarten dieser Stufe können mehrere Gründe, wie fehlerhafte Entwürfe und Vorgaben oder ungeeignete Werkzeuge, ein unzureichendes Wissen oder eine inadäquate Arbeitsumgebung, haben (Meister 1971, S. 26). Sie können aber auch das Resultat eines vorhandenen „design errors“ sein, der nicht entdeckt und dadurch als Vorgabe hingenommen wurde (Meister 1971, S. 26). Die Folge aus dieser Fehlerart sind dann unter anderem höhere Kosten für das Unternehmen, Betriebsmittel, die entweder nutzlos sind oder überarbeitet werden müssen, und Produktionsverspätungen (Meister 1971, S. 26).

Der Teststufe hat Meister (1971, S. 26) den „operating error“, „installation error“ und „maintenance error“ zugeordnet. „Installation errors“ treten im Allgemeinen nur auf, wenn das System installiert wird, und reduzieren sich drastisch, wenn die Einführung des Systems vorbei ist (Meister 1971, S. 42). „Maintenance errors“ auf der anderen Seite können jederzeit auftreten und nehmen häufig zu, wenn das System in die Jahre kommt (Meister 1971, S. 42). „Operating errors“ geschehen in der Teststufe, wenn bei dem Test des Systems durch die Mitarbeiter festgestellt wird, dass das System nicht in der Art und Weise funktioniert, wie es geplant war (Meister 1971, S. 43). Die Fehler dieser Stufe treten zum Beispiel durch unzureichende oder unvollständige technische Daten, ungeeignete Logistik, eine mangelhafte Arbeitsumgebung und ein, in der Designstufe mangelhaft entworfenes, technisches Design auf (Meister 1971, S. 26). Diese Fehler resultieren beispielsweise in verspäteten Systemoperationen, Systemabstürzen, einer Verschlechterung der Systemperformance und potenziellen Gefahren bei der Ausführung der Systeme (Meister 1971, S. 26).

In der Ausführungsstufe des Systems durch den Anwender und nicht mehr durch die Mitarbeiter, kann es ebenfalls zu einem „operating error“, „installation error“ oder „maintenance error“ kommen (Meister 1971, S. 26). Die Gründe und Resultate für diese Fehler sind ähnlich denen in der Teststufe. Ein Unterschied liegt darin, dass in dieser

letzten Stufe das System für den eigentlichen Nutzer verfügbar gemacht wurde und dieser nun zum Beispiel einen „operating error“ durch die Nutzung des Systems auslösen könnte (Meister 1971, S. 43). Weitere Fehlerursachen in dieser Stufe können eine geringe Motivation, unzureichendes Wissen über die Ausführung, eine Überforderung durch die Konditionen des Systems oder eine zu hohe Komplexität der Ausführung sein (Meister 1971, S. 26).

Obwohl die Klassifikation nach Meister (1971) schon relativ alt ist, zeigt das Aufgreifen dieser Theorie durch Hofinger (2012), dass diese trotzdem noch einen Aktualitätswert hat. In diesem Zusammenhang kann zudem angemerkt werden, dass diese Systementwicklungsstufen auch bei jeglicher Big Data Technologie vorzufinden sind. Eine Betrachtung von Fehlern mit Blick auf diese Klassifikation macht zudem deutlich, dass die Fehlerursache nicht immer zwangsläufig bei der handelnden Person zu suchen ist, sondern räumlich und zeitlich an ganz anderer Stelle liegen kann (Hofinger 2012, S. 49). Hat ein Programmierer beispielsweise eine selbstlernende Maschine entwickelt, die eine Bilderkennung durchführen soll und hat dieser Maschine aber in der Designstufe die falschen Bilder als Beispiele gezeigt, dann liegt der Fehler nicht daran, dass die Maschine etwas falsch gelernt hat. Interpretiert die Maschine in diesem Zusammenhang das Bild eines Menschen als Bild eines Tieres, liegt die Ursache des Fehlers in diesem Beispiel beim Programmierer. Es wird also deutlich, dass diese Klassifikation sich auch heute noch für die Einordnung von Fehlern eignet.

2.1.2.3. Theoretischer Erklärungsansatz bezogen auf Big Data Risiken

Nach den Ausführungen zu den Klassifikationen von Chapanis (1951) und Meister (1971) folgt in diesem Kapitel eine Betrachtung von Risiken mit Bezug auf Big Data Maßnahmen. Dies ist vor allem für die Fehlerprävention bei Datenanalysen im Bereich des

E-Commerce von Bedeutung. Diese ist nur möglich, wenn der Unternehmung die Risiken bekannt sind. In diesem Kapitel geht es daher darum, eine Klassifikation der Risiken aufzuzeigen, damit Präventionsmaßnahmen im Rahmen eines Risikomanagements bei Datenanalysen im Bereich des E-Commerce systematisiert werden können. Hierbei orientiere ich mich an den Ausführungen des Bundesverbands für Informationswirtschaft, Telekommunikation und neue Medien e.V. (BITKOM) zu den

Risiken bei Big Data Anwendungen (BITKOM 2014, S. 136-146). Die Ausführungen der BITKOM (2014, S. 136) dienen nicht dazu, vor Big Data Methoden zu warnen, sondern sollen eine Aufklärung über die zum Teil neuen Risiken dieses Bereiches bieten. Dies deckt sich mit dem Versuch, in dieser Arbeit Strategien zu Fehlerprävention und -management aufzuzeigen, um ein Fehlerbewusstsein in den Unternehmen zu wecken. Laut BITKOM (2014) ergeben sich trotz der möglichen Erfolgspotenziale beim Einsatz von Big Data vor allem Gefahren der Schädigung oder sogar des Verlusts der Reputation: „Es kann zu einem Vertrauensbruch zwischen Geschäftspartnern sowie zwischen Unternehmen und Kunden kommen.“ (BITKOM 2014, S. 136) Imageschäden entstehen hierbei durch vorhergehende fehlerhafte Handlungen, die mitunter aus menschlichen Fehlhandlungen resultieren (BITKOM 2014, S. 136). Fehler entwickeln sich oft aus unzureichenden Kenntnissen und Fähigkeiten in dem spezifischen Bereich, weshalb es für ein Unternehmen umso wichtiger ist, diese Risikobereiche zu kennen und diesen entgegenzuwirken (BITKOM 2014, S. 136). Die BITKOM (2014, S. 137) macht deutlich, dass eine erfolgsversprechende Umsetzung von Big Data Projekten nur dann erfolgen kann, wenn „Kompetenzen in den Bereichen des Data Managements, des Data Warehousings, der Datenbanken, der IT-Infrastruktur, der Skriptsprachen, des Enterprise-Content- und Document-Managements sowie der Business Intelligence“ vorhanden sind. Außerdem werden Kenntnisse im Bereich des Datenschutzes benötigt (BITKOM 2014, S. 137), was durch die neue DS-GVO im Mai 2018 (Bundesministerium für Wirtschaft und Energie o.D.) nur noch an Bedeutung gewonnen hat. Im Normalfall existieren laut der BITKOM (2014, S. 137) diese Kompetenzen in den Unternehmen an unterschiedlicher Stelle und müssen für Big Data Maßnahmen zusammengeführt werden, um die Risikowahrscheinlichkeiten einzudämmen. Die unterschiedlichen Risikobereiche von Big Data sind in Abbildung zwei dargestellt:

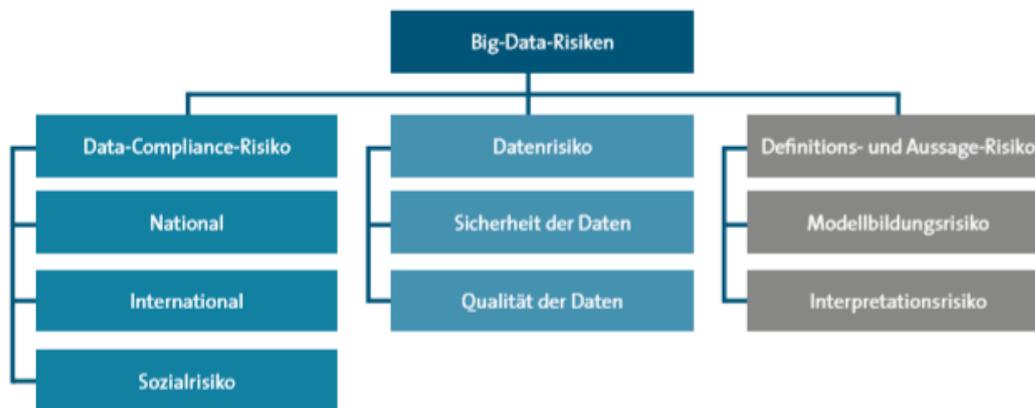


Abbildung 2: Risikobereiche von Big Data (BITKOM 2014, S. 136)

Der Bereich der Data-Compliance-Risiken umfasst die nationalen und internationalen gesetzlichen Regelungen bezogen auf den Datenschutz und die Nutzung von Daten sowie ein Sozialrisiko (BITKOM 2014, S. 137-139). Dieses Sozialrisiko bezieht sich dabei auf wert-, moralbezogene und ethische Fragen im Zusammenhang mit Big Data, weil beispielsweise die Verwendung von Daten aus den sozialen Medien im Zuge von personalisierter Werbung den Nutzern das Gefühl geben könnte, überwacht zu werden (BITKOM 2014, S. 139). Generell sind die gesammelten Daten von Unternehmen in anonymisierter Form zu verwenden, um keine Rückschlüsse auf reale Personen zuzulassen, und personenbezogene Daten dadurch zu schützen (BITKOM 2014, S. 138). Auf die aktuellen gesetzlichen Bestimmungen zur Verarbeitung personenbezogener Daten wird in der Analyse in Kapitel 3 eingegangen. Bevor Daten von Unternehmen verwendet werden, sollten diese sich im Klaren darüber sein, welche Regulierungen vorliegen, um dadurch Stolpersteine zu vermeiden. Ziel sollte es laut der BITKOM (2014, S. 139) sein, „einen Ausgleich und Konsens zwischen den Informations-Interessen der Unternehmen und dem Recht auf informationelle Selbstbestimmung der Betroffenen zu finden.“

Der zweite Risikobereich enthält die Datenrisiken (BITKOM 2014, S. 140). Hierbei geht es um den Sicherheits- und den Qualitätsaspekt der Daten auf technischer Ebene. TCP/IP ist als Basisprotokoll des Internets nicht für die Verarbeitung der enormen und sensiblen Datenmengen geeignet und müsste durch zusätzliche Technologien erweitert werden (BITKOM 2014, S. 140). Bezogen auf den Sicherheitsaspekt der Daten ist es laut der BITKOM (2014, S. 140) wichtig, die Datenansammlungen vor unbefugten Personen zu schützen, um dadurch eine Manipulation der Daten zu verhindern. Dazu

muss sich ein Unternehmen die Fragen stellen, wie, wozu und welche Art von Daten gesammelt werden sollen und wie diese sicher gespeichert und vor unberechtigten Zugriffen durch Schutzmechanismen- und strategien geschützt werden können (BITKOM 2014, S. 140). In diesem Zusammenhang sind effektive Verschlüsselungen und Zugriffsbeschränkungen für den Schutz der Daten enorm wichtig (BITKOM 2014, S. 140). In Bezug auf den Qualitätsaspekt der Daten kann angemerkt werden, dass es bereits in einem frühen Stadium der Datenanalyse von Bedeutung ist, die relevanten Daten für wertvolle Ergebnisse herauszufiltern (BITKOM 2014, S. 140). Hierbei ist eine Prüfung der Daten auf Vollständigkeit nicht ausreichend, da zudem geschaut werden sollte, ob die Daten aktuell, authentisch und zu jeder Zeit verfügbar sind (BITKOM 2014, S. 140-141).

Unter dem Begriff Definitions- und Aussagerisiko werden die Risiken bei Modellbildung und Ergebnisinterpretation zusammengefasst (BITKOM 2014, S. 141-142). Bezogen auf das Modellbildungsrisiko erläutert die BITKOM (2014, S. 141), dass es für Big Data Maßnahmen keine Pauschallösungen gibt, und dass gesetzliche Regelungen beachtet sowie eine genaue Zielvorstellung des Unternehmens und dafür notwendige Analysetechniken festgelegt werden müssen. Wurden Analysetechniken festgelegt, ist es trotzdem weiterhin wichtig, die Ergebnisse dieser Analysen mit Blick auf die festgelegten Ziele des Unternehmens durch die Big Data Maßnahme zu prüfen (BITKOM 2014, S. 142). Stimmen Auswahl, Entwicklung und Anwendung des Modells, bleibt das Interpretationsrisiko (BITKOM 2014, S. 142). Fehlinterpretationen sind vor allem durch unzureichend qualifizierte Mitarbeiter möglich, da viele Modelle die Zusammenhänge zwischen den Daten zwar fehlerfrei und fachgemäß darstellen, aber nicht über eine verständliche Darstellung verfügen (BITKOM 2014, S. 142). „Big Data kann (...) zunächst nur für denjenigen einen Nutzen und ein verwertbares Ergebnis erzeugen, der die Regeln der Interpretation kennt.“ (BITKOM 2014, S. 142).

2.1.3. Zum Fehlermanagement und dessen organisatorischer Verortung in Bezug auf Datenanalysen im Bereich des E-Commerce

Aufgrund der Entwicklungen der letzten Jahre sind die Investitionen durch Unternehmen in die Digitalisierung und dadurch in Daten und Software immer mehr

angestiegen (Ringel et al. 2018, S. 8-9). Dabei stellt sich die Frage, welche neuen Kapazitäten ein Unternehmen braucht, um erfolgreich in diesem Bereich zu agieren. Laut Ringel et al. (2018, S. 9) gibt es dabei unter anderem das Problem, dass neue Mitarbeiter, wie Datenwissenschaftler und Softwareentwickler, gebraucht werden (siehe auch Davenport und Patil 2012). Daraus entsteht das Problem, dass das Wissen in der Industrie über die neuen Möglichkeiten durch Daten und Software begrenzt ist und die neuen Stellen mit fachkundigen Mitarbeitern besetzt werden müssen (Ringel et al. 2018, S. 9). Ein weiteres großes Problem ist die Digitalisierung der Wertschöpfungskette, damit die digitalen Prozesse mit einer adäquaten Geschwindigkeit unterstützt werden können (Ringel et al. 2018, S. 9). Hieraus ergibt sich auch ein Umschwung im Bereich des Fehlermanagements, da die Nutzung von Daten und Software neue Probleme und Fehlerquellen bietet, die noch nicht ausreichend bekannt sind. Auch hierfür werden Fachkräfte und neue beziehungsweise veränderte Abteilungen und Arbeitsgruppen nötig sein, um diese Schwierigkeiten mit dem notwendigen Wissen angehen zu können und so eine solide Basis für ein betriebliches Fehlermanagement in Bezug auf Datenanalysen zu etablieren.

Für ein Unternehmen kann ein Fehler zu negativen Konsequenzen, wie zum Beispiel Zeitverlust oder mangelhafter Produktqualität führen (van Dyck et al. 2005, S. 1228). In wenigen Fällen können Fehler aber auch Anstoß für Innovations- und Lernprozesse sein (van Dyck et al. 2005, S. 1228). In der wissenschaftlichen Literatur zu der Thematik des Fehlers werden laut van Dyck et al. (2005, S. 1228) zum Großteil die negativen Folgen eines Fehlers beleuchtet, ohne den möglichen Lerneffekt aus Fehlern zu benennen. Vielmehr geht es um eine möglichst frühe Fehlerprävention und das Abwehren von Fehlern, sobald Schwachstellen bemerkt werden (van Dyck et al. 2005, S. 1228). Das Fehlermanagement, auf welches van Dyck et al. (2005, S. 1228) eingehen, greift alle möglichen Konsequenzen eines Fehlers auf. Es wird davon ausgegangen, dass Fehler nie ganz verhindert werden können und dass frühzeitige Überlegungen eines Umgangs mit Fehlern wichtig sind. Laut van Dyck et al. (2005, S. 1228) definiert sich das Fehlermanagement hierbei durch ein Reduzieren der negativen Fehlerkonsequenzen und ein Erhöhen der möglichen Lern- und Innovationseffekte, falls ein Fehler auftritt. Ein Unternehmen sollte den Fokus aber sowohl auf das Fehlermanagement, als auch auf die Fehlerprävention legen, um so einen optimalen

Umgang mit Fehlern gewährleisten zu können (van Dyck et al. 2005, S. 1229). Diese Ansicht spiegelt sich auch in der Thematik dieser Masterarbeit wider, da es sowohl um das Management von Fehlern bei Datenanalysen im Bereich des E-Commerce als auch um die Fehlerprävention gehen soll. Anhang eins zeigt, an welchen Stellen zwischen Handlung, Fehler und Fehlerkonsequenz Fehlerprävention und -management integriert sind (Frese und Keith 2015, S. 666). In Anlehnung an Anhang eins ist anzumerken, dass die Fehlerprävention vor dem eigentlichen Fehler angeordnet ist. Diese hat eher eine negative Sicht auf fehlerhafte Handlungen. Die Intention liegt jedoch darin, Fehlerhäufigkeiten auf ein Minimum zu reduzieren. In Anlehnung an Anhang eins setzt das Fehlermanagement an, wenn ein Fehler aufgetreten ist. Vor dem Hintergrund, dass Menschen nie fehlerfrei arbeiten, kümmert es sich um eine effektive Aufarbeitung von Fehlern. In der Folge können Unternehmen aus den aufgetretenen Fehlern lernen, um dadurch die zukünftige Fehlerrate zu senken. Das Fehlermanagement bietet Unternehmen außerdem laut van Dyck et al. (2005, S. 1229) folgende Möglichkeit: „Organizations may be able to benefit from simultaneously pursuing the goal of control and the goal of learning, and we propose that error management is well suited for supporting such an approach.“ So sind sowohl das Controlling als auch das Risikomanagement Bereiche im Unternehmen, die sich mit fehlerhaften Handlungen befassen. Nach van Dyck et al. (2005, S. 1229) kann ein solches Fehlermanagement dazu führen, dass Fehlerquellen frühzeitig erkannt, Fehlerraten minimiert und die Fehlerkonsequenzen besser kontrolliert werden. Auf der anderen Seite setzt sich das Fehlermanagement damit auseinander, die als fehlerhaft gesehene Handlung auch zu Lernzwecken zu nutzen und unterstützt zudem das Experimentieren und Erkunden neuer Handlungswege, was Innovationen innerhalb des Unternehmens begünstigt (van Dyck et al. 2005, S. 1229).

Hagen (2017, S. 181) stellt sich die Frage, ob ein Fehlermanagement nur für Risikobereiche, die die Verantwortung für Menschenleben umfassen, von Bedeutung ist oder für alle Unternehmen. In diesem Zusammenhang kommt er zu dem Schluss, dass theoretisch jedes Unternehmen einen Risikobereich hat, da Geschäftsabläufe, Erfolge des Unternehmens und Arbeitsplätze von Mitarbeitern durch das Unternehmen verantwortet werden (Hagen 2017, S. 181). Dementsprechend sollte in allen Unternehmen laut Hagen (2017, S. 181) nach einer niedrigen Fehlerquote gestrebt,

potenzielle Fehlerquellen ausgeschaltet und bereits geschehene Fehler behoben werden. Auch Hagen (2017, S. 181-182) erläutert, dass das moderne Fehlermanagement fehlerhafte menschliche Handlungen als unvermeidlich ansieht und akzeptiert. Fehler werden nach diesem modernen Verständnis analysiert und es wird dafür gesorgt, dass der Fehler zukünftig nicht mehr passieren kann (Hagen 2017, S. 181-182). Um ein konkretes Fehlermanagement im Sinne dieser modernen Auffassung umzusetzen, muss laut Hagen (2017, S. 182-185) ein Umdenken bezogen auf Fehler geschehen, damit eine fehlerhafte Handlung nicht allein verurteilt wird und Konsequenzen für die Mitarbeiter und das Unternehmen in einer sinkenden Fehlerrate resultieren. Dabei hängt es davon ab, dass die weisungsbefugten Führungskräfte dieses Verständnis des Fehlermanagements vorleben und ihre Mitarbeiter so dazu motivieren, sich stetig zu verbessern und die vorgefallenen Fehler zukünftig zu vermeiden (Hagen 2017, S. 184). Das Fehlermanagement ist damit auf höchster Ebene im Unternehmen anzusiedeln, da Führungskräfte auch diejenigen sind, die Handlungen als fehlerhaft bewerten und die Konsequenzen aus den fehlerhaften Handlungen ziehen. Hagen (2017, S. 188-189) versteht unter einem modernen Fehlermanagement die Aufteilung in drei Phasen. Auch hier ist festzuhalten, dass die ersten zwei Phasen die Fehlerprävention in den Vordergrund stellen, da hier die Grundlagen dafür geschaffen werden, dass Fehler auf das Minimum reduziert werden. In der ersten Phase geht es laut Hagen (2017, S. 188) um das Schaffen einer Sicherheitsgrundsituation und das Festlegen von Verfahrensstandards, damit durch Trainings der Mitarbeiter und abzuarbeitende Checklisten die Fehlerrate minimal bleibt. In Phase zwei werden potenzielle Fehlerquellen aufgelistet und während der Ausführung der unternehmerischen Geschäftsaktivitäten stetig ergänzt (Hagen 2017, S. 189). Diese Vorsichtsmaßnahmen von Unternehmen sollten bereits ganz an den Anfang der Wertschöpfungskette stehen und als Grundlage für alle weiteren wertschöpfenden Aktivitäten gelten. Zudem wird in der zweiten Phase laut Hagen (2017, S. 189) im besten Fall die Handlung jedes einzelnen durch eine zweite Instanz kontrolliert. Ab diesem Punkt und in der dritten Phase geht es dann eher um das Controlling und in diesem Zusammenhang auch um Evaluation von geschehenen Fehlern, um diese zum Anlass zu nehmen, dass solche Fehler nicht nochmals auftreten. Hagen (2017, S. 189) erläutert ebenfalls, dass die dritte Phase die Ansprache der aufgetretenen Fehler im Team umfassen sollte.

2.2. Begriff, Einordnung und Ziele von Big Data im Zusammenhang mit Datenanalysen im Bereich des E-Commerce

2.2.1. Begriffssystematisierung von Informationen und Daten

Daten und die daraus ableitbaren Informationen bilden den Kern jeder Datenanalyse. Clarke (2016, S. 79) erläutert, dass eine eindeutige Definition des Datenbegriffs im Allgemeinen nicht möglich ist und auch nicht aufrechterhalten werden kann, da sich die Bedeutung des Begriffs je nach Verwendung der Daten verändern und es so zu unklaren und mehrdeutigen Definitionen des Datenbegriffs kommen kann. Für ihn gibt es aber unterschiedliche Hauptfaktoren, die während der Nutzung einer Datenmenge eine Definition der Daten möglich machen. Zum einen muss eindeutig sein, mit welcher realen Identität beziehungsweise Instanz die Daten verbunden sind (Clarke 2016, S. 79). Zum anderen muss Klarheit darüber herrschen, welche realen Attribute die Daten abbilden sollen und im Weiteren, welchen Stand der Inhalt dieser festgelegten Attribute repräsentieren soll (Clarke 2016, S. 79). In seinen Ausführungen merkt Clarke (2016, S. 79-80) allerdings Folgendes an: „At the time of use, the more or less clear definition is overlaid by the perspectives and interpretations of the data’s user(s).“ Die Definition des Datenbegriffs hängt demnach von der individuellen Interpretation und Anwendung des Datenbestandes ab.

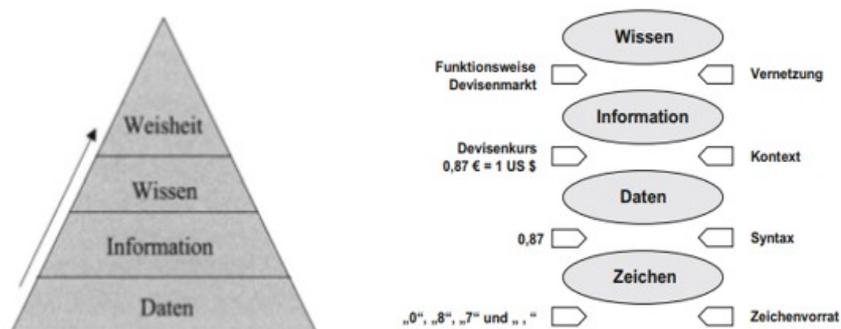


Abbildung 3: Linke Abbildung: Informationspyramide nach Voß und Gutenschwager (2001, S. 13), Rechte Abbildung: Begriffshierarchie nach Krcmar (2015, S. 4)

Um den Begriff der Daten weiter zu systematisieren und die Verarbeitung von großen Datenmengen zu einem wirtschaftlichen Mehrwert im Kontext von Big Data besser zu verstehen, beziehe ich mich in dieser Arbeit auf die Informationspyramide von Voß und Gutenschwager (2001, S. 13) und auf die Begriffshierarchie nach Krcmar (2015, S. 4).

Abbildung drei zeigt die Informationspyramide nach Voß und Gutenschwager (2001, S. 13) und die Begriffshierarchie nach Krcmar (2015, S. 4).

Voß und Gutenschwager (2001, S. 9-10) greifen zunächst Wittmanns Definition (1959, S. 14) auf, der den Begriff der Information mit zweckorientiertem Wissen gleichsetzt (Wittmann 1959, S. 14). Dieses Wissen bezieht sich auf einen zukünftigen Zeitpunkt, an dem ein bestimmter Zweck erreicht werden soll. Der Begriff umfasst Beobachtungen und Erfahrungen, auf deren Grundlage anschließend neue Schlüsse gezogen werden können (Wittmann 1959, S. 14). Voß und Gutenschwager (2001, S. 9-10) erweitern diese Definition um den Prozess des Verstehens von Daten, da die Relevanz bestimmter Daten nicht direkt erkenntlich ist. Informationen dienen demnach dazu, ein zweckorientiertes Wissen zu bilden. Laut Voß und Gutenschwager (2001, S. 10) liegt allem wirtschaftlichen Handeln und Entscheiden somit Wissen zugrunde. Den Prozesscharakter des Verstehens von Daten, um zu Wissen zu gelangen, stellen Voß und Gutenschwager (2001, S. 13) in einer Informationspyramide dar. Die Pyramide (Voß und Gutenschwager 2001, S. 13) zeigt, dass auf Grundlage von gesammelten Daten Informationen gewonnen werden können. Aus diesen Informationen ist es möglich, Wissen zu gewinnen, wobei sich die Datenmenge je nach Pyramidenebene stetig reduziert (Voß und Gutenschwager 2001, S. 13). Durch ein Verstehen von Prinzipien kann in einem letzten Schritt Weisheit erreicht werden (Voß und Gutenschwager 2001, S. 13). Allerdings sollte die Pyramide nicht so verstanden werden, dass die einzelnen Ebenen der Pyramide zwangsläufig aufeinander aufbauen (Voß und Gutenschwager 2001, S. 13). Während der Prozess des Datensammelns keine Ansprüche an die Fähigkeiten der beteiligten Personen stellt, ist die Gewinnung von Informationen aus diesen Daten komplizierter (Voß und Gutenschwager 2001, S. 13). An dieser Stelle kann die Unterteilung in relevante und nicht relevante Daten für die Entscheidungsunterstützung ein Problem sein. Die beteiligten Personen müssen in diesem Schritt in der Lage sein, Beziehungen zwischen Daten zu erkennen, diese herauszustellen und somit für mehr Transparenz zu sorgen (Voß und Gutenschwager 2001, S. 13). Voß und Gutenschwager (2001, S. 13) merken in diesem Schritt zudem an, dass die beteiligten Personen hierbei in ihren Entscheidungen über Beziehungen zwischen Datenmengen übereinstimmen müssen, um grundsätzlich einen Mehrwert generieren zu

können. Die letztliche Entscheidungsgewalt liegt in diesem Zusammenhang trotzdem bei einer Entscheidungsinstanz innerhalb des Unternehmens. Diese ist durch das Verständnis der Bezüge in der Lage, das Firmenwissen zu mehrern (Voß und Gutenschwager 2001, S. 13-14). Laut Voß und Gutenschwager (2001, S. 14) erfordert die Stufe der Weisheit, dass alle Zusammenhänge zwischen den Daten und somit das zugrundeliegende System für die Entscheidungsinstanz bekannt und verstanden sind. Anzumerken ist, dass sich dieses System stetig erweitert, da immer mehr Daten generiert werden und verarbeitet werden müssen (Voß und Gutenschwager 2001, S. 14).

Krcmar (2015, S. 3-4) macht deutlich, dass die Nutzung von Informations- und Kommunikationstechnologien eine Abgrenzung der Begriffe Daten, Information und Wissen erfordert, da für diese Technologien die Termini der Daten-, Informations- und Wissensverarbeitung nicht synonym genutzt werden können. In der Begriffshierarchie steht laut Krcmar (2015, S. 4) das Zeichen an der untersten Stelle, da es die Grundlage der übergeordneten Begriffe bildet. Aufgrund einer Anordnung von Zeichen in einen „definierten, strukturierten Zusammenhang“ kann laut Krcmar (2015, S. 4) von Daten gesprochen werden. Durch eine Anreicherung der daraus gewonnenen Daten mit Kontext, ergeben sich Informationen, die durch eine weitere Vernetzung mit anderen vorhandenen Informationen zu Wissen führen (Krcmar 2015, S. 4). Ein Beispiel für diese Hierarchie zeigt Abbildung drei. Krcmar (2015, S. 5) macht deutlich, dass eine Information auch als Produktionsfaktor in der betrieblichen Leistungserstellung gesehen werden kann und somit als immaterielle Ressource in der Betriebswirtschaftslehre gilt. Shapiro und Varian (2008, S. 3) definieren den Begriff Information sehr allgemein als alles, was digitalisiert werden kann. Auch sie sehen die Information als Produkt und machen deutlich, dass es zwar kostenintensiv ist ein Informationsprodukt herzustellen, die Reproduktion allerdings sehr günstig ist (Shapiro und Varian 2008, S. 3).

2.2.2. Zur Begriffsdefinition und Einordnung von Big Data

2.2.2.1. Begriffsdefinition von Big Data

In dieser Masterarbeit beziehe ich mich auf Big Data Methoden im Bereich des E-Commerce, da sich das Internet über die Jahre zu einem zentralen Marktplatz für die

Transaktionen von Produkten und Dienstleistungen entwickelt hat (Verhoef et al. 2016, S. 2). Der Begriff Big Data ist in diesem Zusammenhang der zentrale Ausgangspunkt, wenn es um Datenanalysen innerhalb dieses Bereiches geht. Buhl et al. (2013, S. 24) merken an, dass 2013 70 Prozent der akademisch relevanten Literatur zu Big Data in den zwei Jahren zuvor publiziert wurden und von einem Big Data Hype gesprochen werden kann. Aus den vorangehenden Gründen wird der Begriff Big Data in diesem Kapitel genauer untersucht und seine Bedeutung für diese Arbeit herausgestellt.

Durch die Nutzung des Begriffs Big Data in unterschiedlichen wissenschaftlichen Disziplinen und Kontexten, ist dieser zu einem allgegenwärtigen Terminus geworden (Mauro et al. 2015, S. 97). Problematisch ist hierbei, dass dadurch eine einheitliche und strukturierte Begriffsdefinition, um die Bedeutung von Big Data zu erfassen, schwierig geworden ist (Mauro et al. 2015, S. 97). Aus diesem Grund haben sich Mauro et al. (2015, S. 97) die Kontexte und Disziplinen angeschaut, in denen der Begriff Big Data Verwendung findet und in diesem Zusammenhang vier Hauptthemen herausgestellt, die im Zusammenhang mit Big Data genannt werden. Hierzu gehören die Themen Informationen, Technologien, Methoden und Auswirkungen, die in den von Mauro et al. (2015, S. 97) betrachteten Verweisen auf Big Data am häufigsten angebracht werden. Im weiteren Verlauf ihrer Ausführungen erläutern Mauro et al. (2015, S. 98), dass auf Grundlage der von ihnen betrachteten Definitionen und der vier herausgestellten Hauptthemen in Bezug auf Big Data Folgendes festgehalten werden kann: Um Informationen verarbeiten und nutzen zu können, werden bestimmte Technologien und analytische Methoden benötigt (Mauro et al. 2015, S. 103). Diese führen im Idealfall zu Erkenntnissen mit einem wirtschaftlichen Wert für das Unternehmen, wodurch die Auswirkungen von Big Data auf Unternehmen und die Gesellschaft deutlich werden (Mauro et al. 2015, S. 103). Alles in allem fassen Mauro et al. (2015, S. 103) ihre gewonnenen Erkenntnisse in folgender Definition zusammen: „Big Data represents the Information assets characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value.“ In Kapitel 2.2.3 wird mit Blick auf die Charakteristika Volume, Velocity and Variety genauer auf Disziplinen eingegangen, in denen der Begriff Big Data genutzt wird.

Cavanillas et al. (2016, S. 3) definieren Big Data in ähnlicher Art und Weise als „the emerging field where innovative technology offers new ways of extracting value from the tsunami of new information.“ Hierbei sehen sie die Fähigkeit, aus Datenmengen Wissen zu extrahieren und die Informationen sinnvoll einzusetzen, als einen Hauptfaktor für Wettbewerbsvorteile (Cavanillas et al. 2016, S. 3). Außerdem fügen Cavanillas et al. (2016, S. 3) hinzu, dass die Verwendung von Big Data Technologien für viele Unternehmen eine Voraussetzung geworden ist, um auf dem Markt überleben und Wettbewerbsvorteile gegenüber der Konkurrenz etablieren zu können. Im Weiteren zitieren sie die EU-Kommissarin Kroes, die Big Data als „das neue Öl“ bezeichnet, welches dank neuer digitaler high-performance Technologien gesteuert, manipuliert und genutzt werden kann (Cavanillas et al. 2016, S. 5). Zudem führen Cavanillas et al. (2016, S. 10) an, dass es sich bei Big Data um eines der Hauptwirtschaftsgüter der Zukunft handelt und ein Ausschöpfen und Verstehen des Potenzials nicht nur den Wettbewerb europäischer Unternehmen verbessert, sondern auch für wirtschaftliches Wachstum und mehr Jobs sorgen wird. Dieser Ausbau des Wettbewerbs führt in Europa nicht nur zu einem Drang nach weiterer Innovation, sondern hilft auch dabei, dass Europa weiterhin an der Spitze der globalen, datengetriebenen, digitalen Wirtschaft bleibt (Cavanillas et al. 2016, S. 10).

Für diese Arbeit ergibt sich durch die angeführten Definitionen von Mauro et al. (2015, S. 103) und Cavanillas et al. (2016, S. 3) folgende Sichtweise auf den Begriff Big Data: Mit diesem Schlüsselbegriff wird auf Informationen als immaterielle Ressourcen des Unternehmens angespielt, die sich aus Datenanalysen ergeben. In diesem Kontext werden die Informationen aus Daten gewonnen, die sehr vielfältig sind, sich schnell verändern und in großem Ausmaß vorhanden sind. Um aus den Datenmengen Informationen generieren zu können werden spezifische innovative Technologien und Analysemethoden benötigt. Nur so können wertvolle Ergebnisse für das Unternehmen gewonnen werden.

2.2.2.2. Einordnung von Big Data im aktuellen Kontext

Daten werden bereits seit mehreren Jahrzehnten generiert. Die Besonderheit ist, dass im Zeitalter der Digitalisierung die Menge an den generierten Daten gänzlich andere Dimensionen annimmt als noch vor 30 oder 40 Jahren (Verhoef et al. 2016, S. 1). So

wird die jährlich generierte Menge an digitalen Daten für das Jahr 2025 auf 175 Zettabyte geschätzt, wobei es sich 2018 noch um ein Datenvolumen von 33 Zettabyte handelte (Reinsel et al. 2018, S. 3). Bereits in den 1990er Jahren investierten Firmen in die Erstellung großer Kundendatenbanken, um das Kaufverhalten und andere Kundencharakteristika abbilden zu können (Verhoef et al. 2016, S. 1-2). Vor allem die Einführung des Internets und der sozialen Medien haben zu einem enormen Zuwachs der produzierten Datenmengen und der Möglichkeit des Abrufs von täglichen beziehungsweise Echtzeit-Daten geführt (Verhoef et al. 2016, S. 2). Dadurch wird laut Verhoef et al. (2016, S. 2) dem Generieren von Mehrwert aus den großen Datenmengen eine wichtige Rolle in der wirtschaftlichen Entwicklung der nächsten Jahre zugesprochen.

Allerdings ist die Nutzung von Big Data nicht ganz unproblematisch (Buhl et al. 2013, S. 25). Ein erstes Problem bildet der große Aufwand, der mit den „ständigen Produkt- und Neuerhebungen der Kundendaten zur Produkt- und Serviceverbesserung über verschiedene Kanäle hinweg“ verbunden ist (Buhl et al. 2013, S. 25). Diese notwendige Aktualität der Datenmengen führt dann unter anderem zu einem sehr kosten- und zeitintensiven Datenmanagement, das sich entweder in einer Erhöhung der Kosten für die Kunden widerspiegeln kann (Buhl et al. 2013, S. 25) oder einen Eintritt in Big Data Analysen für kleinere Unternehmen schwer bis unmöglich macht, da die Kapazitäten hierfür fehlen. Hinzu kommt, dass Big Data Herausforderungen, wie Datenvolumen, -geschwindigkeit, -vielfalt und -glaubwürdigkeit mit sich bringt (Buhl et al. 2013, S. 26). Die wachsende Nutzung mobiler Endgeräte fördert zudem nicht nur die Anzahl der verfügbaren Datenquellen, sondern sorgt auch dafür, dass stetig neue Quellen betreut werden müssen, um eine Echtzeitverarbeitung der relevanten Daten überhaupt möglich zu machen (Buhl et al. 2013, S. 26). Aus diesen Gründen merken Buhl et al. (2013, S. 26) an, dass fast die Hälfte aller Big Data Projekte vorzeitig beendet wird. Dies war auch zwei Jahre später noch der Fall, da Marr (2015, S. 1) anmerkt, dass er ein Scheitern der Hälfte aller Big Data Projekte vorhersagt. Auch Verhoef et al. (2016, S. 3) merken an, dass es viele Unternehmen gibt, denen es schwerfällt, einen wirklichen wirtschaftlichen Mehrwert aus den enormen Datenmengen zu generieren. Ein weiteres Problem im Zusammenhang mit Big Data sind die zunehmend einschränkenden Datenschutzverordnungen, wie die Europäische Datenschutz-Grundverordnung (DS-

GVO) vom 25.05.2018 (Bundesministerium für Wirtschaft und Energie o.D.). Dieser Datenschutz auf europäischer Ebene soll „Lösungen zu Fragen, die sich durch „Big Data“ und neue Techniken oder Arten der Datenverarbeitung wie Profilbildung, Webtracking oder dem Cloud Computing für den Schutz der Privatsphäre“ bieten (Bundesministerium für Wirtschaft und Energie o.D.).

Trotzdem erläutern Buhl et al. (2013, S. 27), dass Untersuchungen gezeigt haben, dass Big Data sowohl für die Wissenschaft als auch für die Praxis relevant ist und eher als Chance statt als Bedrohung der Wirtschaft aufgefasst werden sollte. Das zeigt sich beispielsweise auch an dem Untertitel des Jahresberichts der Boston Consulting Group zu den „most innovative companies 2018“, der „innovators go all in on digital“ lautet (Ringel et al. 2018). Verhoef et al. (2016, S. 3) führen ebenfalls an, dass Big Data in der Wirtschaft laut Presseberichten zur Norm geworden ist, da Unternehmen verstanden haben, dass sie durch eine intelligente Analyse von Datenmengen in der Lage sind, im Wettbewerb stärker gegenüber ihrer Konkurrenz aufzutreten. Größere Unternehmen, wie Amazon, Google oder Netflix, analysieren bereits die Datenmengen, die ihre Nutzer generieren (Verhoef et al. 2016, S. 3). Buhl et al. (2013, S. 27) nennen als Beispiel für eine erfolgreiche Analyse von Datenbeständen den „Versandhausriesen Otto“ oder das US-amerikanische Telekommunikationsunternehmen Verizon.

Skeptiker fragen laut McAfee und Brynjolfsson (2012, S. 63) unter anderem, wo der Beweis dafür ist, dass eine intelligente Nutzung von Big Data zu einer besseren Unternehmensperformance führt. Durch eine Untersuchung fanden sie heraus: „(...) companies in the top third of their industry in the use of data-driven decision making were, on average, 5% more productive and 6% more profitable than their competitors.“ (McAfee und Brynjolfsson 2012, S. 64) Der wirtschaftliche Wert, von dem in Kapitel 2.2.2.1 als Ergebnis von Big Data Methoden gesprochen wurde, charakterisiert sich in der Praxis dann beispielsweise durch „speedier, more personalized promotions“ (McAfee und Brynjolfsson 2012, S. 64-65). Neue Technologien, wie Quantum-Computing oder In-Memory-Datenbanksysteme, machen es möglich, dass die Kundendaten in gänzlich neuen Dimensionen wirtschaftlich verarbeitet und verwaltet werden können, wenn die Technologien im Einklang mit den Geschäftsprozessen sind (Buhl et al. 2013, S. 28). Außerdem ist es nötig, dass sich ein

Unternehmen, welches Datenanalysen verwenden möchte, vorab durch „innovative Geschäftsmodelle den Zugang zu neuen Kundengruppen, deren Plattformen und somit zu neuen Datenquellen“ sichert (Buhl et al. 2013, S. 28).

2.2.3. Charakteristika von Big Data

Der Begriff Big Data wird in unterschiedlichen Disziplinen genutzt, wobei sich die Frage stellt, aus welcher Perspektive der Begriff verwendet wird und welche Formen er annimmt. Während Klein et al. (2013) den Begriff Big Data aus informationstechnischer Sicht durch die Merkmale Volume, Velocity und Variety charakterisieren, bewerten McAfee und Brynjolfsson (2012, S. 62-63) und Kollmann (2019, S. 10-11) die drei Merkmale aus betriebswirtschaftlicher Sicht. Die Charakteristika von Big Data sollen im Folgenden näher erläutert werden.

Mit dem Merkmal Volume werden in diesem Zusammenhang die großen Datenmengen angesprochen, die in der heutigen Zeit zum Beispiel durch soziale Medien generiert werden (Klein et al. 2013, S. 320). Als Beispiel für diese großen Datenansammlungen nennen Klein et al. (2013, S. 320) Facebook. Pro Minute werden in diesem sozialen Medium unter anderem 650.000 unterschiedliche Inhalte veröffentlicht. McAfee und Brynjolfsson (2012, S. 62) halten fest, dass 2012 jeden Tag 2,5 Exabytes an Daten generiert wurden, wobei ein Exabyte einer Billion Gigabyte entspricht. Kollmann (2019, S. 10) spricht in Anlehnung an die BITKOM (2014) zudem von einem stetig wachsenden Umfang dieser Datenmengen.

Das Merkmal Velocity umfasst den Aspekt der Geschwindigkeit. Die großen Datenmengen werden auf der einen Seite sehr schnell produziert, aber müssen auch schnell weiterverarbeitet werden, damit ein Unternehmen möglichst effizient darauf reagieren kann (Klein et al. 2013, S. 320). Diese Schnelligkeit der Verarbeitung der generierten Datenbanken, machen Klein et al. (2013, S. 320) daran deutlich, dass die Verarbeitung im Minuten- oder sogar im Sekundenbereich erfolgen muss. Pro Minute werden laut ihnen zum Beispiel zwei Millionen Suchanfragen bei Google getätigt. Kollmann (2019, S. 11) macht in diesem Kontext in Anlehnung an die BITKOM (2014) deutlich, dass die besondere Herausforderung im Kontext des Merkmals Velocity in der

Notwendigkeit von Echtzeitanalysen und Antworten im Sekundenbereich besteht. McAfee und Brynjolfsson (2012, S. 63) erläutern im Zusammenhang mit dem Merkmal Velocity, dass die Möglichkeit von Echtzeitinformatoren durch die Schnelligkeit in Bezug auf die Daten einen Wettbewerbsvorteil für Unternehmen gegenüber ihrer Konkurrenz ausmachen kann.

Das dritte Merkmal Variety umschreibt die Vielzahl der vorhandenen Daten (Klein et al. 2013, S. 320). Dabei geht es laut Klein et al. (2013, S. 320) aber nicht um strukturierte Daten, sondern um sehr verschiedene und oftmals nicht strukturierte Datenmengen. „Im Rahmen von Big Data werden (...) alle vorhandenen Daten, ob strukturiert oder nicht, zusammengefasst und gemeinsam analysiert. Das hierdurch erzeugte, in sich unstrukturierte Datenkonglomerat kann in drei Kategorien klassifiziert werden.“ (Klein et al. 2013, S. 320) Sie erläutern, dass in der ersten Kategorie Daten vorhanden sind, die sich aus einer Kommunikation zwischen Personen ergeben, wie zum Beispiel in den sozialen Medien (Klein et al. 2013, S. 320). In die zweite Kategorie lassen sich Daten einordnen, die sich aus der Kommunikation zwischen Personen und Diensten oder Maschinen generieren, wie zum Beispiel bei E-Commerce Transaktionen (Klein et al. 2013, S. 320). Drittens gibt es Daten, die zwischen Diensten oder Maschinen entstehen, wie zum Beispiel Überwachungsbilder (Klein et al. 2013, S. 320). McAfee und Brynjolfsson (2012, S. 63) spielen, mit Bezug auf das Merkmal Variety, ebenfalls auf den Aspekt der vielen unterschiedlichen Datenquellen an und erwähnen zudem, dass all diese Quellen relativ neu sind, wie zum Beispiel die sozialen Netzwerke oder mobile Endgeräte. Dies führt laut McAfee und Brynjolfsson (2012, S. 63) zu einem neuen Umgang mit den unstrukturierten Datenmengen, da die bisherigen Datenbanken für die Speicherungen der enormen Mengen ungeeignet sind und so bisher teure und datenintensive Ansätze in der Speicherung und Verarbeitung wirtschaftlich werden. Kollmann (2019, S. 10-11) nimmt in Anlehnung an die BITKOM (2014) eine Unterteilung in vier Oberkategorien für die Daten aus diesen vielfältigen Datenquellen vor. Dabei spricht er von unstrukturierten, semistrukturierten, strukturierten und polystrukturierten Daten (Kollmann 2019, S. 10-11). Die letzte Kategorie der polystrukturierten Daten umfasst zum Beispiel Kombinationen von unternehmens-internen Daten mit Daten aus den sozialen Medien (Kollmann 2019, S. 11).

Klein et al. (2013, S. 321) nennen im Gegensatz zu McAfee und Brynjolfsson (2012) noch ein viertes Merkmal von Big Data: Veracity. Dies umschreibt die Zuverlässigkeit der Daten, welche oftmals unterschiedliche Ursprünge haben und daher nicht zwangsweise zuverlässig und genau sein müssen. Auch die Schnelligkeit der Verarbeitung spielt laut Klein et al. (2013, S. 320) hierbei eine wichtige Rolle, da die Daten oft unter Zeitdruck verarbeitet werden und dadurch nicht ausreichend bereinigen werden können, wodurch sie unsicher und ungenau bleiben. Auf dieses vierte Merkmal geht auch Gentsch (2019, S. 9) in seinen Ausführungen ein und erläutert, dass es sich hierbei um das Merkmal handelt, welches durch Unternehmen noch unzureichend berücksichtigt wird. Als Beispiele für solche Daten, denen es an Glaubwürdigkeit fehlt, nennt er unter anderem Phishing Mails oder Fake News (Gentsch 2019, S. 9). Auch Kollmann (2019, S. 11) spricht in Anlehnung an die BITKOM (2014) ebenfalls noch von einem weiteren Merkmal, das als Analytics bezeichnet wird und auf die unterschiedlichen Analysemethoden zur Aufdeckung „von Mustern, Zusammenhängen und Bedeutungen innerhalb der Datenmengen“ anspielt. Darunter fallen automatisierte „Verfahren der Statistik, Prognosemodelle, Optimierungsalgorithmen, Data Mining, Data Warehousing, Semantik- und Bildanalysen.“ (Kollmann 2019, S. 11).

In der Literaturrecherche zum Thema Big Data wird deutlich, dass eine Definition aus betriebswirtschaftlicher Sicht eher in die Richtung geht, wie die enormen Datenmengen für betriebswirtschaftliche Zwecke genutzt werden können. In dem Artikel von McAfee und Brynjolfsson (2012, S. 62-63) wird erläutert, dass das Merkmal Velocity für einen Wettbewerbsvorteil stehen kann, wenn diese Informationen in Echtzeit generiert werden. Klein et al. (2013, S. 320) nehmen aus informationstechnischer Sicht lediglich darauf Bezug, dass es sich um die Schnelligkeit der Generierung von Datenmengen und die daraus resultierende Forderung nach einer hohen Performance in der Datenverarbeitung handelt. An diesem Beispiel wird deutlich, dass McAfee und Brynjolfsson (2012) bereits einen Schritt weitergehen und die Möglichkeit von Echtzeitinformationen durch Datenverarbeitung auf unternehmerische Zwecke beziehen.

2.2.4. Ziele des strategischen Einsatzes von Big Data im Wertschöpfungsprozess

2.2.4.1. Strategisches Management in Bezug auf Datenanalysen im Bereich des E-Commerce

2.2.4.1.1. Big Data aus Sicht des Market based View

Im Zuge des strategischen Managements stellt sich die Frage, wie Unternehmen Wettbewerbsvorteile gegenüber den Wettbewerbern des Marktes schaffen und dadurch besser performen (Barney und Arikan 2001, S. 124). Die „Industrie 4.0“ wird laut Bischof et al. (2016, S. 143) als die Lösung zur Produktivitätssteigerung und zum Generieren eines einzigartigen Wertversprechens durch die Anwendung innovativer Informationen und Kommunikationstechnologien gesehen. Big Data wird in diesem Zusammenhang als ein zentrales Element dieser Industrie betrachtet (Bischof et al. 2016, S. 143). Für die Strategie eines Unternehmens ist es daher wichtig, sich die Strukturen des Marktes anzuschauen, auf dem das Unternehmen vertreten ist und auf dem Wettbewerbsvorteile erzielt werden sollen. Um die Strukturen innerhalb des Marktes zu verstehen, verweisen Bischof et al. (2016, S. 146-147) auf die „five competitive forces“ nach Porter (2008, S. 83), die in Anhang zwei zu sehen sind und im Folgenden mit Blick auf das Stichwort Big Data betrachtet werden.

Porter (2008, S. 83) erläutert zu der Verhandlungsmacht der Käufer, dass Kunden in der Lage sein können, die Preise zu drücken, bessere Qualität und besseren Service zu fordern und somit die Unternehmen im Wettbewerb gegeneinander auszuspielen. Bei solchen einflussreichen Kunden handelt es sich um Kunden, die eine Verhandlungsmacht, wie das am Wettbewerb teilnehmende Unternehmen, haben (Porter 2008, S. 83). Kunden haben diese Macht in Bereichen, in denen es nur wenige Käufer gibt, in denen Käufer große Produktmengen abnehmen, in denen es um standardisierte Produkte geht, die überall gekauft werden können, und in denen es um Produkte geht, die notfalls auch selbst hergestellt werden können (Porter 2008, S. 83-84). Die Transparenz des Internets sorgt dafür, dass Produktangebote von verschiedenen Herstellern offen gelegt werden, was das Vergleichen von Produktpreisen einfacher gestaltet und darin resultiert, dass die Kunden preissensitiver geworden sind (Porter 2008, S. 84). Big Data hat in diesem Zusammenhang laut Bischof et al. (2016, S. 146)

das Potenzial, die Verhandlungsmacht der Kunden durch Möglichkeiten der Differenzierung anhand von nutzerspezifischen Daten zu verringern. Dies kann laut den Autoren beispielsweise durch individuelle Produktangebote, nutzerspezifische Preismodelle, eine genauere Marktsegmentierung oder eine verbesserte und engere Kundenbeziehung, die die Wechselkosten erhöht, erreicht werden. Allerdings muss laut Bischof et al. (2016, S. 146) auch beachtet werden, dass die Verhandlungsmacht durch vermehrte Transparenz und ein besseres Verständnis der Produktperformance auch ansteigen kann.

Die Rivalität zwischen den bestehenden Wettbewerbern äußert sich laut Porter (2008, S. 85) in Preisrabatten, neuen Produkteinführungen, Werbekampagnen und Leistungsverbesserungen. Je größer diese Rivalität ist, desto geringer ist die Profitabilität. Eine große Rivalität ergibt sich dann, wenn es viele Wettbewerber oder unterschiedlich große Wettbewerber gibt, wenn sich ein Markt nur langsam entwickelt, wenn die Austrittsbarrieren des Marktes hoch sind oder wenn die Bindung an den Markt hoch ist (Porter 2008, S. 85). Zudem kann es zwischen Wettbewerbern desselben Marktes zu einer Preisrivalität kommen, wenn Produkte oder Dienstleistungen ähnlich sind, Fixkosten für die Produkte hoch und Stückkosten niedrig sind, eine Notwendigkeit der Kapazitätserweiterung besteht, die Produkte verderblich sind oder schnell von neuen Innovationen und Informationen überholt werden (Porter 2008, S. 85-86). Eine Rivalität zwischen Wettbewerbern kann laut Porter (2008, S. 86) aber auch positive Auswirkungen haben, wenn jeder Wettbewerber das Ziel verfolgt, ein individuelles Kundensegment mit einem individuellen Mix aus Preis, Produkt, Service, Eigenschaften und Markenidentitäten zu bedienen. In diesem Zusammenhang kann der Einfluss von Big Data dazu führen, dass es weniger um eine Preisrivalität, sondern vielmehr um personalisierte Angebote für spezifischere Marktsegmente geht (Bischof et al. 2016, S. 146). Außerdem kann ein mehrwertschaffender Service dafür sorgen, dass noch genauer differenziert wird und sich das Unternehmen dadurch gegen Rivalität absichern kann (Bischof et al. 2016, S. 146). Dies führt allerdings dazu, dass die Fixkosten der Unternehmen ansteigen, da in neue Technologien investiert werden muss, was wiederum zu einer Rivalität zwischen den Unternehmen bezogen auf die Fixkosten führen kann (Bischof et al. 2016, S. 146). Die Gefahr von substitutiven Gütern und Dienstleistungen ist im Wettbewerb stetig präsent und kann verschiedene Formen annehmen (Porter 2008, S. 84). Ersatzprodukte ver-

ringern nicht nur den Gewinn, sondern können Produkte komplett aus dem Markt verdrängen, wie zum Beispiel im Fall der Einführung von kabellosen Telefonen als Alternative zum Kabeltelefon (Porter 2008, S. 84). Die Gefahr von Ersatzprodukten auf dem Markt ist hoch, wenn das neue Produkt ein besseres Preis-Leistungsverhältnis hat und die Wechselkosten zum neuen Produkt gering sind. Zudem wird durch Ersatzprodukte die Profitabilität der Branche gesenkt, da hierdurch ein Höchstpreis festgelegt wird (Porter 2008, S. 84). Die Auswirkungen von Big Data sind wie bei der Rivalität zwischen Wettbewerbern auch für diesen Bereich nicht genau abzuschätzen (Bischof et al. 2016, S. 146). Einerseits ermöglicht die zunehmende Differenzierungsmöglichkeit ein geringeres Risiko für Ersatzprodukte, da Produkte spezifischer sind. Auf der anderen Seite verschwimmen hierdurch auch die Grenzen zwischen den unterschiedlichen Branchen zunehmend, wodurch vorher zusammenhangslose Produkte als Ersatzprodukte fungieren könnten (Bischof et al. 2016, S. 146).

Zulieferer mit einer gewissen Verhandlungsstärke sind in der Lage, höhere Preise zu fordern, die Qualität und den Service zu limitieren oder die Kosten auf andere Teilnehmer zu verlagern (Porter 2008, S. 82). Ein verhandlungsstarker Zulieferer zeichnet sich unter anderem dadurch aus, dass der Umsatz nicht nur von einer Branche abhängt, dass die Kunden bei einem Zuliefererwechsel hohe Wechselkosten haben, dass der Zulieferer vor allem differenzierte und nicht nur allgemeine Produkte anbietet und dass es keinen Ersatz für das Angebot des Zulieferers gibt (Porter 2008, S. 82-83). In diesem Zusammenhang kann Big Data dafür sorgen, dass sich der bisherige produktzentrierte Ansatz zu einem informations- und servicebasierten Ansatz umwandelt (Bischof et al. 2016, S. 147). Dadurch gewinnen Zulieferer, die diesen Trend verfolgen, an Verhandlungsstärke und die bisherigen Zulieferer verlieren an Stärke (Bischof et al. 2016, S. 147).

Porter (2008, S. 81) erläutert zu der Gefahr von potenziellen neuen Konkurrenten, dass diese das Gewinnpotenzial einer Branche eingrenzt und die Wettbewerber bei einer hohen Gefahr ihre Preise klein halten oder investieren müssen, um die potenziellen Konkurrenten abzuschrecken. Die Gefahr ist höher, je niedriger die Eintrittsbarrieren in den Markt sind (Porter 2008, S.81). Deshalb ist es für ein Unternehmen, dass in einem Markt Fuß fassen möchte, essenziell wichtig, sich vorher über die Eintrittsbarrieren für

den jeweiligen Markt klar zu werden und zu kalkulieren, ob sich der Markteintritt lohnt (Porter 2008, S. 81). Unternehmen, die bereits in einer anderen Branche etabliert sind, haben den Vorteil, dass sie beim Eintritt in einen neuen Markt von bereits bestehenden Kapazitäten profitieren können (Porter 2008, S. 80). Um in den Wettbewerb rund um Daten und Informationen einzusteigen, müssen allerdings Barrieren überwunden werden: „They consist of a complex and hybrid product design incorporating physical and software components, high upfront investments in IT-related infrastructure and skills, and, finally a high level of differentiation.“ (Bischof et al. 2016, S. 147) Neue, auf Big Data basierende, Geschäftsmodelle können für Unternehmen aber auch Markteintrittsbarrieren reduzieren oder ganz aufheben (Bischof et al. 2016, S. 147).

2.2.4.1.2. Big Data aus Sicht des Resourced based View

Nach der genauen Betrachtung von Marktstrukturen im Zusammenhang mit Big Data, ist eine Betrachtung des strategischen Einsatzes der unternehmerischen Ressourcen sinnvoll. Der Resourced based View fokussiert sich auf die internen Ressourcen des Unternehmens, um Stärken und Schwächen innerhalb des Unternehmens aufdecken zu können, da Unternehmen als Bündel verschiedener Ressourcen gesehen werden (Bischof et al. 2016, S. 148). Ressourcen sind laut Barney und Hesterly (2008, S. 74) „the tangible and intangible assets that a firm controls that it can use to conceive of and implement its strategies.“ Barney (1991, S. 101) erläutert zudem in Anlehnung an Daft (1983), dass diese Strategien dann zu einer Verbesserung der Effizienz und der Effektivität des Unternehmens führen. Die benannten Ressourcen sind für Unternehmen im Sinne des strategischen Managements die Möglichkeit, nachhaltige Wettbewerbsvorteile gegenüber der Konkurrenz zu generieren (Barney 1991, S. 99). Auch Bischof et al. (2016, S. 148) halten fest, dass der Resourced based View die strategische Bedeutung der Identifikation, des Managements und des Nutzens von einzigartigen Ressourcen eines Unternehmens zur Generierung von Wettbewerbsvorteilen in den Vordergrund stellt. Wettbewerbsvorteile liegen laut Barney (1991, S. 102) in folgendem Fall vor: „A firm is said to have a competitive advantage when it is implementing a value creating strategy not simultaneously being implemented by any current or potential competitors.“ Ein nachhaltiger Wettbewerbsvorteil liegt zusätzlich dann vor, wenn außerdem für die Konkurrenz auch keine Möglichkeit besteht, diese Strategie in Zukunft nachzuahmen (Barney 1991, S.

102). Allerdings muss mit Blick auf diese zwei Definitionen angemerkt werden, dass als Konkurrenz eines Unternehmens nicht nur aktuelle, sondern auch potenzielle neue Wettbewerber gesehen werden (Barney 1991, S. 102). Zudem merkt Barney (1991, S. 102) an, dass das Aufrechterhalten eines Wettbewerbsvorteils davon abhängt, wie lang es der Konkurrenz nicht möglich ist, diesen nachzubilden. Eine genaue Angabe, wie lang ein Wettbewerbsvorteil anhält ist hierbei aber nicht möglich und ein nachhaltiger Wettbewerbsvorteil muss auch nicht zwangsläufig für immer bestehen bleiben (Barney 1991, S. 102-103).

Barney und Hesterly (2008, S. 74-75) benennen in ihren Ausführungen zusätzlich zu den Ressourcen eines Unternehmens „capabilities“, welche sie als „tangible and intangible assets that enable a firm to take full advantage of the other resources it controls“ beschreiben (Barney und Hesterly 2008, S. 74). Diese Fähigkeiten führen zwar nicht direkt zur Strategiebildung, aber sie ermöglichen es dem Unternehmen Strategien zu bilden, wofür sie als Beispiel Kenntnisse im Bereich des Marketings nennen (Barney und Hesterly 2008, S. 74). Ferner kategorisieren sie die Ressourcen und Fähigkeiten eines Unternehmens in vier Bereiche: Finanzielle, physische, individuelle und organisatorische Ressourcen (Barney und Hesterly 2008, S. 74). Finanzielle Ressourcen umfassen hierbei alle Geldquellen des Unternehmens, die es zur Bildung von Strategien nutzen kann (Barney und Hesterly 2008, S. 74). Physische Ressourcen sind zum Beispiel alle technischen Hilfsmittel und Anlagen, gelagerte Materialien oder der Standort eines Unternehmens (Barney und Hesterly 2008, S. 74). Zu den individuellen Ressourcen zählen die Mitarbeiter des Unternehmens und ihre Erfahrung, Intelligenz, Beziehungen, Bewertung, Einsicht und Fortbildung, wobei sowohl der Topmanager als auch der normale Mitarbeiter eine wichtige menschliche Ressource des Unternehmens bilden kann (Barney und Hesterly 2008, S. 74-75). Organisatorische Ressourcen sind Gruppen aus Individuen, die durch Barney und Hesterly (2008, S. 75) wie folgt definiert werden: „Organizational resources include a firm’s formal reporting structure; its formal and informal planning, controlling, and coordinating systems; its culture and reputation; as well as informal relations among groups within a firm and between a firm and those in its environment.“

Strategische Ressourcen haben laut Barney (1991, S. 105-106) vier Charakteristika: Diese müssen wertschaffend für das Unternehmen, selten in der Wettbewerbslandschaft des Unternehmens, nicht durch die Konkurrenz imitierbar und nicht ersetzbar durch strategisch gleiche Produkte für diese Ressource sein. Bischof et al. (2016, S. 148-150) greifen diese vier Charakteristika auf und setzen sie in den Kontext von Big Data. Für den IT-Bereich ist es schwierig den Wert für die dadurch entstehenden Möglichkeiten der Informationstechnik für die strategische Positionierung abzuschätzen (Bischof et al. 2016,

S. 149). Hierzu gehört auch der Bereich rund um das Schlagwort Big Data, wobei Bischof et al. (2016, S. 149) äußern, dass Unternehmen, die Big Data Methoden nutzen, ihre Wettbewerber übertreffen werden. Außerdem machen Bischof et al. (2016, S. 149) deutlich, dass der Einsatz von Big Data als strategische Ressource Investitionen in den sogenannten „Big Data Stack“ erfordert. Hierzu gehören „modified hardware, specific database technology and innovative software applications in order to store, process and analyze the data.“ (Bischof et al. 2016, S. 149). In diesem Zusammenhang wird laut Bischof et al. (2016, S. 149) deutlich, dass es im Kontext von Big Data durch leicht zugängliche und standardisierte Hard- und Software-Applikationen schwer wird, Wettbewerbsvorteile zu bilden. Rückblickend auf die vier Charakteristika einer strategischen Ressource nach Barney (1991, S. 105-106) erfüllen diese Big Data Technologien somit nicht die Forderung danach, dass sie selten, nicht ersetzbar und nicht nachahmbar sind (Bischof et al. 2016, S. 149). Sie können lediglich dafür sorgen, dass andere strategische Ressourcen besser und schneller eingesetzt werden können und dadurch die Performance verbessern (Bischof et al. 2016, S. 149). Um den gewünschten strategischen Nutzen im Kontext von Big Data zu erlangen schlagen Bischof et al. (2016, S. 149) Folgendes vor: „The major strategic leverage from Big Data following the rationale of the Resource-Based View can be realized by integrating Big Data with other resources of the company, like IT systems (e.g. ERP or CRM systems), products, processes, and personal skills.“ Dadurch bietet sich die Möglichkeit, dass Big Data in Kombination mit strategischen Ressourcen des Unternehmens zu einer Ressource im Sinne des Resourced based View wird, die die vier Kriterien nach Barney (1991, S. 105-106) erfüllt. Durch diese möglichen Kombinationen können Ressourcen entstehen, die so viele Datenzusammenhänge abbilden und so komplex sind, dass sie für Wettbewerber nicht nachvollziehbar und dadurch nicht nachahmbar sind (Bischof et al.

2016, S. 149). Außerdem führt zum Beispiel eine Zusammenführung von Produkten und Daten dazu, dass neue einzigartige hybride Ressourcen entstehen, die dadurch einen Seltenheitswert haben (Bischof et al. 2016, S. 149). Des Weiteren sind diese Kombinationen wertschaffend für ein Unternehmen, da beispielsweise Vorhersagen zur Instandhaltung generiert werden können, was Wartungskosten und die Gefahr von Produktionsausfällen minimieren kann (Bischof et al. 2016, S. 149). Daten über das Kundennutzungsverhalten können es dem Unternehmen zusätzlich erleichtern, strategische Chancen zu nutzen und eine starke strategische Position gegenüber den Wettbewerbern über einen längeren Zeitraum zu halten (Bischof et al. 2016, S. 149). Die Kombination von Big Data mit anderen Ressourcen ist zwar sehr spezifisch, Wettbewerber könnten allerdings eine ähnliche Strategie für ihre Datensammlungen nutzen (Bischof et al. 2016, S. 149). Da es sich aber um kundenspezifische Daten handelt, müsste ein Wettbewerber aber auch die zugrundeliegende Strategie des Unternehmens oder auch das Geschäftsmodell übernehmen, was die Kombination quasi nicht ersetzbar macht (Bischof et al. 2016, S. 149). Es lässt sich demnach festhalten, dass Big Data als eine strategische Ressource Unternehmen mehrere Möglichkeiten bietet, um Wettbewerbsvorteile durch die internen unternehmerischen Ressourcen zu generieren oder auch um neue Märkte zu erschließen, beziehungsweise sich auf bestehenden Märkten zu positionieren (siehe Kapitel 2.2.4.1.1).

2.2.4.2. Risikomanagement in Bezug auf Datenanalysen im Bereich des E-Commerce

Ziel eines unternehmerischen Risikomanagements ist der Aufbau eines Systems, das es Unternehmen ermöglicht, Risiko und Unsicherheiten zu bewältigen (Dionne 2013, S. 154). Dabei macht Dionne (2013, S. 154) deutlich, dass Risiken in nahezu jeder unternehmerischen Aktivität präsent sind und deren Identifikation, Beurteilung und Handhabung Teil der strategischen Entwicklung des Unternehmens und an höchster Stelle des Unternehmens angesiedelt sind. „An integrated risk management approach must evaluate, control, and monitor all risks and their dependences to which the company is exposed.“ (Dionne 2013, S. 154) Ein Risiko, wie es in dem Zitat angesprochen wird, charakterisiert sich durch die Kombination der Auftretenswahrscheinlichkeit eines Ereignisses und der darauf resultierenden Folgen,

die meistens negativ sind (Dionne 2013, S. 154). Eine Unsicherheit hingegen ist schwieriger zu greifen, da hierbei nicht klar ist, wie wahrscheinlich das Ereignis auftritt und welche Folgen es hat (Dionne 2013, S. 154). Deshalb ist es für Unternehmen wichtig, sowohl vorsorgliche Maßnahmen für die Minimierung von Unsicherheiten, als auch präventive Maßnahmen zur Eindämmung von Risiken zu treffen (Dionne 2013, S. 154). Die Hauptaufgaben des Risikomanagements sind „diversification and risk hedging using various instruments, including derivatives and structured products, market insurance, self-insurance, and self-protection.“ (Dionne 2013, S. 154).

Bannerman (2008, S. 2118) macht deutlich, dass Software-Projekte sehr riskante Aktivitäten sind und dass das Risikomanagement in diesem Bereich die Möglichkeit bietet, die Ergebnisse der Projekte zu verbessern. Auch er verweist auf die Definition, dass sich ein Risiko dadurch ableitet, wie hoch die Wahrscheinlichkeit ist, dass ein nicht erwünschtes Ereignis auftritt und welche Folgen dann aus diesem resultieren (Bannerman 2008, S. 2119). Dabei werden die Folgen meist mit etwas Negativem assoziiert und Bannerman (2008, S. 2119) erläutert, dass potenzielle Risikofaktoren bei Software Projekten im Allgemeinen bereits vor Beginn des Projects identifiziert werden sollten. Dadurch können die grössten Risiken während des Projektverlaufs im Auge behalten werden, um die Wahrscheinlichkeit des Auftretens dieser Risiken zu minimieren (Bannerman 2008, S. 2119). Allerdings merkt Bannerman (2008, S. 2119) auch an, dass es vier Einschränkungen dieser Sichtweise auf Risiken gibt: Erstens orientiert sich diese Sichtweise nicht an den realen Verhaltensweisen von Managern, da diese das Risiko weniger präzise betrachten und glauben, sie könnten Risiken beispielsweise allein durch ihre Fähigkeiten abwehren (Bannerman 2008, S. 2119). Die zweite Einschränkung ist, dass ein Abschätzen der Auswirkungen vieler Risikofaktoren in der Praxis, vor allem bei Software Projekten, sehr schwierig ist und daher oftmals eher von Unsicherheiten gesprochen wird (Bannerman 2008, S. 2119-2120). Drittens wird das Risikoereignis in ein Abhängigkeitsverhältnis zu den Risikofolgen gesetzt, was die unternehmenseigenen Möglichkeiten und Kapazitäten außer Acht lässt (Bannerman 2008, S. 2120). Die vierte Einschränkung sieht Bannerman (2008, S. 2120) darin, dass in der Definition lediglich die Risiken angesprochen werden, die das Unternehmen abschätzen kann. Dabei sind die Risiken, die nicht abgeschätzt werden können und die dadurch die Vorbereitung von Gegenmaßnahmen für dieses Risiko unmöglich machen,

für das Unternehmen die viel größeren Gefahren. (Bannerman 2008, S. 2120). Das Risikomanagement für Software Projekte definiert sich laut Bannermann (2008, S. 2120) in Anlehnung an Boehm (1989, 1991), Charette (1989) und Kerzner (2003) durch „a set of principles and practices aimed at identifying, analyzing and handling risk factors to improve the chances of achieving a successful project outcome and/or avoid failure.“

Gerade bei Softwaretechnologien, die für Datenanalysen genutzt werden, können viele Risiken vorher nicht abgeschätzt werden (Bannerman 2008, 2118). Wird als Beispiel ML genommen, dann wird deutlich, dass in diesem Zusammenhang ein Abschätzen des Risikos enorm schwierig ist. Da diese Technologien fähig sind, ihre Performance stetig zu verbessern, ohne dass Menschen ihr neue Aufgaben erklären müssen (McAfee und Brynjolfsson 2017, S. 4), kann die zukünftige Entwicklung dieser Technologie nicht genau abgeschätzt werden. Ein Risikomanagement in diesem Bereich zu etablieren, stellt das Unternehmen somit vor ganz neue Herausforderungen, die es zu bewältigen gilt. Die Literaturrecherche hat gezeigt, dass es beim Risikomanagement auch darum geht, Fehler zu vermeiden und sich Fehlerquellen, die Risiken bergen, bewusst zu machen. Hierdurch ist die Fehlerprävention als Teil des Risikomanagements von nicht zu unterschätzender Bedeutung für das Unternehmen und seinen Erfolg.

2.2.5. Möglichkeiten und Notwendigkeit eines gezielten

Datenmanagements im E-Commerce

Im Zeitalter der Digitalisierung, wo virtuelle Kontakte normal sind und sich jeder über seine individuelle IP-Adresse in ein digitales Datennetz einwählen kann, übernimmt der Teilnehmende eine aktive Rolle im digitalen Kommunikationsaustausch (Kollmann 2019, S. 46). Hierbei besteht laut Kollmann (2019, S. 46) die Möglichkeit, dass jeder Teilnehmende dabei sowohl in der Lage ist Informationen zu erstellen als auch Informationen abzurufen. Kollmann (2019, S. 46) nutzt hierfür den Begriff der „Interaktivität“, der unter anderem das kooperative Agieren und die wechselseitige Kommunikation umfasst. Durch diese Interaktivität entsteht für alle Teilnehmenden im Verlauf des Kommunikationsaustausches die Möglichkeit der individuellen Reaktion und Aktion auf die anderen Teilnehmenden (Kollmann 2019, S. 46). Welchen Grad diese Interaktivität zwischen den Kommunikationspartnern annimmt, ist allerdings abhängig

von den durch die Software zugelassenen Möglichkeiten der Interaktion und ob eine Online- und Offline-Technologie verwendet wird (Kollmann 2019, S. 46). Diese Veränderung hin zur interaktiven Kommunikation bietet laut Kollmann (2019, S. 46) große Möglichkeiten für wirtschaftliche Aktivitäten, da es meistens um aktive Einzeltransaktionen geht, in der jeder Anwender eine eigenständige Informationsadresse besitzt. Dadurch ergibt sich die Möglichkeit der Selektion und Ansteuerung des Einzelnen, die sich von der bisherigen Massenansprache absetzt und die Wirkung einer Kommunikation erhöht (Kollmann 2019, S. 46). Somit ermöglicht das Internet nicht nur die Interaktivität, sondern auch eine Individualisierung und Personalisierung des Angebots für die Kunden (Kollmann 2019, S. 48). Im Vergleich zu Offline-Kanälen bietet die digitale Kommunikation laut Kollmann (2019, S. 48) zudem den Vorteil, dass „sämtliche Bewegungen, Transaktionen und Informationen der Nutzer in Form von digitalen Daten gespeichert werden können“ und die Anbieter dadurch direkt auf einen Kunden reagieren können, wenn dieser beispielsweise die Internetseite des Unternehmens verlässt, ohne etwas zu kaufen, sich vorher aber ein bestimmtes Produkt angeschaut hat. Dabei ist anzumerken, dass die Personalisierung umso besser ist, je mehr die Teilnehmenden miteinander interagieren und je mehr preisgegeben wird, da ein höherer Datenbestand ein höheres Individualisierungs- und Personalisierungspotenzial für beispielsweise Werbemaßnahmen bedeutet (Kollmann 2019, S. 48).

In der Realität äußert sich die Interaktivität zwischen den Kommunikationsteilnehmern beispielsweise wie folgt: Ein Arbeiter bei LinkedIn, Jonathan Goldman, war so interessiert an den produzierten und bis dahin ungenutzten Datenmengen, dass er begann sich die Beziehungen der Menschen bei LinkedIn anzuschauen und so Potenziale für eine Verwendung dieses Wissens zu sehen (Davenport und Patil 2012, S. 72). Eine Möglichkeit, der er nachging, war die Funktion, dass Nutzern Menschen angezeigt wurden, die sie vermutlich aufgrund ihrer Schulzeit oder ihres Arbeitslebens kennen könnten (Davenport und Patil 2012, S. 72). Allein für dieses Feature, das von Goldman entwickelt und bald zum Standardfeature wurde, war die Klickrate 30 Prozent höher als für andere Features, die der Onlinedienst bietet (Davenport und Patil 2012, S. 72).

Diese Personalisierung eines Angebots für den Kunden kann dabei helfen, die Interaktivität zu steigern, da sie es ermöglicht, ein grundlegendes Bedürfnis des Kunden nach Individualität und nutzerbezogenen Informationen abzudecken (Kollmann 2019, S. 48). In seinen Ausführungen spricht Kollmann von einer „One-to-One-Beziehung“ oder einer „One-to-All-Beziehung“ zwischen der Handelsebene und den Kunden, die entweder direkt oder indirekt und virtuell oder real sein können (Kollmann 2019, S. 49). In eine direkte reale One-to-One-Beziehung ist kein weiteres Subjekt involviert, sondern lediglich zwei Kommunikationspartner, die sich an demselben Ort befinden (Kollmann 2019, S. 49). Bei der indirekten realen One-to-All-Beziehung geht es um eine Massenansprache von erreichbaren Marktteilnehmern, welche im Fokus der Kommunikationsaktivität stehen (Kollmann 2019, S. 49). In einer indirekten virtuellen One-to-All-Beziehung geht es um eine virtuelle Kommunikation an eine breite Masse, wie zum Beispiel bei einer Werbung in einem Onlineshop über Werbebanner (Kollmann 2019, S. 49). Die direkte virtuelle One-to-One-Beziehung wird durch die individuellen Netzadressen der Netzteilnehmer möglich, da jeder Teilnehmer individuell über beispielsweise seine E-Mail-Adresse angesprochen werden kann (Kollmann 2019, S. 49). Durch die Möglichkeiten des Internets beschränkt sich die Kommunikation hierbei nicht auf einen bestimmten Personenkreis, sondern bietet eine globale Reichweite und ist dabei unabhängig von Raum und Zeit möglich (Kollmann 2019, S. 49). Um diesen Bereich der direkten virtuellen One-to-One-Beziehung und die damit verbundenen Möglichkeiten der individuellen Personalisierung geht es, wenn in dieser Arbeit über Datenanalysen im Bereich des E-Commerce gesprochen wird. Eine Möglichkeit für eine Nutzung von generierten Informationen aus gesammelten Datenmengen im E-Commerce ist das sogenannte Retargeting: „Firms can now offer personalized recommendations to consumers who return to their website, using consumers’ previous browsing history on that website.“ (Lambrecht und Tucker 2013, S. 561) Dies äußert sich in der Praxis dann beispielsweise so, dass den Nutzern Werbung für Produkte im Internet angezeigt wird, die diese sich kurz vorher in einem Onlineshop oder auf einer Webseite angeschaut haben (Lambrecht und Tucker 2013, S. 561).

2.3. Grundlagen und Kennzahlen von Geschäftsmodellen im E-Commerce

2.3.1. Grundlagen und Kennzahlen von Geschäftsmodellen des E-Commerce

Durch die zunehmende Nutzung des Internets und damit verbunden durch die zunehmenden betriebswirtschaftlichen Aktivitäten über das Internet als neuen Vertriebskanal, ist dieses Feld für die Betriebswirtschaft sowohl in der Wissenschaft als auch in der Praxis stetig attraktiver geworden (Wirtz und Becker 2002, S. 85). Die Folge davon ist, dass sich durch neue Geschäftsideen für diesen Kanal auch neue Geschäftsmodelle ergeben, da die traditionellen Modelle der Realwirtschaft nicht problemlos übertragen werden können (Wirtz und Becker 2002, S. 85). Eine gute Idee für ein neues Geschäft ist demnach nur dann erfolgversprechend, wenn die Investoren und potenziellen Neukunden durch die Innovation der Idee und ein gut durchdachtes und nachhaltiges Geschäftsmodell überzeugt werden (Wirtz und Becker 2002, S. 85). Ein Geschäftsmodell zeichnet sich laut Wirtz und Becker (2002, S. 85) durch Folgendes aus: „Das Geschäftsmodell charakterisiert, welche externen Ressourcen in die Unternehmung fließen und wie diese durch den innerbetrieblichen Leistungserstellungsprozess in vermarktungsfähige Informationen, Produkte und/oder Dienstleistungen transformiert werden.“ Demnach bildet es neben dem Wertversprechen auch die Wertschöpfungsarchitektur, durch die der Wert geschaffen werden soll, und das Erlösmodell des Unternehmens ab. Auch Becker (2019, S. 18) erläutert, dass ein Geschäftsmodell „das Grundprinzip, nach dem Unternehmen Wertschöpfung betreiben“ beschreibt, um so die langfristige Unternehmensexistenz zu sichern.

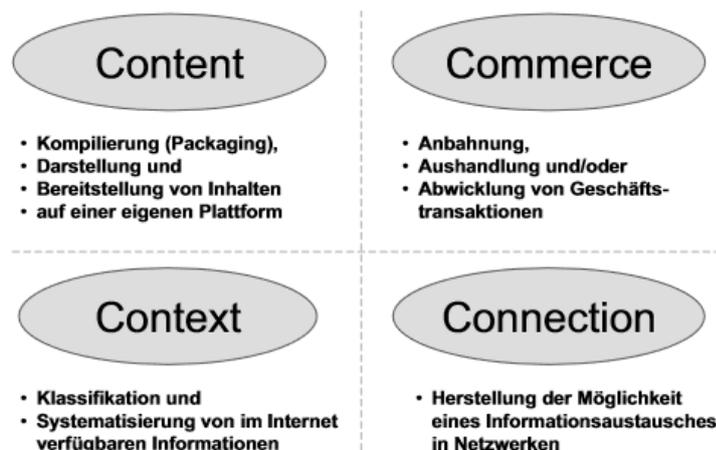


Abbildung 4: Basisgeschäftsmodelle des 4C-Net-Business-Modell in der Internetökonomie (Wirtz und Becker 2002)

Wirtz und Becker (2002, S. 85-86) erläutern in ihren Ausführungen, dass es bei einer Geschäftsmodelltypologie anhand spezifischer Charakteristika zu Überlappungen zwischen den verschiedenen Typen kommen kann, da die Grenzen nicht immer trennscharf sind. Jeder Typus sollte an sich homogen und bezogen auf die anderen Typen heterogen sein, um eine Differenzierung, Orientierung und Klassifizierung überhaupt möglich zu machen (Wirtz und Becker 2002, S. 86). Für das von ihnen gebildete 4C-Net-Business-Modell haben sie die Geschäftsmodelle anhand des Kriteriums „Leistungsangebot“ voneinander abgegrenzt. Hierdurch ist gegeben, dass innerhalb eines Typus ähnliche Leistungs- und Wertschöpfungsprozesse der Geschäftsmodelle vorhanden sind (Wirtz und Becker 2002, S. 86). Das 4C-Net-Business-Modell ist in Abbildung vier zu sehen und umfasst die vier grundlegenden Geschäftsmodelltypen Content, Commerce, Context und Connection. Zusätzlich zu dieser Typisierung, gibt es für jeden Geschäftsmodelltypus mehrere Geschäftsmodellvarianten, wodurch laut Wirtz und Becker (2002, S. 86) die Möglichkeit besteht, alle bestehenden Geschäftsmodelle der Internetökonomie einordnen und systematisieren zu können.

Im Folgenden soll kurz auf die einzelnen Geschäftsmodelltypen und ihre jeweiligen Varianten eingegangen werden. Der Geschäftsmodelltyp Content umfasst, das kompilieren, darstellen und bereitstellen von Inhalten auf einer eigenen Plattform (Abb. 4: Basisgeschäftsmodelle des 4C-Net-Business-Modell in der Internetökonomie (Wirtz und Becker 2002)). Ziel ist es laut Wirtz und Becker (2002, S. 86) „den Nutzern Inhalte einfach, bequem, visuell ansprechend aufzubereiten und online zugänglich zu machen.“

Unterteilt wird dieser Typus in die Varianten E-Information, E-Entertainment und E-Education, wobei E-Information und E-Entertainment zusätzlich eine hybride Variante bilden können, die E-Infotainment genannt wird (Wirtz und Becker 2002, S. 86-87). Ein Beispiel für einen solchen Geschäftsmodelltyp, welches der Variante des E-Infotainment zugeordnet werden kann, wäre laut Wirtz und Becker (2002, S. 86) www.kicker.de. Der zweite Geschäftsmodelltyp Commerce beschreibt die Anbahnung, Aushandlung und/oder Abwicklung von Geschäftstransaktionen (Abb. 4: Basisgeschäftsmodelle des 4C-Net-Business-Model in der Internetökonomie (Wirtz und Becker 2002)). Dieser Modelltyp wird von Wirtz und Becker (2002, S. 87-88) in die Varianten Attraction (Anbahnung), Bargaining/Negotiation (Aushandlung) und Transaction (Abwicklung) unterteilt. Ein Beispiel für einen Geschäftsmodelltypen der Variante Aushandlung wäre www.ebay.com (Wirtz und Becker 2002, S. 87). Der Geschäftsmodelltyp Context umschreibt Modelle, die der Klassifikation und Systematisierung von im Internet verfügbaren Informationen dienen (Abb. 4: Basisgeschäftsmodelle des 4C-Net-Business-Model in der Internetökonomie (Wirtz und Becker 2002)). Aufgeteilt wird der Typ Context von Wirtz und Becker (2002, S. 88-89) in die Varianten Suchmaschinen und Web-Kataloge. Ein allgemein bekanntes Beispiel einer solchen Suchmaschine wäre www.google.com (Wirtz und Becker 2002, S. 89). Der letzte Geschäftsmodelltyp von Wirtz und Becker (2002, S. 89) wurde von ihnen Connection genannt und bietet die Möglichkeit eines Informationsaustausches über Netzwerke. Hierunter fallen laut ihnen die Varianten der Intra- und der Inter-Connection, wobei www.hotmail.com als Mailing Service beispielsweise unter die Variante der Intra-Connection fallen würde, da dies das Angebot einer kommunikativen Dienstleistung innerhalb des Internets ist (Wirtz und Becker 2002, S. 89-90).

Auch Picot et al. (2001, S. 19) erläutern, dass sich aus der Entwicklung von E-Business und E-Commerce Veränderungen in den traditionellen Geschäftsprozessen zwischen dem Unternehmen und dem Kunden ergeben. Die Neuerung bezieht sich hierbei auf eine „stufenweise Integration“ des Kunden durch Möglichkeiten wie der Präsentation des Unternehmens bezogen auf spezifische Zielgruppen, elektronischen Produktkatalogen mit inbegriffenen Unterstützungssystemen in der Entscheidungsfindung, dem Online-Shopping und einem elektronisch unterstütztem After-Sales-Service (Picot et al. 2001, S. 19). Die großen Potenziale des E-Commerce

liegen hierbei demnach in der elektronischen Geschäftsabwicklung (Picot et al. 2001, S. 20). Die Gründe für das Wachstum und die steigende Beliebtheit des E-Commerce sehen Picot et al. (2001, S. 20) in der zunehmenden Nutzung des Internets, der besseren Bedienbarkeit von Internetseiten und in den sinkenden Kosten für einen Internetzugang. Hierdurch wächst auch das Interesse an Geschäftsmodellen des E-Commerce stetig, da diese durch die elektronische Unterstützung die Abwicklung eines Geschäfts beschleunigen und Möglichkeiten der verbesserten Kundenbindung, der Differenzierung von der Konkurrenz und das Herausstellen von Wettbewerbsvorteilen ermöglichen (Picot et al. 2001, S. 20). Anhang drei zeigt mögliche Realisierungsformen des elektronischen Handels nach Picot et al. (2001, S. 21). Unterschieden wird hier anhand von vier Ebenen, die unabhängig von der Form des Handels sind und somit den Übergang vom traditionellen Handel (conventional commerce) zum elektronischen Handel illustrieren (Picot et al. 2001, S. 21). Für diese Masterarbeit sind die Bereiche des indirect und direct E-Commerce von Bedeutung.

2.3.2. Grundlagen und Kennzahlen datenzentrischer Geschäftsmodelle

Big Data und die damit verbundenen Möglichkeiten der Datenverarbeitung und einer wirtschaftlichen Nutzung des Datenbestandes legen den Grundstein für eine Entwicklung neuer Geschäftsmodelle vor dem Hintergrund der Internetökonomie (Dorfer 2018, S. 20). Vorhandene Daten werden bei diesen von Dorfer (2016, S. 308) als datenzentrische Geschäftsmodelle bezeichneten Modellen in neuer Kombination für die Wertschöpfung des Unternehmens genutzt. Unter diese Geschäftsmodelle fallen unter anderem Social Networking Sites, Suchmaschinen oder auch Bewertungsplattformen, da diese die Daten verschiedener Quellen sammeln, verarbeiten und wertschaffend einsetzen (Dorfer 2016, S. 308). Laut Dorfer (2016, S. 315) sind datenzentrische Geschäftsmodelle als Teilmenge von E-Business-Geschäftsmodellen interpretierbar, weil „kommerziell orientierte Akteure eine Marktleistung vollständig oder weit überwiegend mit Internet-Technologien realisieren.“ Die angesprochenen Geschäftsmodelle haben eine intermediäre Funktion, da sie den Endnutzern meist kostenfrei datenzentrische Dienste oder Produkte anbieten, um hierdurch an Nutzerdaten zu kommen, die dann Unternehmenskunden zum Zweck personalisierter Werbung oder personalisierter

Produktangebote angeboten werden und die Einnahmequelle dieser Geschäftsmodelle bilden (Dorfer 2018, S. 25). Diese Modelle funktionieren demnach durch eine Umwandlung von Daten in Informationen, wobei die Daten als Produktionsfaktor angesehen werden (Dorfer 2018, S. 25). Durch das große Potenzial, welches datenzentrische Geschäftsmodelle beispielsweise durch neue Wertschöpfungsprinzipien bieten, handelt es sich hierbei laut Dorfer (2016, S. 308) um ein Thema, welches sowohl wirtschaftswissenschaftlich als auch unternehmenspraktisch sehr relevant ist. Allerdings finden sich in der wissenschaftlichen Literatur eher Beiträge zu spezifischen Unternehmen mit Blick auf die zugehörige Datensammlung, -nutzung und -monetarisierung (Dorfer 2016, S. 308). Zudem merkt Dorfer (2018, S. 25) an, dass in ihrer Synopse zu vorhandenen Konzeptionen des beschriebenen Geschäftsmodelltypus deutlich wird, dass die Forschungsarbeiten zu diesem Geschäftsmodelltypus entweder den Terminus „Daten“ oder „Informationen“ in den Vordergrund stellen, obwohl derselbe Typus beschrieben wird. Hieraus schließt sie, dass sich Konzeptionen, die den Begriff „Daten“ in den Vordergrund stellen, eher auf den Wertschöpfungsprozess und somit auf die Verarbeitung der Daten fokussieren. Bei einer Verwendung des Terminus „Informationen“ orientiert sich die Konzeption laut Dorfer (2018, S. 25) eher an dem Output des Wertschöpfungsprozesses und somit an dem durch die Datenverarbeitung geschaffenen Wert für den Kunden. Eine Begriffssystematisierung von Daten und Informationen wurde in Kapitel 2.2.1 genauer durchgeführt, wo die Begriffe voneinander abgegrenzt wurden.

Dorfer (2016, S. 309) strebt in ihrer Untersuchung an, „datenzentrische Geschäftsmodelle im Sinne eines deskriptiven Wissenschaftsziels zu erfassen und präzise zu beschreiben.“ So stellt sie bezogen auf die Merkmale datenzentrischer Geschäftsmodelle die folgenden drei Merkmale heraus (Dorfer 2016, S. 316): Das erste Merkmal ist ein „Angebot einer kommerziellen Marktleistung, die vollständig oder weit überwiegend über Internettechnologien realisiert wird.“ Als zweites Merkmal erläutert Dorfer (2016, S. 316), dass die Datenintermediation als Leistungsangebot im Zentrum der Geschäftsmodelle steht. Als drittes Merkmal nennt Dorfer (2016, S. 316-317) „symbiotische Marktbeziehungen zwischen Endnutzer- sowie Unternehmenskundenmärkten bzw. Privathändlermärkten“, da der Nutzen eines Teilnehmers auf einem Markt steigt, wenn sich bei einem angekoppelten Markt die Datenmenge erhöht.

Anhang vier bildet die von Dorfer (2016) gebildete Klassifikation datenzentrischer Geschäftsmodelle ab, wobei anzumerken ist, dass die Geschäftsmodelle zur sozial-interaktiven Stiftung kognitiven Nutzens in der Mitte stehen, da diese einen hybriden Charakter vorweisen (Dorfer 2016, S. 327).

Wie in Kapitel 2.3.1 erklärt, bildet ein Geschäftsmodell das Wertversprechen, die Wertschöpfungsarchitektur und das Erlösmodell eines Unternehmens ab, worauf in diesem Absatz mit Blick auf datenzentrische Geschäftsmodelle kurz eingegangen werden soll. Laut Dorfer (2016, S. 329) besteht das Wertversprechen darin, dass die Kundengruppen der Nutzer und die Kundengruppen der Unternehmenskunden beziehungsweise der Privathändler durch den beschriebenen Geschäftsmodelltypus Effizienz- und Effektivitätsvorteile aufgrund des Datenmanagements gewinnen. Die Nutzer haben hierdurch zum Beispiel ein „Mittel, um über die Produktion und Rezeption von Daten ihre kognitiven und sozial-interaktiven Bedürfnisse zu befriedigen“ (Dorfer 2016, S. 329), während die Unternehmenskunden und Privathändler Unterstützung im Rahmen ihres datenbasierten Marketings erhalten. Zur Wertschöpfungsarchitektur von datenzentrischen Geschäftsmodellen erläutert Dorfer (2016, S. 330-331), dass am Schluss des Wertschöpfungsprozesses Datengüter als Marktleistung stehen und die Bedürfnisse der Akteure befriedigen sollen. Außerdem klassifiziert sie Datengüter als Verbrauchsgüter, deren Wert durch jeden erneuten Konsum sinkt. Für die Kundengruppe der Nutzer sind diese Güter zudem als Konsumgüter zu sehen, während es sich für die Kundengruppe der Unternehmenskunden und Privathändler um Investitionsgüter für die Leistungserstellung handelt (Dorfer 2016, S. 331). Im weiteren Verlauf ihrer Untersuchung unterscheidet sie zwischen zwei Grundtypen der Wertschöpfung datenzentrischer Geschäftsmodelle: Dem Modell der direkten und der indirekten Transaktion (Dorfer 2016, S. 331-333). Bei der direkten Transaktion erfolgen keine parallelen Transaktionen, sondern es erfolgt eine „unmittelbare Datenintermediation zwischen den Nachfragern und den zahlungspflichtigen Anbietern“ (Dorfer 2016, S. 331). Bei dem Modell der indirekten Transaktion ist ein Unternehmen an die Transaktion gekoppelt, welches Daten sammelt, die bei der Interaktion vom Nutzer preisgegeben werden und dadurch datenbasierte Marketingleistungen erstellen kann, wodurch dieses Modell für den privaten Nutzer meist kostenfrei ist (Dorfer 2016, S. 332-333). Um die angesprochenen Datengüter für die Kunden zu generieren, „erbringen

datenzentrische Geschäftsmodelle eine dynamische, intermediäre Wertschöpfung“ (Dorfer 2016, S. 333). Dorfer (2016, S. 334) bezieht sich in diesem Zusammenhang auf den elektronischen Wertschöpfungsprozess nach Kollmann, welcher für diese Masterarbeit ebenfalls die Grundlage der Analyse in Kapitel drei bildet und der in Kapitel 3.1 genau beschrieben wird (Kollmann 2019). Zum Erlösmodell der datenzentrischen Geschäftsmodelle merkt Dorfer (2016, S. 335) an, dass es sich hierbei um „vorwiegend indirekte, transaktions(un)abhängige Erlöse unter Ausnutzung der symbiotischen Marktbeziehungen“ handelt. Das Konsumgut für den Nutzer ist somit kostenlos, während das Investitionsgut für den Privathandel oder Unternehmenskunden kostenpflichtig ist (Dorfer 2016, S. 335-336). Investitionsgüter sind hierbei aus dem Bereich Werbung, Marktforschung und Sales und werden refinanziert durch Werbe-, Kommissions- und Data-Mining-Erlöse und Einstellgebühren (Dorfer 2016, S. 336).

2.4. Grundlagen, Kennzahlen und Einordnung des E-Commerce in Bezug auf Datenanalysen

2012 lag der weltweite Umsatz des Onlinehandels erstmals bei einer Billion US-Dollar (Walsh et al. 2013, S. 49; Große Holtforth 2017, S. 4) und für das Jahr 2018 wurden ein Umsatz von drei Billionen US-Dollar für den weltweiten Onlinehandel prognostiziert (Große Holtforth 2017, S. 4). Der Bundesverband für E-Commerce und Versandhandel Deutschland e.V. gibt in einer Pressemitteilung vom 22.04.2019 bekannt, dass die E-Commerce Umsätze für das erste Quartal 2019 die Erwartungen übertroffen haben und mit 11,2 Prozent über dem Vorjahreswert des ersten Quartals 2018 von 10,6 Prozent liegen (bevh 2019, S. 1). Die deutschen Kunden kauften in Onlineshops von Januar bis März 2019 somit Waren im Wert von 16.244 Millionen Euro inklusive Umsatzsteuer (bevh 2019, S.1). Durch die stetig wachsenden E-Commerce Transaktionen steigt auch die Masse an produzierten Daten und damit die Forderung nach einem wirtschaftlichen Nutzen dieser Datenmengen, beispielsweise durch Datenanalysen, was die Aktualität der Thematik dieser Masterarbeit nochmals hervorhebt.

E-Commerce umfasst laut Turban et al. (2018, S. 7) den Handel, Transport, Verkauf oder Kauf von Dienstleistungen, Gütern oder Daten über Netzwerke, wie das Internet.

Große Holtforth (2017, S. 4) definiert E-Commerce als einen „Absatzkanal, der auf digitalen-vernetzten Handelsplattformen den Abschluss von Kaufverträgen ermöglicht.“ Der Begriff des E-Commerce ist abzugrenzen von dem Begriff des E-Business, der über die genannte Definition des E-Commerce hinausgeht (Turban et al. 2018, S. 7). Zum E-Business zählen alle Business-Aktivitäten, die via Internet gemacht werden (Turban et al. 2018, S. 7). Für Picot et al. (2001, S. 19) geht es beim E-Business um eine unternehmensprozessorientierte Betrachtung während das E-Commerce die vertriebszentrierte Perspektive ins Auge fasst. Generell ist es in der wissenschaftlichen Literatur laut Turban et. al (2018, S. 7) allerdings schwierig eine einheitliche Abgrenzung der beiden Begriffe zu finden, da diese teilweise im selben Kontext verwendet werden, manchmal aber das E-Commerce auch als Teil des E-Business gesehen wird.

Große Holtforth (2017, S. 3) erläutert, dass E-Commerce wesentliche Vorteile, wie einen günstigeren, flexibleren, innovativeren und somit dynamischeren Prozess, im Vergleich zum stationären Handel bietet. In digital-vernetzten Handelsplattformen gekaufte Güter werden per Paketversand- und Warenlogistik an den Konsumenten geschickt, welcher die Bezahlung ortsungebunden und über digitale Zahlungssysteme abwickelt (Große Holtforth 2017, S. 4). Hierdurch bietet sich für den Kunden durch ein transparentes Angebot mehrerer Unternehmen, einem zeitlich ungebundenen Einkauf und einer fehlenden Standortgebundenheit ein deutlich höherer Freiheitsgrad als beim stationären Handel (Große Holtforth 2017, S. 4). Allerdings bietet E-Commerce laut Große Holtforth (2017, S. 4) nicht nur den Kunden Vorteile, sondern auch für den Händler, da dieser durch das Onlinegeschäft zum Beispiel Kosten für Personal und Miete spart. Zudem vergrößert sich der Kundenkreis durch die zeitliche und örtliche Ungebundenheit und die Bezahlung wird durch digitale Bezahlssysteme vereinfacht und automatisiert. Außerdem entstehen Skaleneffekte, „die bei der Bereitstellung von digitalen Informationen und Prozessschritten zu einer besonderen Wachstumsdynamik führen können.“ (Große Holtforth 2017, S. 4-5) Turban et al. (2018, S. 16) zählen zu den Vorteilen des E-Commerce auf Seiten der Verkäufer unter anderem die globale Reichweite, die Kostenreduktion, die zeitliche Ungebundenheit, die Verbesserung der Logistikketten, die Möglichkeit des personalisierten Angebots und die einfachere Kommunikation mit den Kunden. Die Vorteile für die Kunden sehen sie beispielsweise in der großen, dauerhaft zur Verfügung stehenden Auswahl, in der Allgegenwart der

Onlineshops, in der Transparenz des Angebots und in dem zunehmenden Komfort des Einkaufserlebnisses für den Kunden (Turban et al. 2018, S. 16). Es bleibt anzumerken, dass E-Commerce nicht nur Vorteile für Käufer und Verkäufer bietet. Für den Kunden ist es beispielsweise bei manchen Produkten schwierig, die Qualität der Produkte oder des Kaufprozesses zu beurteilen (Große Holtforth 2017, S. 5). Außerdem können Bewertungen von Onlinehändlern manipuliert sein, was eine Beurteilung des Händlers vorab erschwert (Große Holtforth 2017, S. 5). Benötigt ein Kunde ein Produkt sofort, besteht zudem das Problem, dass nur wenige Onlinehändler eine direkte Zustellung anbieten und wenn, dann für einen hohen Aufpreis, was für eine Unsicherheit zwischen Erwerb und Lieferung des Produkts sorgt (Große Holtforth 2017, S.5). Auch für den Verkäufer entsteht ein Problem, da die Markteintrittsbarrieren ins Onlinegeschäft sehr gering sind und es darauf ankommt, welcher Onlinehandel es am besten schafft, den Kunden über das Internet zu erreichen (Große Holtforth 2017, S. 5).

Für diese Masterarbeit wird der Bereich des E-Commerce in drei Bereiche aufgeteilt, die in der Analyse in Kapitel drei genauer betrachtet werden sollen, wenn es um mögliche Fehlerquellen beim Einsatz von Big Data geht. Gebildet werden die drei Bereiche anhand der Umsatzgrößen von Warengruppen in Deutschland im ersten Quartal 2019, die der Bundesverband für E-Commerce und Versandhandel Deutschland e.V. in der Pressemitteilung bekannt gibt (bevh 2019). Hierbei werden die Versendertypen im E-Commerce-Handel nicht berücksichtigt, da es lediglich um das Einkaufsverhalten der Kunden und somit um die Warengruppen mit dem stärksten Umsatz gehen soll. An erster Stelle steht die Warengruppe „Bekleidung inkl. Schuhe“, deren Umsatz sich im ersten Quartal 2019 auf 4.181 Millionen Euro inklusive Umsatzsteuer belief (bevh 2019, S 1). An zweiter Stelle steht mit einem Umsatz von 2.542 Million Euro inklusive Umsatzsteuer die Warengruppe „Einrichtung“, worunter Möbel, Lampen, Deko, Haus- und Heimtextilien und Haushaltswaren und -geräte fallen (bevh 2019, S. 1). Die Warengruppe „Freizeit“ steht mit einem Umsatz von 1.861 Millionen Euro inklusive Umsatzsteuer an dritter Stelle und umfasst DIY und Blumen, Spielwaren, alles rund um das Auto, Motorrad oder Zubehör und Hobby- und Freizeitartikel (bevh 2019, S. 1).

3. Analyse von Strategien zur Fehlerprävention und zum Fehlermanagement bezogen auf Datenanalysen im Bereich des E-Commerce

3.1. Einsatzbereiche von Big Data entlang des elektronischen Wertschöpfungsprozesses

3.1.1. Informationssammlung als erste Stufe des Wertschöpfungsprozesses

Der elektronische Wertschöpfungsprozess umfasst die gesamte elektronische Wertschöpfungskette, an deren Ende ein fertiges elektronisches (Informations-)Produkt steht (Kollmann 2019, S. 60). Die elektronischen Wertschöpfungsaktivitäten sind allerdings nicht vergleichbar mit den Aktivitäten einer realen Wertschöpfungskette, da letztere lediglich einen unterstützenden Charakter für die elektronische Kette hat, um dieses Angebot realisieren zu können (Kollmann 2019, S. 59-60). Nichtsdestotrotz lässt sich festhalten, dass es, in Anlehnung an Porter (1998), auch bei der elektronischen Wertschöpfungskette darum geht, dass eine Reihe von wertschöpfenden Aktivitäten einen Wert generieren und aufbauen, der im besten Fall an die Kunden weitergegeben werden kann (Miller und Mork 2013, S. 2). Eine „Data Value Chain“, wie Curry (2016, S. 31) sie bezeichnet, charakterisiert sich durch mehrere Schritte, die aus einem Datenbestand einen Wert und nützliche Einblicke in die Zusammenhänge zwischen Daten generieren (Curry 2016, S. 31). Er zitiert die Europäische Kommission, die diese Wertschöpfungskette als Zentrum der zukünftigen Informationsökonomie sieht, und Möglichkeiten der digitalen Entwicklung in verschiedenste Bereiche bringt (Curry 2016, S. 31). Der Wandel zum elektronischen Wertschöpfungsprozess, bleibt nicht ohne Folgen für den traditionellen Handel und sorgt dafür, dass es durch den E-Commerce zu deutlichen Veränderungen im Wertschöpfungsprozess kommt (Picot et al. 2001, S. 21). Der Vorteil des elektronischen Marktes liegt vor allem in der Entkopplung von Raum und Zeit, wodurch die Abwicklung deutlich beschleunigt werden kann und somit die Daseinsberechtigung von Handelsvertretern auf traditionellen Märkten in Frage gestellt wird (Picot et al. 2001, S. 22). Zusammenfassend halten Picot et al. (2001, S. 23) fest, dass „durch E-Commerce in der Wertschöpfungskette des Handels sowohl eine starke Entbündelung als auch Substitution von Funktionen des Handels erfolgt.“ Allerdings merken sie auch an, dass neue Stufen in der elektronischen Kette entstehen, die

Intermediation beziehungsweise Hypermediation genannt werden (Picot et al. 2001, S. 23) und auf deren Funktion innerhalb von datenzentrischen Geschäftsmodellen in den Ausführungen in Kapitel 2.3.2. zu Dorfers Untersuchungen (2016) eingegangen wurde. Aufbauend auf dem elektronischen Wertschöpfungsprozess nach Kollmann (2019, S. 59-62) werden in diesem und in den folgenden zwei Kapiteln Grundlagen, Technologien und Geschäftsaktivitäten zu den Prozessstufen genauer betrachtet. Anhang fünf zeigt den elektronischen Wertschöpfungsprozess mit der integrierten elektronischen Wertschöpfungskette nach Kollmann (2019, S. 61).

Im ersten Schritt des elektronischen Wertschöpfungsprozesses, der Informationssammlung, geht es zunächst darum, dass relevante Daten als Input für die weitere Wertschöpfung gesammelt werden, um so einen nutzbaren Datenbestand aufzubauen (Kollmann 2019, S. 62). Durch die Sammlung der Daten soll eine Effektivitätssteigerung erreicht werden, da durch eine unkomplizierte, schnelle und umfassende Informationsgewinnung über die individuellen Ansprüche oder Vorstellungen potenzieller Kunden die Basis „für die Realisierung eines auf die individuellen Wünsche zugeschnittenen Leistungsangebotes“ geschaffen werden soll (Kollmann 2019, S. 62). Zudem bietet die frühzeitige und umfassende Kenntnis der Kundenansprüche und -vorstellungen dem Unternehmen die Möglichkeit, diese mit in die Produktgestaltung einfließen zu lassen (Kollmann 2019, S.62).

Das Ziel ist der Aufbau eines qualitativ hochwertigen Datenbestandes im Sinne der Vollständigkeit, Glaubwürdigkeit, Zusammensetzung, Aktualität und Richtigkeit der Datenquellen (Miller und Mork 2013, S. 2). Daten, die von Beginn an keine Aussagekraft haben, werden auch in der Analyse keinen Mehrwert für das Unternehmen bieten (Miller und Mork 2013, S. 4). Der Dateninput, der dann die Grundlage der späteren Datenanalyse bildet, kommt meist, in Anlehnung an das in Kapitel 2.2.3. herausgestellte Merkmal Variety, aus vielen unterschiedlichen Datenquellen. Daher sind Plattformen notwendig, die die Datenmengen erfassen, speichern und im späteren Verlauf analysieren können (Lyko et al. 2016, S. 41). Für diesen Informationsinput, der aus den Datenquellen hervorgeht, gibt es vordefinierte Protokolle innerhalb eines Rahmenplans der Datenerfassung, wo es um die Datenspeicherung geht (Lyko et al. 2016, S. 41). In ihren Ausführungen erläutern Lyko et al. (2016, S. 42), dass es neben

öffentlich zugänglichen Protokollen auch zum Teil unternehmensspezifische Protokolle gibt, die nicht einsehbar sind. Als Beispiel eines der am häufigsten verwendeten Protokolle nennen Lyko et al. (2016, S. 42-44) unter anderem das Advanced Message Queuing Protocol (AMQP), auf das hier kurz zum besseren Verständnis dieser Protokolle eingegangen werden soll. AMQP wurde aufgrund des Bedarfs nach einem verfügbaren Protokoll entwickelt, das die Anforderung großer Unternehmen an die Datenerfassung erfüllt (Lyko et al. 2016, S. 42). 2012 wurde dieses Protokoll zu einem Standard der Organization for the Advancement of Structured Information Standards (OASIS) und erfüllt die Anforderungen nach Allgegenwärtigkeit, Sicherheit, Genauigkeit, Einsetzbarkeit, Kompatibilität und Handhabbarkeit (Lyko et al. 2016, S. 42-43). Laut der OASIS (o.D.) bietet das Protokoll Unternehmen einen einfachen und sicheren Ansatz, um Datenflüsse und Geschäftstransaktionen in Echtzeit weiterzuleiten. Das Ziel von AMQP ist Folgendes: „The goal of AMQP is to ensure information is safely and efficiently transported between applications, among organizations, across distributed cloud computing environments, and within mobile infrastructures.“ (OASIS o.D.). Außerdem nennen Lyko et al. (2016, S. 44-50) Softwarewerkzeuge, die für die unterschiedlichen Schritte der Datenerfassung genutzt werden können. Für die Datenakquisition eignen sich beispielsweise Technologien wie Kafka oder Flume (Lyko et al. 2016, S. 47-49). Kafka wurde von LinkedIn entworfen, damit die große Masse an Websiteaktivitäten nachverfolgt werden kann (Lyko et al. 2016, S. 47). Dies ist laut Lyko et al. (2016, S. 47) aus folgendem Grund wichtig: „These activity events are critical for monitoring user engagement as well as improving relevancy in their data-driven products.“ Zudem bietet Kafka die Möglichkeit, die gesammelten Daten zeitgleich in die Open Source Software Hadoop zu laden, welche zur Datenverarbeitung genutzt wird (Lyko et al. 2016, S. 47). Bei Flume handelt es sich um ein System, das folgenden Zweck erfüllen soll: „The purpose of Flume is to provide a distributed, reliable, and available system for efficiently collecting, aggregating, and moving large amounts of log data from many different sources to a centralized data store.“ (Lyko et al. 2016, S. 48) Außerdem bezeichnen es Lyko et al. (2016, S. 48) als ein robustes und fehlertolerantes System, das über Zuverlässigkeits-, Ausfallsicherungs- und Wiederherstellungsmechanismen verfügt.

3.1.2. Informationsverarbeitung als zweite Stufe des Wertschöpfungsprozesses

Die Informationsverarbeitung charakterisiert sich durch die Bearbeitung des gesammelten Datenbestandes aus der ersten Prozessstufe, um ein Informationsprodukt für den Kunden zu erstellen (Kollmann 2019, S. 62). Hierbei steht das Ziel der Effizienzsteigerung laut Kollmann (2019, S. 63) im Fokus, da die zeitnahe, einfache und ausführliche Verarbeitung, der aus den Daten gewonnenen Informationen, die unternehmensinternen Prozesse verbessern und Kosten einsparen kann.

Domingue et al. (2016, S. 63) machen deutlich, dass Datensätze umso besser maschinell weiterverarbeitet werden können, desto mehr Struktur sie haben. Datenanalysen befassen sich deshalb mit der Strukturierung großer Datenmengen, um Zusammenhänge zwischen den Daten zu erkennen und das hierdurch generierte Wissen für Unternehmenszwecke anwenden zu können (Domingue et al. 2016, S. 63). In der Industrie werden laut Domingue et al. (2016, S. 67) heutzutage groß angelegte maschinelle Lernprogramme und andere Algorithmen für die Analyse der Datenmengen genutzt. Kombiniert werden diese „with complex event processing and stream processing“ (Domingue et al. 2016, S. 67). Diese Kombinationen sind so ausgelegt, dass sie die komplizierten Prozesse in Echtzeit analysieren (Domingue et al. 2016, S. 67). Diese maschinellen Lernprogramme und Algorithmen können dann in Verbindung mit einer kompatiblen Big Data Software, wie Hadoop, angewendet werden, um dadurch große Datenmengen schnell zu wertvollen Ergebnissen weiterverarbeiten zu können (Domingue et al. 2016, S. 69-70). Die Software zur Verarbeitung von sehr großen Datenmengen hat sich in den letzten Jahre stetig weiterentwickelt (Domingue et al. 2016, S. 74). Hadoop und die MapReduce Technologie werden hierbei vor allem wegen der einfachen Nutzbarkeit und Kompatibilität dieser Plattform von einer Vielzahl an Unternehmen genutzt (Domingue et al. 2016, S. 75). Die Open Source Software Hadoop wird unter anderem von Facebook verwendet und bietet die Möglichkeit der parallelen Verarbeitung von Daten, die innerhalb eines bestimmten Zeitabschnitts gesammelt wurden (Lyko et al. 2016, S. 49). Für eine Echtzeitverarbeitung von Daten werden laut Lyko et al. (2016, S. 45-47) Plattformen wie Storm oder S4 genutzt. Das Ziel der Datenanalyse durch die angesprochenen Technologien ist im Allgemeinen

Folgendes: „Large volumes of data which may be heterogeneous with respect to encoding mechanism, format, structure, underlying semantics, provenance, reliability, and quality is turned into data which is usable.“ (Domingue et al. 2016, S. 83)

Walsh et al. (2013, S. 48) halten fest, dass durch den Einsatz von Big Data im E-Commerce neue Chancen und Möglichkeiten vor allem da entstehen, wo Datenauswertungen bisher an ihre Grenzen gestoßen sind. Als Beispiel nennen sie das Retourenmanagement von E-Commerce-Unternehmen, welches durch Datenanalysen die Möglichkeit einer einfacheren Identifikation von Kunden mit einer hohen Neigung zu Retouren und möglichen Auslösern für diese hat (Walsh et al. 2013, S. 48). Durch eine elektronische Abwicklung der Interaktion mit dem Kunden besteht im E-Commerce die Möglichkeit, dass von den Kunden getätigte Transaktionen und die Kundenhistorie genau protokolliert werden können (Walsh et al. 2013, S. 49). Allerdings besteht laut Walsh et al. (2013, S. 49) das Problem, aus diesen Protokollen Handlungsempfehlungen für die Kunden auszuwerten und abzuleiten. Zudem erläutern Walsh et al. (2013, S. 49), dass die für die Auswertungen genutzten Technologien als Geschäftsgeheimnis gesehen werden und diese fehlende Transparenz der Datenverarbeitung es erschwert, den Datenanalysen von Unternehmen auf den Grund zu gehen. Der Begriff Big Data umfasst in diesem Kontext den Technologieursprung einer Reihe von technologischen Entwicklungen im Zusammenhang mit der Datenverarbeitung, wozu zum Beispiel das Cloud-Computing, das In-Memory-Computing und Hadoop zählen (Walsh et al. 2013, S. 49).

Neben der Analyse der gesammelten Datenmengen, geht es auch um die Pflege des Datenbestandes (Freitas und Curry 2016, S. 87). Unter dem Begriff der „data curation“ verstehen Freitas und Curry (2016, S. 87) „the methodological and technological data management support to address data quality issues maximizing the usability of the data.“ Dabei geht es laut Freitas und Curry (2016, S. 87) in Anlehnung an Craign et al. (2007) darum, dass der Datenbestand aktiv und fortlaufend bearbeitet wird, um die Datenqualität für eine eventuelle Wiederverwendung aufrechtzuerhalten und einen sinnvollen Datenbestand aufzubauen. Die Prozesse der Datenpflege umfassen dabei „content creation, selection, classification, transformation, validation, and preservation.“ (Freitas und Curry 2016, S. 88).

Eine ausreichende Datenpflege impliziert allerdings auch, dass die Daten adäquat gespeichert werden. Strohbach et al. (2016, S. 121) merken in diesem Kontext an, dass individuelle, technische Möglichkeiten zur Datenspeicherung eine bessere Skalierbarkeit und geringere Komplexität und Kosten als ältere Systeme zur Handhabung von Datenbanken bieten. Beispielsweise ist das Hadoop Distributed File System (HDFS) eine effiziente und kostengünstige Alternative, wenn mehrere Speichersysteme miteinander verbunden werden, um die Charakteristika Volume, Velocity und Variety von Big Data zu bewältigen (Strohbach et al. 2016, S. 120). Das HDFS ermöglicht zum Beispiel eine zuverlässige Speicherung und Aufbewahrung von großen Mengen unstrukturierter Daten auf standardisierter Hardware, da es sich hierbei um einen wesentlichen Bestandteil der Open Source Software Hadoop handelt (Strohbach et al. 2016, S. 123). Aber auch das Cloud Computing wächst vor allem mit Blick auf den Kontext rund um Big Data in seiner Popularität (Strohbach et al. 2016, S. 126). Allerdings haben sowohl HDFS als auch andere Systeme zur Datenspeicherung ihre Schwachstellen und Strohbach et al. (2016, S. 119) erläutern, dass ein perfektes System zur Datenspeicherung fähig wäre, eine virtuell unbegrenzte Menge an Daten aufzubewahren und sowohl mit strukturierten als auch unstrukturierten Daten umgehen kann. Zudem müsste es Folgendes können: „The ideal big data storage system would (...) cope both with high rates of random write and read access, flexibly and efficiently deal with a range of different data models (...) and for privacy reasons, only work on encrypted data.“ (Strohbach et al. 2016, S. 119) Hier ist anzumerken, dass sich die Datenpflege und -speicherung auf den gesamten elektronischen Wertschöpfungsprozess beziehen und nicht allein der Prozessstufe der Informationsverarbeitung zugeordnet werden können. Auch in der Informationssammlung und -übertragung spielen die Pflege und die Speicherung und Aufbewahrung der Daten eine große Rolle.

3.1.3. Informationsübertragung als dritte Stufe des Wertschöpfungsprozesses

Die Informationsübertragung als letzte Stufe des Wertschöpfungsprozesses umfasst die Übertragung, des durch die Datenanalyse erlangten Wissens, auf Unternehmensentscheidungen (Kollmann 2019, S. 63). Das Resultat stellt wertvolle Informationen für

den Kunden dar, wobei es erneut um das Ziel der Effektivitätssteigerung geht (Kollmann 2019, S. 63). Durch die einfache, schnelle und umfassende Übertragung der erlangten Informationen besteht die Möglichkeit einer „verbesserten Wahrnehmung der Vorteilhaftigkeit eines Angebotes“ durch den Kunden (Kollmann 2019, S. 63). Hierdurch ergibt sich ebenfalls die Möglichkeit, sich von der Konkurrenz abzusetzen und Wettbewerbsvorteile durch ein vorteilhafter wirkendes Angebot zu erlangen (Kollmann 2019, S. 63). Um eine Nutzung des Outputs der Datenanalyse und die vorherige Verarbeitung zu ermöglichen, liegt die Herausforderung vor allem darin, „Informationskonsumenten Daten und Informationen zur Verfügung zu stellen, die korrekt, nützlich, verständlich und zugänglich sind.“ (Wölfl et al. 2019, S. 217) Die Datenqualität und auch die Qualität des generierten Informationsoutputs für den Kunden ist hierbei eng mit den Datenverarbeitungsprozessen verbunden (Wölfl et al. 2019, S. 217).

Wenn sich ein Unternehmen für datenverarbeitende Maßnahmen entscheidet, geht es in den meisten Fällen um die Unterstützung bei Geschäftsentscheidungen (Becker 2016, S. 143). Der Prozess, in dem die Analyseergebnisse der Datenverarbeitung in Unternehmensentscheidungen einfließen, umfasst hierbei „reporting, exploration of data (browsing and lookup), and exploratory search (finding correlations, comparisons, what-if scenarios, etc.).“ (Becker 2016, S. 143) Die dahintersteckende Informationslogistik trägt durch eine Kontrolle über die Wertschöpfungskette und durchgängiger Transparenz zum Unternehmenswert bei (Becker 2016, S. 143). Im Kontext von Big Data ist anzumerken, dass Big Data Einfluss auf die Gültigkeit von datengetriebenen Entscheidungen nimmt, da Daten zum Beispiel für kurzfristige oder aber langfristige Empfehlungen geeignet oder die Daten entweder aktuell beziehungsweise historisch sein können (Becker 2016, S. 143). Eine der Hauptauswirkungen der Nutzung von Big Data und damit einhergehend des elektronischen Wertschöpfungsprozesses, ist das Aufdecken von Zusammenhängen und Abhängigkeiten in den analysierten Daten (Becker 2016, S. 146). Dies bietet für die Übertragung dieser gewonnenen Informationen nicht nur unternehmerische Möglichkeiten und führt zu mehr Effizienz, sondern gewährt durch ein besseres Verständnis der Zusammenhänge mehr Transparenz in den verschiedenen unternehmerischen Entscheidungsprozessen (Becker 2016, S. 146).

Für die Übertragung der gewonnenen Informationen ist eine Untersuchung der Ergebnisse wichtig, die zum Beispiel durch eine geeignete Visualisierung durchgeführt werden kann (Becker 2016, S. 152). Eine gut durchdachte und mit Bedacht ausgewählte visuelle Präsentation der Ergebnisse kann hierbei entscheidend sein, um große Mengen von Ergebnissen für unternehmerische Entscheidungen, und dem damit zusammenhängenden Erkennen von Datenbeziehungen, nutzbar zu machen (Becker 2016, S. 152). Je nachdem, wie komplex die zu verarbeitenden Datenmengen sind, kann eine geeignete Visualisierung kostenintensiv sein und bei großer Komplexität an die Grenzen der Interaktivität stoßen (Becker 2016, S. 152). Vor allem, wenn die Ergebnisse der Datenanalysen zum Beispiel für die Übertragung an viele Endnutzer bestimmt sind, ist die Komplexität der Ergebnisdaten aus der Datenanalyse oft genauso hoch wie die der gesammelten Daten (Becker 2016, S. 152). Daher stoßen statistische Visualisierungsverfahren oft an diesem Punkt an ihre Grenzen, was interaktive Möglichkeiten der Visualisierung nötig macht (Becker 2016, S. 152). Diese sollen laut Becker (2016, S. 152) in Anlehnung an Shneiderman (1996) sieben Aufgaben erfüllen: „overview, zoom, filter, details-on-demand, relate, history, and extract.“ Im Rahmen von Datenmodellen auf Basis von Algorithmen des ML wird die Qualität der Klassifikationen, Gruppierungen, Empfehlungen und Vorhersagen durch die Datenanalysen anhand von bereits verstandenen und geprüften Datensätzen getestet (Becker 2016, S. 152).

Für die Analyse dieser Masterarbeit lässt sich der in Kapitel 3.1.1 bis 3.1.3 vorgestellte elektronische Wertschöpfungsprozess und die darin eingebundene elektronische Wertschöpfungskette nach Kollmann (2019) methodisch nutzen. Die Kette umfasst hierbei die wertschöpfenden Geschäftsaktivitäten innerhalb von informationsverarbeitenden Prozessen, die zur Erstellung eines elektronischen (Informations-)Produktes beitragen und durch den elektronischen Wertschöpfungsprozess zusammengefasst werden (Kollmann 2019, S. 59). Zu den Aktivitäten gehören die Sammlung, die Systematisierung, die Auswahl, die Kombination, die Verteilung, der Austausch, die Bewertung und das Angebot von Informationen (Kollmann 2019, S. 60). Allerdings bleibt kritisch anzumerken, dass Kollmann (2019, S. 59-61) in seinen Ausführungen lediglich den Terminus Information benutzt. In Anlehnung an Kapitel 2.2.1 ist festzuhalten, dass eine Differenzierung zwischen den Begriffen Daten und Informationen sinnvoll ist, um die

Umwandlung von Daten in Informationen, dann in Wissen und im besten Fall in Weisheit besser aufzeigen zu können. Methodisch können die Stufen von Kollmanns Wertschöpfungsprozess (2019) noch jeweils um eine Prä- und eine Post-Stufe ergänzt werden, um herauszustellen, was kurz vor beziehungsweise nach den Wertschöpfungsstufen von einem Unternehmen beachtet werden muss, um Fehlerquellen zu vermeiden. Auf die in Kapitel 2.4 herausgestellten Bereiche des E-Commerce – „Bekleidung inkl. Schuhe“, „Einrichtung“ und „Freizeit“ – wird in der Analyse der Fehlerquellen immer wieder Bezug genommen, um durch Beispiele aus diesen Bereichen die getätigten Aussagen erläutern zu können. Mit Blick auf die Thematik dieser Arbeit, werden in der Analyse nicht nur Praxisbeispiele von Fehlern rund um Datenanalysen im Bereich des E-Commerce im Sinne eines Fehlermanagements gezeigt, sondern auch mögliche Risikobereiche für Fehlerquellen bezogen auf die Fehlerprävention. Dadurch kann sowohl aufgezeigt werden, welche Herausforderungen bereits für E-Commerce Unternehmen problematisch waren, aber auch, welche Herausforderungen problematisch sein können. Der Begriff Fehlerquelle ist daher nicht direkt gleichzusetzen mit einem zwangsläufigen Fehler, sondern kann auch Risikobereiche abbilden. Generell ist anzumerken, dass ein Fehler, in Anlehnung an Anhang eins, zu jeder Zeit im Wertschöpfungsprozess eines Unternehmens auftreten kann. Ein Fehlermanagement kann somit zu jeder Zeit des Wertschöpfungsprozesses für ein Unternehmen nötig sein. Die Fehlerprävention hat hingegen bereits ab Beginn eines Projektes eine große Bedeutung für den Erfolg, da es hierbei darauf ankommt, die Fehlerrate auf ein Minimum zu reduzieren und die Risiken des Projektes abzuschätzen.

3.2. Stufen des elektronischen Wertschöpfungsprozesses mit Blick auf mögliche Fehlerquellen von Datenanalysen im Bereich des E-Commerce

3.2.1. Fehlerquellen in der Informationssammlung

3.2.1.1. Fehlerquellen in der Prä-Informationssammlung

Bevor ein Unternehmen Geschäftsaktivitäten zur Erstellung eines unternehmerischen Werts ausführt, sollten zunächst die gesetzlichen Rahmenbedingungen für eine Nutzung von Big Data Technologien als Grundlage einer solchen Maßnahme betrachtet werden

(BITKOM 2014, S. 137). Da der Umfang dieser Arbeit eingeschränkt ist, wird in diesem Absatz lediglich auf die Datenschutz-Grundverordnung (DS-GVO) der Europäischen Union eingegangen und nicht zusätzlich noch auf internationale Bestimmungen.

Am 14.04.2016 wurde die DS-GVO mit großer Mehrheit vom europäischen Parlament angenommen und trat nach einer Übergangsphase von zwei Jahren am 25.05.2018 innerhalb der Europäischen Union in Kraft (Bundesministerium für Wirtschaft und Energie o.D., S. 1). Bei der Sammlung von Daten müssen Unternehmen damit seit etwas mehr als einem Jahr darauf achten, dass sie ihre Geschäftsabläufe an diese neue Rechtsgrundlage angepasst haben (Bundesministerium für Wirtschaft und Energie o.D., S. 1). Ziel der DS-GVO ist mit Blick auf die zunehmende Digitalisierung der Gesellschaft das Schaffen einer Balance zwischen Wirtschafts- und Verbraucherinteressen (Bundesministerium für Wirtschaft und Energie o.D., S. 1). Durch die Verordnung soll eine Stärkung des Grundrechts „auf informationelle Selbstbestimmung durch höhere Transparenz und mehr Mitbestimmung der Bürgerinnen und Bürger mit Blick auf ihre Daten“ geschaffen werden (Bundesministerium für Wirtschaft und Energie o.D., S. 1). Für die in Kapitel 2.3.2 erläuterten datenzentrischen Geschäftsmodelle und generell für datenverarbeitende Unternehmen bietet die DS-GVO zudem einen zukunftsorientierten europaweiten Rechtsrahmen, wodurch Wettbewerbsverzerrungen und Marktzugangsbarrieren aufgrund verschiedener nationaler Datenschutzrechte beseitigt werden (Bundesministerium für Wirtschaft und Energie o.D., S. 1). Zu den zentralen Elementen der DS-GVO zählt zum Beispiel die Anonymisierung der Daten, damit keine Rückschlüsse auf reale Personen gezogen werden können (Bundesministerium für Wirtschaft und Energie o.D., S. 1). Außerdem lässt die DS-GVO eine Datenverarbeitung auch ohne Einwilligung des Kunden zu, wenn dies zur Vertragserfüllung beiträgt, oder berechnete Interessen vorliegen (Bundesministerium für Wirtschaft und Energie o.D., S. 1). In diesen Fällen, muss der Zweck allerdings durch ein Unternehmen gut begründet werden können (Bundesministerium für Wirtschaft und Energie o.D., S. 1). Zudem muss ein Kunde, der von einer Verarbeitung der eigenen Kundendaten betroffen ist, vorher gut informiert sein, was genau mit den Kundendaten geschieht und aus welchen Gründen diese gesammelt werden (Bundesministerium für Wirtschaft und Energie o.D., S. 1). Durch diese neuen

Bestimmungen soll die DS-GVO dazu beitragen, dass den Nutzern durch die Anonymisierung mehr Privatsphäre, und durch die Auskunftspflicht zur Verarbeitung mehr Transparenz geboten wird (Bundesministerium für Wirtschaft und Energie o.D., S. 1).

Da sich sowohl Fahrrad XXL und IKEA in ihren Datenschutzerklärungen (Fahrrad XXL o.D., Kapitel 3; IKEA o.D., Kapitel 4.2) auf Artikel sechs Absatz eins der DS-GVO beziehen, wenn es um die Rechtsgrundlage zur Verarbeitung personenbezogener Informationen geht, soll an dieser Stelle beispielhaft ein genauerer Blick auf diesen Artikel geworfen werden. In der DS-GVO heißt es in Artikel sechs Absatz eins (Europäisches Parlament 04.05.2016, S. 36) wie folgt: Eine Verarbeitung von Daten ist nur rechtmäßig, wenn die betroffene Person dieser zuvor zugestimmt hat oder die Verarbeitung zur Erfüllung eines Vertrages beziehungsweise vorvertraglicher Maßnahmen notwendig ist. Außerdem ist eine Verarbeitung rechtmäßig, wenn sie zur Erfüllung von rechtlichen Verpflichtungen beiträgt oder dadurch lebenswichtige Interessen betroffener Personen geschützt werden (Europäisches Parlament 04.05.2016, S. 36). Wenn die Verarbeitung zur Erledigung einer im öffentlichen Interesse liegenden Aufgabe beiträgt oder in der Ausübung einer öffentlichen Gewalt erfolgt, dann ist sie ebenfalls gerechtfertigt (Europäisches Parlament 04.05.2016, S. 36). Als letzter Punkt wird genannt, dass die Verarbeitung rechtmäßig ist, wenn sie „zur Wahrung der berechtigten Interessen des Verantwortlichen oder eines Dritten erforderlich“ ist, solange diese Interessen die betroffene Personen und hierbei vor allem Kinder nicht in ihren Grundrechten und Grundfreiheiten verletzen (Europäisches Parlament 04.05.2016 S. 36). IKEA (IKEA o.D., Kapitel 4.2) erläutert hierzu beispielsweise, dass sie ihre Datenverarbeitung durch eine Interessenabwägung legitimieren, da sie das Interesse haben den Kunden eine „reibungslose Lieferung zu den für Dich bestmöglichen Zeiten zu ermöglichen.“

Die Betrachtung des Artikels sechs Absatz eins der DS-GVO zu Beispielzwecken macht deutlich, wie wichtig eine genaue Betrachtung der gesetzlichen Rahmenbedingungen durch das Unternehmen für den Erfolg von Big Data Maßnahmen ist, da nur durch Kenntnisse des Möglichen das maximale Potenzial ausgeschöpft werden kann. Ein unzureichendes Wissen der gesetzlichen Grundlagen und eine daraus resultierende fehlerhafte Durchführung von Big Data Maßnahmen im Unternehmen, kann dazu

führen, dass zum Beispiel eine unrechtmäßige Verarbeitung von Kundendaten auch noch im Nachhinein zu gravierenden Problemen für das Unternehmen führen kann. Die DS-GVO hat aber auch für Schwierigkeiten bei der Umsetzung der neuen Rechtsgrundlagen gesorgt, da die Umstellung zeit- und kostenintensiv war und daher vor allem für kleine und mittelständische Unternehmen eine Herausforderung darstellte (Augsburger Allgemeine 2019, S. 1). Der Druck des letzten Jahres, die Geschäftsaktivitäten eines Unternehmens bis zum 28.05.2018 DS-GVO konform zu machen (Bundesministerium für Wirtschaft und Energie o.D., S. 1), wirft zudem die Frage auf, wie gründlich die individuellen Auseinandersetzungen mit der neuen Rechtsgrundlage waren.

3.2.1.2. Fehlerquellen in der Informationssammlung

Sind alle gesetzlichen Rahmenbedingungen zur Kenntnis genommen und verstanden worden, geht es nun um die eigentliche Stufe der Informationssammlung. Um Fehlerquellen in dieser Stufe zu vermeiden ist es wichtig, dass das Unternehmen sich klar ist, wie und bei welchen Unternehmenszielen die Big Data Methoden einen unterstützenden und verbessernden Charakter annehmen kann. Wölfl et al. (2019, S. 220) erläutern hierzu, dass sich die Entscheider im Klaren über die Ziele und den Zweck einer Datenverarbeitung für das jeweilige Unternehmen sein müssen. Auch Liggesmeyer et al. (2014, S. 109) sagen bezogen auf Unternehmensziele, dass eines der Hauptprobleme von Big Data Analysen „in der Verknüpfung von Geschäftsmodellen und -zielen mit den potenziell benötigten Daten und Informationen [liegt, S.W.], um die richtigen Entscheidungen für die Organisation (...) treffen zu können.“ Die Entscheider der unternehmerischen Maßnahmen sollten daher unter anderem wissen, welche Daten bereits verfügbar sind, um fehlende Daten und Möglichkeiten zur Sammlung dieser herausstellen zu können (Wölfl et al. 2019, S. 219-220). Allerdings sind viele Nutzer aufgrund von Datenschutzbedenken nicht gewillt, ihre Daten preiszugeben. Buhl et al. (2013, S. 28) empfehlen daher, dass bereits 99 Prozent der relevanten Daten aus anderen Quellen über den Kunden gesammelt werden, damit jeder individuelle Kunde nur noch ein Prozent nachtragen muss. Dieser eine Prozent kann dann über Websiteaktivitäten oder ähnliches gesammelt werden. Da viele Unternehmen, um weiterhin aktiv am Wettbewerb mitwirken zu können, zudem schnellstmöglich versuchen Big Data Maßnahmen in ihrem Unternehmen zu etablieren, kommt noch ein

weiteres Problem hinzu: Laut Liggesmeyer et al. (2014, S. 110) werden viele Daten über verschiedene Systeme gesammelt und oft ist nicht klar, wie diese Daten im Sinne einer strategischen Ausrichtung und mit Blick auf die Rechtsgrundlage eingesetzt werden können. An dieser Stelle kann auf Kapitel 3.2.1.1 und damit auf die Wichtigkeit der Rechtsgrundlagen für eine erfolgreiche Anwendung der Maßnahmen verwiesen werden. Bezogen auf die strategische Ausrichtung kann in Anlehnung an Wölfl et al. (2019, S. 220) gesagt werden, dass diese mit den Überlegungen beginnen sollte, „welche Entscheidungen und Prozesse man verbessern könnte, wenn die relevanten Informationen dafür verfügbar wären.“

Um eine zielgerichtete Datensammlung, -verarbeitung und -übertragung durchzuführen, ist der Aufbau personeller Ressourcen von großer Bedeutung für jedes Unternehmen. In Anlehnung an Tambe (2014) erläutern Wölfl et al. (2019, S. 218), dass Unternehmen, die in fortgeschrittene Technologien und notwendiges fachkompetentes Personal investieren, einen größeren Nutzen aus dem elektronischen Wertschöpfungsprozess ziehen können. Problematisch kann es in diesem Zusammenhang werden, wenn das Personal fehlende Kenntnisse und ein mangelndes Verständnis über die Vorgänge und Möglichkeiten von Big Data Maßnahmen aufweist. Die Herausforderung liegt laut Wölfl et al. (2019, S. 222) in Anlehnung an Watson (2014) hierbei darin, dass das Unternehmen ein „Zusammenspiel verschiedener Mitarbeiter mit unterschiedlich ausgeprägten Fähigkeiten der Datenanalyse und Dateninterpretation von Datennutzern bis hin zu Data Scientists“ organisiert. Für eine effektive Datenanalyse mit erfolgversprechendem Ausgang ist für jedes Unternehmen daher eine grundlegende und hohe Personalqualität besonders wichtig, um solche Maßnahmen überhaupt möglich zu machen.

Wie in Kapitel 3.1.1 erwähnt, dienen technische Systeme, wie Kafka, unter anderem zur Sammlung von Websiteaktivitäten. Eine Sammlung dieser Aktivitäten kann hierbei beispielsweise durch die Nutzung von Cookies erfolgen. Fahrrad XXL (o.D., Kapitel zwei) definiert Cookies als „kleine Textdateien, die auf dem Computer des Nutzers gespeichert werden und die eine Analyse der Benutzung der Webseite ermöglichen.“ Das Unternehmen verwendet in diesem Zusammenhang sowohl eigene Cookies als auch Cookies von Drittanbietern, um eine personalisierte Werbung zu ermöglichen

(Fahrrad XXL o.D., Kapitel 2). Auf mögliche Fehlerquellen in der Kooperation mit Drittanbietern wird in Kapitel 3.2.2.2 noch genauer eingegangen. Auch Zalando (2019, Kapitel 16) verwendet Cookies für eine Erfassung des Kundennutzungsverhalten. Problematisch ist, dass durch das Nachverfolgen der individuellen Websiteaktivitäten jedes Nutzers eine enorme Datenmenge entsteht, deren Verarbeitung zu sinnvollen Informationen eine große Herausforderung für die Unternehmen darstellt. Dies führt zu teilweise unnötigen Erhebungen und der Speicherung von irrelevanten Daten für das eigentliche Ziel, welches das Unternehmen durch eine Big Data Maßnahme verfolgt (Wölfl et al. 2019, S. 220). Auch hier zeigt sich die Bedeutung der systematischen Planung der Datensammlung vor Beginn und vor dem Hintergrund des Unternehmensziels, um den Nutzen der Daten für die unternehmerischen Zwecke zu maximieren und die Gültigkeit der Daten gewährleisten zu können (Wölfl et al. 2019, S. 220). Andernfalls kann eine schlecht geplante Datensammlung dazu führen, dass viele irrelevante Daten ohne Qualität gesammelt werden, oder aber dass Datenquellen übersehen werden, die qualitativ hochwertige und relevante Daten hervorgebracht hätten (Wölfl et al. 2019, S. 220). Buhl et al. (2013, S. 29) sprechen von der Notwendigkeit einer „Data Governance“, um klare Vorgaben zur Datenqualität und verbundenen Datenverantwortlichkeiten und Datenqualitätsmanagementprozessen zu haben.

3.2.1.3. Fehlerquellen in der Post-Informationssammlung

Wenn sich ein Unternehmen für datenverarbeitende Wertschöpfungsaktivitäten entscheidet, dann liegen dieser Entscheidung immer individuelle Erwartungen an die Potenziale von Big Data Maßnahmen für eine Verbesserung der Unternehmensperformance zugrunde. Wie bereits, bezogen auf das Big Data Charakteristikum Volume herausgestellt wurde, wird im Kontext der Datenverarbeitung von einer enormen Masse an Daten gesprochen, die zu sinnvollen Informationen verdichtet werden sollen. Bevor eine Verarbeitung der Daten aber stattfindet, ist eine der größten Hauptaufgaben eines Unternehmens, dass nach Filtern gesucht wird, die die gesammelten rohen Daten sortieren (Jagadish et al. 2014, S. 89). Hierbei müssen die Filter so gewählt werden, dass keine sinnvollen Informationen verloren gehen und dass nach dem Filtern eine Datenmenge bleibt, deren Speicherung mit der vorhandenen Technologie des Unternehmens möglich ist (Jagadish et al. 2014, S. 89). Dies fördert auch die Übersichtlichkeit der Datenanalyse, da irrelevante Daten ausgeschlossen

werden und im Optimalfall nach diesem Schritt nur noch relevante Daten für die bevorstehende Datenverarbeitung zur Verfügung stehen. Hat ein Unternehmen zuvor eine systematische Planung der Datenerhebung durchgeführt, sollte dieser Punkt zu keinem Problem werden, da bereits relevante und irrelevante Datenquellen festgelegt wurden.

Für eine erfolgreiche Einbindung von Big Data in die Unternehmensstruktur ist die Datenqualität laut Buhl et al. (2013, S. 28) von enormer Relevanz. Nur durch eine hohe Datenqualität können E-Commerce Unternehmen, wie zum Beispiel Zalando, gewährleisten, dass sie zielgerichtet werben können und den Kunden nur mit Angeboten und Werbung bespielen, die für diesen tatsächlich auch eine Relevanz hat (Zalando 2019, Kapitel neun). Miller und Mork (2013, S. 59) erläutern, dass Daten, die keine Aussagekraft haben, auch in der Verarbeitung keine qualitativen Ergebnisse liefern werden. Somit können Sammlungen voller Daten nicht direkt gleichgesetzt werden mit wertvollen Informationen für die Nutzer beziehungsweise für das Unternehmen (Miller und Mork 2013, S. 59). Zudem ist die Glaubwürdigkeit der gesammelten Daten und im weiteren Verlauf auch der hieraus gewonnenen Informationen kritisch zu betrachten. Die Daten stammen oft aus unterschiedlichen Quellen und eine fehlende Transparenz führt in diesem Zusammenhang dazu, dass keine Informationen über die Art der Datensammlung beziehungsweise über die Qualitätssicherung bekannt sind (Liggesmeyer et al. 2014, S. 110). Die Qualität kann hierbei laut Liggesmeyer et al. (2014, S. 110) durch die Charakteristika der Vollständigkeit, Konsistenz und Aktualität der Daten beurteilt werden. Für kurzfristige Entscheidungen ist die Aktualität der Daten von enormer Bedeutung, während für langfristige strategische Entscheidungen die Vollständigkeit, Genauigkeit und Konsistenz im Vordergrund steht (Liggesmeyer et al. 2014, S. 110).

Jagadish et al. (2014, S. 89) merken an, dass die gesammelten Daten oftmals nicht in einer für die Analyse passenden Form vorliegen. Angenommen, es wurden sowohl Video-, Bild- als auch Textdaten gesammelt, wird es schwierig, diese in den unterschiedlichen Formen effektiv zu analysieren (Jagadish et al. 2014, S. 89). Vielmehr ist die technische Herausforderung hierbei, dass Erkenntnisse aus diesen Datenquellen extrahiert und in eine einheitliche strukturierte Form gebracht und diese Daten dann

analysiert werden (Jagadish et al. 2014, S. 89). Hierbei muss genau festgelegt werden, was aus den unterschiedlichen Datenquellen und -formaten extrahiert werden soll, um dann diese Daten zielführend weiteranalysieren zu können und hieraus wertvolle Informationen für das Unternehmen zu gewinnen (Jagadish et al. 2014, S. 89). Da diese Datenquellen, in Anlehnung an das in Kapitel 2.2.3 vorgestellte Big Data Charakteristika Veracity, nicht für ihre Zuverlässigkeit bekannt sind, können extrahierte relevante Daten aus unterschiedlichen Datenformaten ein erster Schritt zur Datenbereinigung sein (Jagadish et al. 2014, S. 89). Allerdings bleibt anzumerken, dass diese Maßnahmen sowohl von der verwendeten Technologie, als auch von den unterschiedlichen zur Verfügung stehenden Datenquellen abhängig, und daher nicht ohne Bedingungen umsetzbar sind (Jagadish et al. 2014, S. 89). Jagadish et al. (2014, S. 89) merken an, dass hierbei Folgendes helfen kann: „A set of data transformation and integration tools helps the data analyst to resolve heterogeneities in data structure and semantics.“ Diese Auflösung der Heterogenität des Datenbestandes sorgt dafür, dass die Daten in die standardisierten Abläufe und zu den analytischen Voraussetzungen passen (Jagadish et al. 2014, S. 89). Die Herausforderung entsteht hierbei, dass der Datenanalyst automatisch die richtigen Metadaten zur Beschreibung gesammelter Datensätze generiert (Jagadish et al. 2014, S. 90). Dafür gibt es laut Jagadish et al. (2014, S. 91) Systeme zur Erstellung dieser Metadaten, um den Datenanalysten zu entlasten. In diesem Zusammenhang merken Jagadish et al. (2014, S. 91) bezogen auf eine „data provenance“ an: „Recording information about the data at its birth is not useful unless this information can be interpreted and carried along through the data analysis pipeline.“ Datensysteme, welche die Herkunft der Daten und ihrer Metadaten transparent gestalten, ermöglichen das schnelle Auffinden von Prozessfehlern durch fehlerhafte oder irrelevante Daten während der Datenanalyse (Jagadish et al. 2014, S. 91).

3.2.2. Fehlerquellen in der Informationsverarbeitung

3.2.2.1. Fehlerquellen in der Prä-Informationsverarbeitung

Um eine korrekte Auswertung der in Stufe eins des elektronischen Wertschöpfungsprozesses nach Kollmann (2019) gesammelten Datenmengen durchzuführen, sollte es in der Prä-Informationsverarbeitung zunächst darum gehen, welche Technologien dem

Unternehmen zur Verfügung stehen, um die gesammelten Datenmengen verarbeiten zu können. An dieser Stelle sei angemerkt, dass eine geeignete Infrastruktur schon mit den passenden Technologien für die Datensammlung beginnt. Aus Begrenzungen dieser Arbeit kann an dieser Stelle allerdings nur auf die benötigte Infrastruktur zur Datenverarbeitung eingegangen werden, wodurch die Bedeutung der technischen Strukturen für das Unternehmen und eine effektive Umsetzung von Big Data Maßnahmen deutlich werden sollte.

Gentsch (2019, S. 10) merkt an, dass bei der Generierung von Informationen und Wissen aus semi- und unstrukturierten Datenmengen der Einsatz von geeigneten AI Methoden, wie ML, Natural Language Processing oder Computervision erforderlich ist. Vor allem ist aber wichtig, dass Systeme zur Datenverarbeitung die Anforderungen von Big Data erfüllen, die unter anderem durch die in Kapitel 2.2.3 vorgestellten Charakteristika von Big Data auftreten. Diese müssen in der Lage sein, die enormen Datenmengen aus den vielen unterschiedlichen Quellen vor dem Hintergrund aufzuarbeiten, dass 90 Prozent aller generierten Daten laut Wölfl et al. (2019, S. 218) in Anlehnung an Dobre und Xhafa (2014) unstrukturiert sind und daher auch auf ihre Glaubwürdigkeit geprüft werden müssen. Die Herausforderung in diesem Zusammenhang besteht darin, für eine zunehmende Komplexität der Systeme gerüstet zu sein und die umfangreichen Datenmengen dieser Systeme korrekt auszuwerten (Liggesmeyer et al. 2014, S. 105). In diesem Zusammenhang stellen Liggesmeyer et al. (2014, S. 109) heraus, dass gerade kleine und mittelständische Unternehmen nicht die Mittel haben, um sich eine geeignete Infrastruktur für die komplexen Datenanalysen anzuschaffen. An dieser Stelle müssen laut Liggesmeyer et al. (2014, S. 109) neue Lösungsansätze gefunden werden, die zum Beispiel Big Data Analysen mithilfe einer Cloud umfassen könnten oder auch zeitweise Anmietungen von Infrastrukturen. Eine Möglichkeit wäre zudem die bereits in Kapitel 3.1.2 erläuterte Open Source Software Hadoop, die öffentlich zugänglich ist und durch ihre einfache Handhabung und Kompatibilität von vielen Unternehmen genutzt wird (Domingue et al. 2016, S. 75).

Die im Rahmen dieser Masterarbeit betrachteten E-Commerce Händler in den Bereichen „Bekleidung inkl. Schuhe“, „Einrichtung“ und „Freizeit“ kooperieren zur Verarbeitung von gesammelten Daten oftmals mit Drittanbietern. Fahrrad XXL (o.D.)

nutzt beispielsweise unter anderem den Webanalysedienst Google Analytics, Google Ads Conversion Tracking, soziale Plugins von YouTube und Facebook und Retargeting Elemente der Criteo GmbH und der Belboon GmbH. Durch einen Überblick über einige der Kooperationspartner von Fahrrad XXL sollte beispielhaft deutlich geworden sein, dass

E-Commerce Händler, die nicht gerade wie Amazon zu den Weltmarktführern gehören, für eine erfolversprechende Datenverarbeitung auf Unternehmen mit einer ausreichenden Infrastruktur für die jeweiligen Schritte des Wertschöpfungsprozesses angewiesen sind. IKEA (o.D., Kapitel drei) macht im Gegenzug deutlich, dass alle gesammelten Kundendaten nur von IKEA und dadurch von unternehmensinternen Technologien verarbeitet werden und dass keine Übermittlung an Dritte stattfindet außer rechtliche Vorschriften verpflichten IKEA dazu. An dieser Stelle stellt sich die Frage, wer innerhalb des Unternehmens Zugriffsrechte auf welche Daten hat. Theoretisch müsste gewährleistet sein, dass alle Daten für die zur Datenverarbeitung befugten Personen transparent sind, damit eine Datenanalyse alle Informationen zu den individuellen Kunden mit einbeziehen kann (Wölfl et al. 2019, S. 221). Fraglich ist in diesem Zusammenhang aber, ob Unternehmen, die für die Datenverarbeitung Kooperationen mit Drittanbietern nutzen, diesen Kooperationspartnern auch alle Datenherkünfte transparent zur Verfügung stellen.

3.2.2.2. Fehlerquellen in der Informationsverarbeitung

Um herauszustellen, welche Fehlerquellen sich mit Blick auf die genutzte Technologie ergeben können, soll in diesem Absatz auf ML im Zusammenhang mit Big Data Maßnahmen eingegangen werden. Gentsch (2019, S. 29) benennt eines seiner Kapitel mit dem Titel „Intelligente Agenten werden zur Normalität“. Er hält in diesem aber auch fest, dass bisher noch keine dem Menschen ebenbürtige AI entwickelt werden konnte, die beliebige intellektuelle Aufgaben ausführen kann (Gentsch 2019, S. 29). Nichtsdestotrotz werden aktuell sehr viele Systeme im Kontext von AI entwickelt (Gentsch 2019, S. 29). Außerdem hält er Folgendes fest: „Die – häufig aus der Statistik – entwickelten Methoden des maschinellen Lernens durchdringen im Anwendungsfall heutzutage jegliche AI-Anwendung und können als algorithmische Grundlagen der AI gewertet werden.“ (Gentsch 2019, S. 30) Eine Betrachtung von den Grenzen und Herausforderungen von solchen Algorithmen vor dem Hintergrund möglicher Fehler-

quellen im Bereich der Datenverarbeitung durch diese bietet sich demnach an. Die in Kapitel 3.1.2 erläuterte Open Source Software Hadoop bietet hierfür in der Version Hadoop 3.2.0 die Möglichkeit der Entwicklung, des Trainings und der Anwendung von ML (Parbel 2019, S. 1). Najafabadi et al. (2015, S. 7) machen deutlich, dass diese lernfähigen Algorithmen vor allem dadurch attraktiv sind, dass aus den großen, unbenannten und unüberwachten Datenmengen bedeutungsvolle Zusammenhänge und Muster erkannt werden können. Sie beziehen sich hierbei auf Deep Learning (DL) Algorithmen, da sie anmerken, dass maschinelle Lernalgorithmen nicht effizient genug sind um aus den komplexen und nicht linearen Datenmengen im Kontext von Big Data Erkenntnisse zu gewinnen (Najafabadi et al. 2015, S. 3). Vielmehr erläutern sie, dass „Big Data problems such as semantic indexing, data tagging, fast information retrieval, and discriminative modeling can be better addressed with the aid of Deep Learning.“ (Najafabadi et al. 2015, S. 3). Aber auch diese Algorithmen stellen Unternehmen bei der Nutzung vor Herausforderungen. Eine technische Herausforderung ist der Umgang mit Datenstreaming und einem Erkenntnisgewinn aus diesen Datenströmen, die kontinuierlich neuen Input schaffen (Najafabadi et al. 2015, S. 14). Außerdem kann es für ein Unternehmen einen hohen Aufwand und Kostenfaktor bedeuten, wenn es um hochdimensionale Daten geht, da Bilder zum Beispiel das Volumen der Datenmasse im Gegensatz zu anderen Datenformaten massiv erhöhen und es für den Algorithmus komplizierter ist aus diesem Datentyp zu lernen (Najafabadi et al. 2015, S. 15). An dieser Stelle kann auf Kapitel 3.2.1.3 verwiesen werden und auf die Empfehlung, dass alle Daten in eine einheitliche Form gebracht werden, um die Analyse zu erleichtern. Fraglich ist an dieser Stelle auch, welche Beispiele dem Algorithmus als Input zur Verfügung stehen, um aus diesen Wissen für eine maximale Ausschöpfung des Potenzials der Datenanalyse durch die AI zu gewinnen. Mit Blick auf die vorangegangenen Ausführungen stellt sich ebenfalls die Frage, wer die Konsequenzen trägt, wenn durch die Handlungen einer AI ein Schaden am Unternehmen entsteht. Scherer (2016, S. 356) nennt hier unter anderem das Beispiel eines selbstfahrenden Autos, das einen Unfall verursacht.

Eine weitere Fehlerquelle kann die Kooperation mit Drittanbietern darstellen, auf die im vorherigen Kapitel schon mit Blick auf die Kooperationspartner von Fahrrad XXL eingegangen wurde. Wölfl et al. (2019, S. 221) erläutern in diesem Zusammenhang in

Anlehnung an Klein und Rai (2009), dass eine zentrale unternehmerische Herausforderung darin besteht, vertrauensvolle Geschäftsbeziehungen aufzubauen und die Kooperation transparent zu gestalten, um auf relevante Daten zugreifen zu können. Hier sind vor allem die Themen des Datenschutzes und der Datenhoheit von Interesse, da es unwahrscheinlich ist, dass Unternehmen ihre vollständigen Datenmengen an andere weitergeben werden (Liggesmeyer et al. 2014, S. 111). Die Frage ist, mit Blick auf das Charakteristikum Veracity, ob Ergebnisse aus Datenanalysen, die durch Kooperationspartner ohne vollständigen Zugriff auf die unternehmerischen Datenquellen durchgeführt wurden, glaubwürdig sind. Die angesprochenen Geschäftsbeziehungen können auch weitere negative Konsequenzen für die Unternehmen haben, da ein Vertrauen in Kooperationspartner auch ausgenutzt werden kann. Im März 2018 wurde der Cambridge-Analytical-Skandal bekannt, der aufgrund einer Kooperation von Facebook mit einem Forscher ermöglicht wurde, da dieser Forscher die von Facebook erhaltenen Daten über 87 Millionen Nutzer an Cambridge Analytica verkaufte (Beuth 2019, S. 1). Das Resultat ist eine fünf Milliarden Dollar Bußgeldstrafe für Facebook, die das Unternehmen aufgrund seiner Finanzkraft zwar nicht allzu stark beeinflusst (Beuth 2019, S. 1), für alle kleineren Unternehmen allerdings durchaus existenzgefährdend werden kann.

3.2.2.3. Fehlerquellen in der Post-Informationsverarbeitung

Ist die Datenanalyse abgeschlossen, besteht die Herausforderung darin, dass die Ergebnisse der Analyse zur richtigen Zeit an die befugten Personen weitergegeben werden (Wölfl et al. 2019, S. 223). Auf Grundlage der gewonnenen Erkenntnisse können Unternehmensentscheidungen getroffen beziehungsweise diese für unternehmerische Maßnahmen, wie personalisierte Werbung, genutzt werden. Der zeitliche Aspekt spielt hierbei eine wichtige Rolle, da Wölfl et al. (2019, S. 223) in Anlehnung an Li et al. (2006) deutlich machen, dass Informationen mit der Zeit ihre Nützlichkeit verlieren können. Zudem ist es von Bedeutung, dass die technische Datenanalyse Ergebnisse so ausgibt, dass diese für Entscheider innerhalb des Unternehmens verständlich, leicht zugänglich und gut nutzbar sind (Wölfl et al. 2019, S. 222). Gut nutzbar heißt in diesem Fall, dass die Ergebnisse für das Unternehmen einen Zweck erfüllen, der das Erreichen unternehmerischer Ziele positiv beeinflusst. Dadurch wird deutlich, dass die Ergebnisse immer mit Rücksicht auf das Unternehmen und dessen individuellen Zielen und Vorstellungen betrachtet werden müssen. Wie in Kapitel 3.2.1.2 herausgestellt wurde,

ist eine klare Vorstellung des Einflusses von Big Data und der durchzuführenden Datenanalysen auf die strategische Ausrichtung des Unternehmens von großer Bedeutung. Hier stellt sich die Frage, wer letztlich im jeweiligen Unternehmen entscheidet, welche Ergebnisse für das Unternehmen relevant sind und wie diese gewinnbringend eingesetzt werden können, und auf welcher betrieblichen Hierarchieebene der Umgang mit den Ergebnissen angeordnet ist (Wölfl et al. 2019, S. 222).

Eine weitere Herausforderung ergibt sich in Bezug auf Gruppen, die für die Interpretation der Analyseergebnisse zuständig sind, und deren Gruppendynamiken (Jagadish et al. 2014, S. 93). In der heutigen immer komplexer werdenden Gesellschaft ist es oftmals notwendig, dass mehrere befugte Personen aus unterschiedlichen Unternehmensbereichen in einer Gruppe an der Interpretation der Datenanalyseergebnisse arbeiten, um diese wirklich verstehen zu können (Jagadish et al. 2014, S. 93). Daher muss ein Analysesystem im Rahmen von Big Data Maßnahmen in der Lage sein, den Input durch mehrere befugte Personen und das Teilen ihrer Interpretationsergebnisse zu unterstützen (Jagadish et al. 2014, S. 93). Je nach Größe des Unternehmens, kann es hierbei auch, zum Beispiel durch verschiedene weltweite Standorte, vorkommen, dass die Personen örtlich und zeitlich voneinander getrennt sind (Jagadish et al. 2014, S. 93). In diesem Fall wäre ein Zusammenkommen der Gruppe, um über die Interpretation zu sprechen, zu aufwändig und sie müsste durch die Technologie in der Zusammenarbeit unterstützt werden (Jagadish et al. 2014, S. 93). Technisch erfordert dies laut Jagadish et al. (2014, S. 93) Folgendes: „Technically, this requires us to consider sharing more than raw datasets; we must also consider how to enable sharing algorithms and artifacts such as experimental results.“ Die Ausführungen zeigen, dass die Gruppe der befugten Personen, zur Interpretation der Ergebnisse, im Vorhinein gut überlegt ausgewählt und durch technische Methoden unterstützt werden sollte, um eine positive Gruppendynamik und in diesem Zusammenhang eine erfolgversprechende Arbeitssituation zu schaffen.

3.2.3. Fehlerquellen in der Informationsübertragung

3.2.3.1. Fehlerquellen in der Prä-Informationsübertragung

Nachdem die Datenanalyse durchgeführt wurde, müssen die Ergebnisse von einer oder mehreren hierfür befugten Personen interpretiert werden (Jagadish et al. 2014, S. 90). In Kapitel 3.2.2.3 wurde bereits die Bedeutung der Weitergabe der aus der Analyse gewonnenen Informationen an die richtigen Personen innerhalb des Unternehmens zur richtigen Zeit herausgestellt. Die Interpretation impliziert, dass die befugten Personen in diesem Zusammenhang alle Ergebnisse aus der Datenanalyse genau untersuchen und die durch die Analyse herausgestellten Annahmen zurückverfolgen (Jagadish et al. 2014, S. 90). In diesem Zusammenhang ergeben sich laut Jagadish et al. (2014, S. 90) mehrere Fehlerquellen: Die genutzten technischen Systeme zur Datenanalyse könnten Bugs haben (Jagadish et al. 2014, S. 90), die, in Anlehnung an die erläuterten Bohr- und Mandelbugs in Kapitel 2.1.1.2, große Probleme hervorrufen können und zum Teil sehr schwierig aufzuspüren sind. Außerdem werden Modelle zur Datenanalyse fast immer mit vorherigen Annahmen verbunden (Jagadish et al. 2014, S. 90), die aber nicht zwangsläufig genauso eintreten müssen und zu Verunsicherungen führen können. Wie bereits herausgestellt, ist die Datenqualität für eine effektive Datenanalyse von großer Bedeutung. Auch an dieser Stelle des Wertschöpfungsprozesses kann auf diesen Punkt verwiesen werden, da eine niedrige Qualität des aufgebauten und verarbeiteten Datenbestandes dafür sorgen kann, dass die Ergebnisse der Datenanalyse auf fehlerhaften Datensätzen basieren und dadurch zu fehlerhaften Ergebnissen führen (Jagadish et al. 2014, S. 90). Diese Gründe machen laut Jagadish et al. (2014, S. 90) deutlich, dass verantwortungsbewusste und zur Interpretation der Datenanalyseergebnisse befugte Personen die Autorität nicht an die verwendete Technologie abgeben werden. Dieser Punkt ist vor allem mit Blick auf AI-Technologien, wie das ML, interessant, da impliziert wird, dass einem technischen System immer eine Kontrollinstanz in einer unternehmerischen Hierarchie übergeordnet ist. Diese Instanz wird dann versuchen, die Ergebnisse der Datenanalyse in diesem Prozessschritt zu verstehen und diese zu verifizieren, wobei dieser Schritt mithilfe des Systems einfach durchzuführen sein sollte (Jagadish et al. 2014, S. 90). In diesem Kontext wird durch Jagadish et al. (2014, S. 90) angemerkt: „This is particularly a challenge with Big Data due to its complexity. There are often crucial assumptions

behind the data recorded.“ Da eine analytische Verarbeitung mehrere Schritte umfassen kann, die wiederum auf Annahmen aufbauen, ist es von großer Bedeutung, dass die befugten Personen in der Lage sind, diese Annahmen kritisch zu überprüfen (Jagadish et al. 2014, S. 90). Zusätzlich muss sichergestellt werden, dass die befugten Personen auch in der Lage sein werden aus den Ergebnissen der Datenanalyse Zusammenhänge zu erkennen und dass sie nicht durch die Masse an Ergebnissen überfordert sind (Jagadish et al. 2014, S. 83). An dieser Stelle kann ebenfalls auf die Bedeutung der Personalqualität verwiesen werden, da Kenntnisse und Fähigkeiten rund um Big Data für eine Einschätzung und Überprüfung der Analyseergebnisse wichtig sind. Wird vor diesem Hintergrund das in Kapitel eins erläuterte Beispiel der US-Supermarktkette Target (Beuth 2014, S. 1) betrachtet, wird deutlich, dass eine eingehende Überprüfung dieses Datenanalyseergebnisses einen signifikanten Unterschied hätte machen können und ein solcher Werbebrief voraussichtlich nicht an eine minderjährige Kundin geschickt worden wäre. Dieses Beispiel zeigt auch, dass es nicht empfehlenswert ist, Ergebnisse ohne eingehende Prüfung auf Unternehmensaktivitäten zu übertragen. Vielmehr muss eine befugte Person des Unternehmens durch eine ausreichende Transparenz in der Lage sein, die Analyseergebnisse zurückverfolgen zu können und gegebenenfalls die Analyse mit verschiedenen Annahmen, Parametern und Datensätzen zu wiederholen (Jagadish et al. 2014, S. 90). Dies hat laut Jagadish et al. (2014, S. 90) den Hintergedanken „to better support the human thought process and social circumstances.“ In diesem Zusammenhang stellt sich allerdings die Frage, wer die Hoheit über die gesammelten Datenmengen, die zugehörige Analyse und die Ergebnisse hat und wie dieser Bereich in die Unternehmensstruktur integriert wird, um für genügend Transparenz der gesamten Analyse zu sorgen.

Jagadish et al. (2014, S. 83) erläutern zudem, dass zum Beispiel „ranking and recommendation algorithms“ dabei helfen können, die für den Kunden interessantesten Daten auf Grundlage der individuellen Präferenzen herauszufiltern. Allerdings muss bei der Anwendung dieser Technologien darauf geachtet werden, dass die Kunden letztlich nicht in einer Filterblase enden, da ihnen nur Dinge angezeigt werden, die den bisherigen Erkenntnissen über den individuellen Kunden ähneln (Jagadish et al. 2014, S. 83). Diese Gefahr der Filterblase spricht auch Justizministerin Barley an und fordert ein Pluralismusgebot, um dieser Blasenbildung entgegenzuwirken und die

Wahrnehmung der Endnutzer nicht durch Algorithmen zu verzerren (Müller-Neuhof 2018, S. 1).

Um die genannten Punkte im Sinne einer effektiven Datenanalyse ausführen zu können und damit das Potenzial der Big Data Analyse zudem vollkommen ausgeschöpft werden kann, ist die Wahl einer geeigneten Visualisierung der Ergebnisse sehr wichtig. Diese muss interaktive Möglichkeiten innerhalb der Visualisierung ermöglichen und dadurch die nötige Transparenz bieten (Becker 2016, S. 152). Diesen Punkt greifen auch Jagadish et al. (2014, S. 93-94) auf, weil laut ihnen Visualisierungen menschliche Interpretationen der Ergebnisse und Zusammenarbeiten in der Gruppe unterstützen sollen und es dem Datenanalysten ermöglichen müssen, dass dieser die Herkunft der Ergebnisse aus der Datenanalyse nachverfolgen kann. Auf dieser Basis können dann unternehmerische Entscheidungen gefällt werden.

3.2.3.2. Fehlerquellen in der Informationsübertragung

Nachdem die Datenanalyseergebnisse überprüft und verifiziert wurden, können die Ergebnisse durch die befugten Personen genau betrachtet und Zusammenhänge erkannt werden, die für unternehmerische Entscheidungen zum Zweck des Erreichens der Unternehmensziele genutzt werden können. Ein Beispiel des Einsatzes der Ergebnisse und hieraus erkannter Zusammenhänge bietet Amazon (2019). Sobald eine Anmeldung des Kunden zu seinem oder ihrem individuellen Kundenkonto erfolgt, erscheint für den Kunden eine personalisierte Startseite, die ihm unter anderem Rubriken bietet wie „Zuletzt angesehen“, „Neu für Sie“ oder „Inspiriert durch Ihre Wünsche“ (Amazon 2019). Für diese unterschiedlichen personalisierten Ansprachen auf den individuellen Startseiten der Nutzer, ist kaum vorstellbar, in welchem Ausmaß die Datenerhebung von Amazon stattfindet. Diese Masse an Daten kann allerdings im Zuge der Datenverarbeitung durch eine zunehmende Unübersichtlichkeit der vorhandenen Daten zu Problemen führen. Beispielsweise hat ein Amazon Kunde im Jahr 2018 gemäß der DSGVO Einblick in sämtliche von Amazon über ihn gesammelte Daten gefordert (Fuest 2018, S. 1). Problematisch war, dass Amazon dem Kunden zusätzlich zu seinen eigenen Daten auch 1700 Audiodateien eines anderen Kunden und eine PDF-Datei, der von Alexa erkannten Sprachbefehle dieses Kunden, zuschickte, die alle auf Amazons Cloudservern gespeichert waren (Fuest 2018, S. 1). Laut Amazon handelte es sich

hierbei um einen einmaligen menschlichen Fehler, der durch eine Optimierung der Abläufe nicht wieder passieren soll (Fuest 2018, S. 1). Der rechtlichen Pflicht, den vom Datenleak betroffenen Nutzer innerhalb von 72 Stunden über dieses aufzuklären, sei Amazon laut Fuest (2018, S. 1) ebenfalls nicht nachgekommen, was den Skandal noch intensivierete.

Eine weitere Fehlerquelle ergibt sich vor dem Hintergrund, dass die Informationen aus den Datenanalysen möglichst schnell produziert werden müssen, da diese dem Kunden einen Mehrwert bieten sollen. Die Aktualität der Informationen ist für die Endnutzer hierbei von enormer Relevanz, wobei Wölfl et al. (2019, S. 223) anmerken, dass es oft zu einer zeitlich verzögerten Informationsdistribution kommt. Als Beispiel aus der Praxis kann hier Zalando genannt werden, da das Unternehmen beim Targeting, das heißt, bei der möglichst genauen Zielgruppenansprache, Schwächen aufweist (Jansen 2018, S. 1). Einem Zalando Kunden wird, trotz des vorherigen Kaufes eines Produktes, dieses immer noch als mögliche Kaufempfehlung angepriesen (Jansen 2018, S. 1). Die Personalisierung der Werbung durch Ergebnisse von den in dieser Arbeit thematisierten Datenanalysen wird zwar immer besser, aber Jansen (2018, S. 1) merkt an, dass einige Schwachstellen, wie das Beispiel gezeigt hat, bestehen bleiben. Fraglich ist auch, ob es für die Privatsphäre des Kunden nicht ratsam wäre, wenn diese Schwachstellen zumindest zum Teil bestehen bleiben. Jansen (2018, S. 1) erläutert in diesem Zusammenhang, dass eine Jeans beispielsweise sowohl von Otto als auch von Zalando angeboten werden kann und es nur durch eine total Überwachung der Nutzer möglich wäre zu wissen, wo und wann der Kunde die Hose letztlich kauft und ab wann in diesem Fall die personalisierte Kaufempfehlung abgebrochen werden müsste.

3.2.3.3. Fehlerquellen in der Post-Informationsübertragung

Die vorangegangenen Kapitel haben gezeigt, was zu einer Implementierung von Big Data Maßnahmen im Unternehmen dazugehört und in welchen Zusammenhängen sich Schwierigkeiten aus einer solchen ergeben können. Nachdem die Ergebnisse der Datenanalyse zum Erkennen von Zusammenhängen, und dadurch zu einer möglichen Nutzung dieser Informationen für unternehmerische Zwecke geführt haben, geht es im Schritt der Post-Informationsübertragung um den Umgang mit den aufgebauten Datenbeständen und Erkenntnissen der Datenanalysen. Die Herausforderung hierbei ist, dass

ein System gefunden werden muss, welches, mit Blick auf die Charakteristika Volume, Velocity und Variety, die immer größer werdenden Datenbestände und die Erkenntnisse aus den Analysen speichern und für eine eventuelle Wiederverwendung aufbewahren kann. Die von Strohbach et al. (2016, S. 123-126) erläuterten Systeme wie HDFS oder das Cloud Computing bieten Unternehmen diese Möglichkeiten. Allerdings sind diese Technologien, wie bereits in Kapitel 3.1.2 erläutert, noch nicht perfekt entwickelt und haben beispielsweise bezogen auf ihre Speicherkapazitäten, die Handhabung von strukturierten und unstrukturierten Daten, ihre Flexibilität und Effizienz und den Schutz der individuellen Privatsphäre noch Optimierungsbedarf (Strohbach et al. 2016, S. 119).

In Anlehnung an Kapitel 3.1.2 kann an dieser Stelle auf die Pflege des Datenbestandes verwiesen werden (Freitas und Curry 2016, S. 87-88). Die Datenpflege ist zu jeder Zeit des elektronischen Wertschöpfungsprozesses von Bedeutung. Dies sollte bereits an den Kapiteln zu den Fehlerquellen in der Informationssammlung und Informationsverarbeitung deutlich geworden sein. Allerdings wird auf diesen Punkt an dieser Stelle eingegangen, damit deutlich wird, dass eine Datenanalyse nicht als ein einmaliges Durchlaufen des idealtypischen Wertschöpfungsprozesses zu sehen ist, sondern als ein Prozess, der stetig wiederholt werden sollte (Kollmann 2019, S. 63). Auch Freitas und Curry (2016, S. 87) haben auf die Wichtigkeit der aktiven und fortlaufenden Bearbeitung des Datenbestandes hingewiesen, um eine Wiederverwendbarkeit der Erkenntnisse gewährleisten zu können. Da immer neue Daten generiert werden und in die Datenanalyse einfließen beziehungsweise bisherige Ergebnisse beeinflussen können, müssen neue Daten immer mit den Alten abgeglichen werden. Dadurch kann die Aktualität der Ergebnisse gewährleistet werden (Kollmann 2019, S. 63).

Festzuhalten ist, dass datenbasierte Maßnahmen für ein Unternehmen nur dann einen Mehrwert bieten, wenn die durch die Datenanalyse gewonnenen Informationen auch kontinuierlich in unternehmerischen Entscheidungen Berücksichtigung finden (Wölfl et al. 2019, S. 223-224). Eine Verwendung von Datenanalysen nur aufgrund des Hypes rund um Big Data für Unternehmen, ohne eine systematische Planung des Einsatzes und des Zwecks dieser Maßnahme, wird für ein Unternehmen nicht von Nutzen sein (Wölfl et al. 2019, S. 223). Viele gewonnene Erkenntnisse werden vermutlich in diesem Fall

durch mangelnde Kenntnisse und Planung nicht für unternehmerische Entscheidungen verwendet werden (Wölfl et al. 2019, S. 223). Eine Durchführung von Datenanalysen mit Blick auf den zeitlichen und finanziellen Aufwand würde sich in diesem Kontext als überflüssig für das Unternehmen herausstellen. Zudem merken Wölfl et al. (2019, S. 224) in Anlehnung an Morgan et al. (2005) an, dass sich aus dem Nutzen der Erkenntnisse aus einer Datenverarbeitung kein direkter finanzieller Effekt für die Unternehmensperformance abzeichnen wird, da die Nutzung dieser Ergebnisse vor allem mittel- und langfristigen Einfluss auf die Performance eines Unternehmens hat.

3.2.4. Zusammenfassung der analytischen Erkenntnisse zu den Fehlerquellen im elektronischen Wertschöpfungsprozess

Tabelle eins sorgt für die nötige Übersicht und Systematisierung der Ergebnisse dieser Masterarbeit zu den möglichen Fehlerquellen, die während des elektronischen Wertschöpfungsprozesses in Anlehnung an Kollmann (2019) auftreten können. Diese sollten von jedem Unternehmen, das eine Big Data Maßnahme in Betracht zieht, eingehend diskutiert werden. Es muss geprüft werden, ob eine solche Maßnahme für das jeweilige Unternehmen geeignet ist und einen Nutzen zum Erreichen der Unternehmensziele beitragen kann. An dieser Stelle sei angemerkt, dass die Tabelle nicht den Anspruch der Vollständigkeit aller möglichen Fehlerquellen bei der Durchführung von Datenanalyse im E-Commerce hat, da aus Begrenzungen dieser Arbeit nur auf grundlegende Herausforderungen eingegangen werden konnte. Auf diese könnte in weitergehenden Forschungen zu der Thematik vertiefend eingegangen werden, um weitere Fehlerquellen hinzuzufügen.

	Methode/Geschäftsaktivität bzw. Technologie	Mögliche Fehlerquellen
Prä- Informations- sammlung	Rücksichtnahme auf den rechtlichen Rahmen	Unzureichende Betrachtung der Rechtsgrundlage - z.B. durch Druck in der Umsetzung Nicht rechtskonforme Entscheidungen werden im Verlauf der Datenanalyse getroffen
Informations- sammlung	Festlegung vom Zweck der Big Data Maßnahme für Unternehmensziele	Unklarheiten, inwieweit die Datenanalyse für die strategische Ausrichtung hilfreich ist Unklarheiten, welche Datensammlungen für Unternehmensziele relevant sind
	Aufbau personeller Ressourcen	Fehlende Kenntnisse und Fähigkeiten der Mitarbeiter
	Planung der Datensammlung	Fehlende systematische Planung führt zu: - Irrelevanten Daten von geringer Qualität - Übersehen von relevanten Datenquellen
Post- Informations- sammlung	Filtern der Datenmengen zur Förderung der Übersichtlichkeit der Datenanalyse	Zunehmende Unübersichtlichkeit durch fehlende Filter während der Datenanalyse
	Datenqualität vor der Analyse prüfen	Fehlende Überprüfung des durch die Sammlung erhaltenen Dateninputs auf Vollständigkeit, Konsistenz und Aktualität
	Auswertung der gesammelten Daten und Aufbau eines möglichst homogenen Datenbestandes zur Analyse	Datenquellen liefern nicht alle dieselbe Form der Daten - Fehlende Homogenität der Daten kann zu Problemen in der Analyse führen
Prä- Informations- verarbeitung	Aufbau einer geeigneten Infrastruktur zur Datenverarbeitung	Fehlende Mittel für den Aufbau der benötigten Strukturen - Notwendigkeit von kritisch zu betrachtenden Kooperationen (siehe Informationsverarbeitung)
Informations- verarbeitung	Algorithmen für die Datenverarbeitung z.B. durch ML oder DL	Fehlender Erkenntnisgewinn durch unzureichenden Kenntnisse beim Umgang mit Datenströmen Effektive Verarbeitung hochdimensionaler Daten schwierig Unzureichender Input für die AI als Lerngrundlage
	Kooperation mit Drittanbietern	Fehlendes Vertrauensverhältnis Fehlende Transparenz in der Verarbeitung durch Drittanbieter

	Methode/Geschäftsaktivität bzw. Technologie	Mögliche Fehlerquellen
Post- Informations- verarbeitung	Umgang mit Datenanalyseergebnissen	Unklarheit über personelle Zuständigkeiten in der Datenverarbeitung Gruppendynamik muss stimmig sein
Prä- Informations- übertragung	Datenanalyseergebnisse überprüfen, verifizieren und Zusammenhänge erkennen	Bugs im Analysesystem Fehlerhafte Annahmen im Analyseprozess Fehlerhafte Daten zur Analyse Masse an Analyseergebnissen können zur Überforderung der befugten Personen im Interpretationsprozess führen
	Algorithmen auf Grundlage der individuellen Präferenzen	Endnutzer können in eine Filterblase geraten, die ihre Wahrnehmung verzerrt
	Visualisierung der Ergebnisse	Ungeeignete Visualisierung erschwert Interpretation
Informations- übertragung	Nutzung der aus den Datenanalyseergebnissen gewonnenen Informationen zu Datenzusammenhängen	Unübersichtlichkeit durch Masse an Daten erhöht die Wahrscheinlichkeit eines Fehlers - z.B. Amazon Kunde erhält falsche Dateien
	Schnelle Sammlung, Verarbeitung und Übertragung der Daten	Aktualität der Daten und Informationen ist kritisch zu betrachten - Zalandos Kaufempfehlungen z.T. veraltet
Post- Informations- übertragung	Datenspeicherung	Unzureichende Systemlösungen mit Blick auf die Merkmale Volume, Velocity und Variety
	Pflege des Datenbestandes	Fehlende Datenbestandspflege gefährdet Aktualität der Daten und Ergebnisse
	Evaluation des Nutzens der Datenanalyse	Kein Nutzen durch unzureichende Planung der Maßnahme - Erkenntnisse werden nicht für unter- nehmerische Zwecke verwendet - Maßnahme stellt sich vor dem Hintergrund des finanziellen und zeitlichen Aufwands als nicht sinnvoll heraus

Tabelle eins: Zusammenfassung der analytischen Ergebnisse von Fehlerquellen im elektronischen Wertschöpfungsprozess in Anlehnung an Kollmann (2019)

3.3. Maßnahmen einer Fehlerprävention

3.3.1. Maßnahmen während der Informationssammlung

Kapitel 3.3 bietet einen Überblick zu Fehlerpräventionsmaßnahmen bei Datenanalysen im Bereich des E-Commerce. Aus wirtschaftlicher Sicht ist es aus Kosten- und Zeitgründen sinnvoll, die Fehlerrate in einem Unternehmen auf das Minimum zu reduzieren (siehe Kapitel 2.1.3). Anhand der drei Stufen des elektronischen Wertschöpfungsprozesses nach Kollmann (2019, S. 61) soll in Kapitel 3.3 ein Augenmerk auf präventive unternehmerische Möglichkeiten mit Blick auf die in der Analyse herausgestellten Fehlerquellen gelegt werden, um die Wahrscheinlichkeit dieser Fehlerquellen in dem jeweiligen Unternehmen zu reduzieren.

Die Rechtsgrundlagen sollte mit Blick auf datenbasierende Geschäftsaktivitäten in regelmäßigen Abständen überprüft werden (BITKOM 2014, S. 144), um etwaigen rechtlichen Abweichungen vorzubeugen. Risikovermeidungsmaßnahmen könnten laut der BITKOM (2014, S. 144) hierbei zum Beispiel Hinweisgebersysteme, Anonymisierungen, Aggregationen von Daten zu Metadaten, Pseudonymisierungen oder Schulungen der Mitarbeiter zum Datenschutz sein. Zudem ist es wichtig, dass Daten aus den verwendeten Systemen im Falle einer Löschung komplett gelöscht, diese durch Löschprotokolle nachgewiesen und nicht im Hintergrund archiviert werden, da archivierte Daten ein hohes Risikopotential bergen (BITKOM 2014, S. 144). Das Risiko wird zum Beispiel bezogen auf die DS-GVO deutlich, wenn es um Löschfristen der Daten beziehungsweise um das individuelle Recht des Vergessenwerdens geht (Bundesministerium für Wirtschaft und Energie o.D., S. 1).

Ohne einen gut geplanten und durchdachten strategischen Ansatz lässt sich aus einer Big Data Maßnahme laut Wölfl et al. (2019, S. 225) kein Potenzial schöpfen. Daher sollten Unternehmen nicht erst über einen möglichen Einsatz der aus den Datenanalysen gewonnenen Informationen nachdenken, wenn diese vorliegen, sondern sie sollten sich bereits zu Beginn der Maßnahme im Klaren darüber sein, welches Ziel und welchen Zweck die systematische Datenverarbeitung für das Unternehmen haben soll (Wölfl et al. 2019, S. 225). In diesem Zusammenhang machen Wölfl et al. (2019, S. 225) deutlich, dass Unternehmen am ehesten einen Mehrwert durch datenbasierte

unternehmerische Entscheidungen erreichen werden, wenn diese mit einem kundenzentrierten Einsatz der Daten verbunden sind. In Anlehnung an Schroeck et al. (2012) erläutern sie, dass durch eine Durchführung von Datenanalysen die Möglichkeit einer systematischen Erfassung des Kundenverhaltens entsteht, wodurch die Unternehmen dann Kundenwünsche ableiten und entsprechend dieser Wünsche ihre Leistungen möglichst schnell ausrichten können (Wölfl et al. 2019, S. 225).

Wölfl et al. (2019, S. 225-226) machen in ihren Ausführungen deutlich, dass eine Datenverarbeitung, und die damit zusammenhängenden datenbasierten Entscheidungen vom Top Management eines Unternehmens vorangetrieben werden müssen, um diese Maßnahmen im Unternehmen zu festigen und die Transparenz innerhalb des Unternehmens durch das Teilen von Daten über Funktionsgrenzen hinweg zu erhöhen. Zudem erklären sie, dass die Errichtung einer zentralen organisationalen Einheit Sinn ergibt, „welche die Verantwortlichkeit für Projekte der zielgerichteten Datennutzung hat.“ (Wölfl et al. 2019, S. 226). In dieser zentralen Einheit sollten Mitarbeiter vorhanden sein, die über ein gutes und kompetentes Fachwissen, Kenntnisse und Fähigkeiten rund um Big Data verfügen. Dadurch können sie die unterschiedlichen Funktionsbereiche des Unternehmens in der Datennutzung unterstützen, Synergieeffekte zwischen den Bereichen schaffen und den Gesamtüberblick über die Datenverarbeitung behalten (Wölfl et al. 2019, S. 226). Mit Blick auf die Thematik dieser Arbeit, wäre dies dann auch die zentrale Abteilung innerhalb des Unternehmens, die sich mit der Fehlerprävention und dem Fehlermanagement bezogen auf die Datenanalysen auseinandersetzt und so einen optimalen Umgang mit Fehlern in diesem Unternehmensbereich gewährleisten kann.

Die zusammenfassende Tabelle in Kapitel 3.2.4 zeigt zudem, wie wichtig der Aufbau eines qualitativ hochwertigen Datenbestandes von Beginn an ist. Qualitativ hochwertig heißt in diesem Zusammenhang auch, dass durch diesen Bestand möglichst viele der erläuterten potenziellen Fehlerquellen vermieden werden. Um diesen Datenbestand aufzubauen, ist unter anderem die Auswahl der Datenquellen von Bedeutung. Dabei sollte die Datenqualität mit Blick auf die Konsistenz, Aktualität und Korrektheit dieser geprüft werden (BITKOM 2014, S. 143; Liggesmeyer et al. 2014, S. 110), sodass die Datenanalyse und die hieraus gewonnenen Ergebnisse eine solide Datengrundlage haben. Hierfür

ist laut Wölfl et al. (2019, S. 225) auch die Erarbeitung einer Data Governance sinnvoll. Diese legt fest, „welche Daten erhoben, gespeichert, analysiert und von wem verwendet werden.“ (Wölfl et al. 2019, S. 225) Zudem wird sichergestellt, welche Daten qualitativ hochwertig und von hohem Nutzen für das Unternehmen sind. Hinsichtlich der Ausführungen in Kapitel 3.2.1.3 heißt eine hohe Qualität des Datenbestandes auch, dass heterogene Datensammlungen durch Datenanalysten in eine homogenere Form gebracht werden, da eine fehlende Homogenität zu Problemen während der Datenanalyse führen kann (Jagadish et al. 2014, S. 89). Es kann zudem nicht davon ausgegangen werden, dass alle Daten gleich komplex sind, wodurch die Herausforderung entsteht, dass auch komplizierte Datentypen, wie Graphiken oder ähnliches, durch Big Data Technologien verarbeitet werden können (Domingue et al. 2016, S. 74).

3.3.2. Maßnahmen während der Informationsverarbeitung

Das zentrale Objekt einer erfolgsversprechenden Datenverarbeitung ist die Etablierung einer geeigneten Infrastruktur. Hierzu gehört auch die technische Infrastruktur des Unternehmens. Der Begriff Big Data umfasst in der Regel eine Kombination aus vielen unterschiedlichen, technischen Architekturen (BITKOM 2014, S. 145). Diese Kombination setzt sich aus den Technologien, die für die Durchführung des beschriebenen elektronischen Wertschöpfungsprozesses nach Kollmann (2019, S. 61) notwendig sind, zusammen. Hierunter fallen technische Möglichkeiten im Bereich der AI zur Echtzeitverarbeitung von Datenflüssen, technische Systeme zur Erhebung von Daten, Systeme zur Speicherung und viele mehr, auf die bereits an anderer Stelle eingegangen wurde. Die Herausforderung der Kombination dieser unterschiedlichen technischen Systeme besteht ebenfalls in den zunehmenden technischen Neuerungen (BITKOM 2014, S. 145), die in die Unternehmensstrukturen eingefügt werden sollten, um die Datenverarbeitung aktuell zu halten. Im Fall einer Einführung von künstlichen Intelligenzen in die Datenanalyse wäre es in Anlehnung an die BITKOM (2014, S.145) sinnvoll, wenn die neueren Systeme zunächst einzelne Analyseaufträge der älteren Systeme übernehmen oder ergänzen und sich dann zunehmend zu der Verarbeitung der sehr großen Datenmengen steigern. Dies sorgt auch dafür, dass keine Überforderung durch fehlende Kenntnisse und Fähigkeiten bezogen auf die neuen Systeme eintritt, da sich das Unternehmen zunächst an die Möglichkeiten und Arbeitsabläufe der

Technologien herantastet. Zu einer geeigneten Infrastruktur gehören außerdem auch die personellen Ressourcen. Damit bei den befugten Personen keine Unklarheiten über die Zuständigkeiten während der Datenverarbeitung bestehen, ist der gezielte und geplante strategische Einsatz einer Big Data Maßnahme in einem Unternehmen und ein zentraler Arbeitsbereich für die Datenverarbeitungsmaßnahmen von großer Bedeutung (Wölfl et al. 2019, S. 225-226). Diesen Punkt greift auch die BITKOM (2014, S. 144) in ihren Ausführungen auf, da erläutert wird, dass klare Formulierungen in Bezug auf die allgemeinen Verantwortlichkeiten und Aufgaben der beteiligten Personen und auf vorhandene Schnittstellen, Zugriffsrechte und Nutzerregelungen gegeben sein müssen.

Gentsch (2019, S. 13) erläutert, dass Daten keinen Mehrwert erbringen, sondern, dass erst die Algorithmen, egal ob vordefinierte Mechanismen oder selbstlernende Systeme, einen Wert aus den Daten generieren können. Bezogen auf diese neueren technischen Möglichkeiten der Datenverarbeitung durch Algorithmen, ist die Herangehensweise der Nutzung dieser entscheidend. Najafabadi et al. (2015, S. 17) erläutern in diesem Zusammenhang, dass die generelle Vorgehensweise bei diesen Algorithmen darin besteht, diese zunächst durch kleinere Mengen des vorhandenen Dateninputs für die antizipierte Datenrepräsentation zu trainieren und nicht direkt den gesamten Dateninput zu verwenden (Najafabadi et al. 2015, S. 17). Die anhand dieses Trainings von dem Algorithmus gelernten Muster können dann auf den verbleibenden Dateninput angewandt werden (Najafabadi et al. 2015, S. 17). In diesem Zusammenhang muss herausgefunden werden, welcher Dateninput für ein Training der Algorithmen sinnvoll ist, um eine nützliche Datenrepräsentation zu schaffen, die auch auf spätere Datenmengen in den unterschiedlichen Anwendungsbereichen von Big Data Maßnahmen wieder angewendet werden kann (Najafabadi et al. 2015, S. 17).

Wölfl et al. (2019, S. 221) haben in Anlehnung an Klein und Rai (2009) erläutert, dass eine zentrale Herausforderung im Aufbau von vertrauensvollen Geschäftsbeziehungen besteht. Da einige E-Commerce Händler, wie das Beispiel von Fahrrad XXL (o.D.) gezeigt hat, auf Kooperationen mit Drittanbietern zur Datenverarbeitung angewiesen sind, ist Vertrauen in diesen Beziehungen ein wichtiger Faktor. Um den Kooperationspartnern nicht blind zu vertrauen und auch, um die Kontrolle über Datenanalysen im eigenen Unternehmen zu wahren, bietet sich während des Prozesses der

Datenwertschöpfung immer eine externe Überwachung dieser an, damit Datenmissbrauch und Manipulationen der Datenanalyseergebnisse vermieden werden (BITKOM 2014, S. 144). Außerdem legt die BITKOM (2014, S. 144) nahe, dass der Projektprozess kontinuierlich von einem Betriebsrat und einem Datenschutzbeauftragten geprüft wird, um in den Prozessen rechtskonform zu bleiben. Die Prüfungen haben zudem den Nutzen, dass durch diese Schwachstellen in den internen Prozessen aufgedeckt werden können und die Prozessplanung somit effizienter gestaltet werden kann (BITKOM 2014, S. 144).

Domingue et al. (2016, S. 74-75) stellen außerdem heraus, welchen Anforderungen die nächsten Generationen von Big Data Technologien gewachsen sein müssen. Hierunter fällt zum Beispiel der Umgang mit dem stetigen Wachstum des Internets, da immer mehr Menschen auf Informations- und Kommunikationstechnologien zurückgreifen werden und dadurch auch immer mehr Daten produziert werden (Domingue et al. 2016, S. 74). Neuere Technologien sollten eine Echtzeitverarbeitung der Daten ermöglichen, da ältere Technologien, wie auch Hadoop (Lyko et al. 2016, S. 45), darauf ausgelegt waren „batches of historical data“ zu analysieren (Domingue et al. 2016, S. 75). Hier merken Domingue et al. (2016, S. 75) an, dass ein solches System für Echtzeitverarbeitungen von Daten das bereits angesprochene Storm wäre, welches aber noch Entwicklungsbedarf aufweist (Domingue et al. 2016, S. 75). Eine gleichzeitige Verarbeitung von großen Datenquantitäten stellt Technologien ebenfalls vor eine Herausforderung (Domingue et al. 2016, S. 75). Außerdem sind viele der vorhandenen Technologien nach den Ausführungen von Domingue et al. (2016, S. 75) nicht kompatibel mit Cloud-Systemen. Die Synchronizität der Daten über unterschiedliche Datenbanken zu gewährleisten birgt laut ihnen ebenfalls ein Problemfeld. Eine effiziente Verschlagwortung im Kontext mit Big Data Technologien ist darüber hinaus wichtig, um Dokumente und Datensammlungen besser handhaben zu können und den Überblick zu behalten (Domingue et al. 2016, S. 75).

3.3.3. Maßnahmen während der Informationsübertragung

Wie bereits festgehalten wurde, ist eine zentrale Abteilung in Unternehmen sinnvoll, um für eine nötige Personalqualität und eine ausreichende Transparenz der Datenverarbeitung und dadurch auch der Datenquellen zu sorgen (Wölfl et al. 2019, S. 226). Außerdem kann im Zusammenhang mit der Pflege des Datenbestandes eines Unternehmens festgehalten werden, dass eine neue Generation von Arbeitsplätzen notwendig ist, „where community participants can annotate base data with interpretation metadata, resolve their disagreements and clean up the dataset, while partially clean and partially consistent data may still be available for inspection.“ (Jagadish et al. 2014, S. 90) Um die Ergebnisse der Datenanalyse richtig zu interpretieren, sollten hierzu befugte Personen über die Funktionsweise der in der Datenverarbeitung verwendeten Algorithmen aufgeklärt werden, da dies laut der BITKOM (2014, S. 144) Fehlinterpretationen vorbeugen kann. An dieser Stelle kann auf das in Kapitel 2.1.2.3. verwiesene Interpretationsrisiko verwiesen werden. Fehlinterpretationen sind laut der BITKOM (2014, S. 142) in diesem Zusammenhang vor allem durch unzureichend qualifizierte Mitarbeiter in Kombination mit einer unverständlichen Visualisierung der Ergebnisse möglich.

Ziel der Datenverarbeitung ist es, wie bereits an mehreren Stellen gezeigt wurde, dass eine personalisierte Ansprache des Endnutzers ermöglicht wird, die auf den individuellen Vorstellungen und Wünschen des Nutzers beruht. Gerade deshalb sollte bei einer Datenverarbeitung der Endnutzer mit in den Prozess einbezogen werden, um sich im Klaren über dessen Fähigkeiten und Bedürfnisse zu sein und den Mehrwert der Datenverarbeitung zu erhöhen (Wölfl et al. 2019, S. 226). In Anlehnung an He und King (2008) erläutern Wölfl et al. (2019, S. 226), dass mehrere Studien gezeigt haben, dass die Einbeziehung der Anwender dazu beiträgt, dass diese mit dem letztlich erstellten System zufriedener sind, es häufig nutzen und dieses dementsprechend mehr Erfolg hat. Ein Beispiel eines solchen Systems wäre das von Jonathan Goldman entwickelte LinkedIn-Feature, wodurch dem Nutzer Personen angezeigt werden, die dieser aufgrund seiner preisgegebenen personenbezogenen Daten eventuell kennen könnte (Davenport und Patil 2012, S. 72). Dadurch wird deutlich, dass eine Einbeziehung der Endnutzer es in einem solchen Fall der zielgerichteten Datennutzung vereinfacht „die Systeme, Tools und Anwendungen so auszugestalten, dass sie den

Anforderungen und Bedürfnissen der Datennutzer entsprechen.“ (Wölfl et al. 2019, S. 226). So kann sichergestellt werden, dass die Ergebnisse der Datenverarbeitung von den Endnutzern als relevant wahrgenommen werden und sie keine Probleme in der Bedienung der entwickelten Systeme und Tools haben werden (Wölfl et al. 2019, S. 226-227). Diese problemlose Anwendung ist laut Wölfl et al. (2019, S. 227) in Anlehnung an Davis (1989) ein wichtiger Punkt in der Akzeptanz von Informationssystemen. Wölfl et al. (2019, S. 227) merken an, dass die zahlreichen existierenden Anwendungen zur Datenanalyse an die Bedürfnisse der Endnutzer angepasst werden sollten, um ein Verständnis der Datenanalyseergebnisse durch die Nutzer gewährleisten zu können und die Relevanz der Ergebnisse zu erhöhen.

Zum Abschluss dieses Kapitels kann angemerkt werden, dass der Einsatz von Big Data Technologien allerdings nicht nur Risiken bergen kann, denen präventiv entgegengewirkt werden muss, sondern auch zur Vorhersage von Risiken für das Unternehmen genutzt werden kann (Omri 2015; Walsh et al. 2013). Eine informierte und transparente Entscheidungsfindung führt beispielsweise auch dazu, dass auffällige Geschäftsrisiken frühzeitig erkannt werden können. Effiziente Vorhersagemodelle entstehen in diesem Zusammenhang, „wenn Daten aus den verschiedenen Abteilungen und Informationsquellen intelligent verknüpft und analysiert werden.“ (Omri 2015, S. 110) Somit bieten Big Data Technologien, in Anlehnung an Kapitel 2.2.4.2., prinzipiell Möglichkeiten, das Risikomanagement des Unternehmens effizienter und Geschäftsaktivitäten weniger risikoreich zu machen. Laut Omri (2015, S. 110) können durch die intelligente Verknüpfung von verschiedenen Informationsquellen und Abteilungen nicht nur Betrugsfälle frühzeitig erkannt und vorhergesagt, sondern auch eine Kundenabwanderung erkannt und Risiko-Vorhersagemodelle erstellt werden. Zudem können unter anderem Szenarien und Prognosen gebildet, transparente Risikoanalysen durchgeführt, Kreditwürdigkeiten überprüft und leistungsfähige Werttreibermodelle erstellt werden (Omri 2015, S. 110). Walsh et al. (2013, S. 52) erläutern, dass es für Unternehmen durch die Nutzung von Big Data auch zu Kostensenkungen kommen kann. Im Risikomanagement können zum Beispiel während der Kaufphase Kosten gesenkt werden, da auch Walsh et al. (2013, S. 52) festhalten, dass in diesem Bereich Risiken besser vorhergesagt werden können und ein Missbrauchs- oder Betrugsversuch besser erkannt werden kann. Gerade für das E-Commerce bieten sich

hier vielfältige Möglichkeiten für die Verknüpfung von Daten aus mehreren Quellen, wie zum Beispiel der Onlineverlaufshistorie, dem Kundennutzungsverhalten des Onlineshops oder den Social Media Einträgen.

3.4. Kommunikation in betrieblichen Krisensituationen als Teil des Fehlermanagements

Wie die Beispiele in der vorangegangenen Analyse gezeigt haben, ergeben sich im Zuge der Nutzung von Big Data Maßnahmen für das Unternehmen mehrere Fehlerquellen. Diese können, je nach den Auswirkungen des Fehlers, schwerwiegende Konsequenzen für das Unternehmen haben. Daher ist es im Rahmen des Fehlermanagements von Bedeutung, dass ein Unternehmen, in einer solchen betrieblichen Krisensituation durch ausschlaggebende fehlerhafte Handlungen, angemessen reagiert und den Schaden dadurch so gut wie möglich eingrenzen kann.

Laut Ehmke (2019, S. 116) gibt es keine einheitliche Definition des Begriffes der Unternehmenskrise. Allerdings wurde in der betriebswirtschaftlichen Literatur eine Bedeutungsverengung durchgesetzt, damit nicht jede „gesamtwirtschaftliche Entwicklungsunregelmäßigkeit als Krise definiert wird“ (Ehmke 2019, S. 116). Hierdurch wird laut Ehmke (2019, S. 116) nur dann von einer Krise gesprochen, wenn eine konkrete Existenzgefährdung durch das Verfehlen von Unternehmenszielen eines Unternehmens besteht. Für Ehmke (2019, S. 117) ist zusätzlich der Aspekt des Unternehmensimages in diesem Zusammenhang von Bedeutung, welches durch unterschiedliche Faktoren geschädigt werden kann. Durch eine Schädigung des Unternehmensimages kann dann im schlimmsten Fall die Existenz des Unternehmens gefährdet werden (Ehmke 2019, S. 117). Krystek (2002, S. 89) definiert eine Unternehmenskrise in Anlehnung an gemeinsame Elemente verschiedener Definitionen zum Krisenmanagement als „ungeplante und ungewollte Prozesse von begrenzter Dauer und Beeinflussbarkeit sowie mit ambivalentem Ausgang.“ (Krystek 2002, S. 89) Auch er spricht von einer Gefährdung, die das Unternehmen nachhaltig und in seiner Substanz beeinflussen oder sogar zur Geschäftsaufgabe führen kann. Nichtsdestotrotz merkt Ehmke (2019, S. 118) an, dass eine Krise sehr stark von der individuellen Beurteilung abhängt, da eine Situation für den einen als Krise gesehen werden kann, für den anderen aber zum Beispiel nur als Teil

des normalen Geschäftsverlaufs. Eine Krise bedeutet für das Unternehmen laut Ehmke (2019, S. 133) allerdings nicht zwangsläufig das Ende des Unternehmens. Diese kann auch als Chance dienen, „innovative Änderungen und Konzeptionen herbeizuführen“ (Ehmke 2019, S. 133), was sich mit den Ausführungen zum Fehlermanagement in Kapitel 2.1.3 deckt, da in einer kritischen Situation die Risikobereitschaft eines Unternehmens wächst (Ehmke 2019, S. 133). Auch Krystek (2002, S. 97) greift diesen Punkt auf und erläutert, dass Unternehmen oft erst nach einer bewältigten Krise in der Lage sind, Änderungen im Unternehmen durchzusetzen, neue, zukunftsfähige Konzepte zu entwickeln und mit veralteten Strukturen zu brechen.

Im Falle einer Unternehmenskrise wächst die Bedeutung der Kommunikation (Ehmke 2019, S. 133). Es kann herausgestellt werden, dass die Kommunikation in einer solchen Situation ein bedeutendes Element ist, um den Verlauf einer Krise durch Gegen-darstellung, Dementis oder mit anderen kommunikativen Mitteln zu beeinflussen und so die Krise erfolgreich zu bewältigen (Ehmke 2019, S. 135). Das wichtigste Ziel dieser Krisenkommunikation ist nach dem Erkennen der Krise der Erhalt des Vertrauens in das Unternehmen (Ehmke 2019, S. 135). Dieses sollte allerdings nicht erst mit Beginn der Krise aufgebaut, sondern bereits im Stadium von potenziellen Krisen realisiert werden, da eine kurzfristige kommunikative Maßnahme zur Krisenbewältigung aller Wahrscheinlichkeit nach scheitern wird (Ehmke 2019, S. 135). Im Falle einer Krise ergeben sich laut Ehmke (2019, S. 136) für das Unternehmen zwei Möglichkeiten: Entweder das Unternehmen schweigt zu der Situation oder kommuniziert mit der Öffentlichkeit. Als Lösungsvorschlag rät Ehmke (2019, S. 136-137) zu einer Offenheit und Transparenz über die unternehmerische Situation gegenüber der Öffentlichkeit, die entweder durch eine defensive oder offensive Kommunikationsstrategie realisiert werden kann. Defensiv heißt in diesem Zusammenhang, dass Informationen zur Krise fragmentarisch weitergegeben oder die Krise geleugnet wird (Ehmke 2019, S. 137). Offensiv bedeutet, dass die Öffentlichkeit durch völlige Transparenz und Ehrlichkeit des Unternehmens bereits in einem frühen Stadium über die unternehmerische Krisensituation in Kenntnis gesetzt wird (Ehmke 2019, S. 138).

Um Krisen im Unternehmen zu vermeiden, sollte die Unternehmensführung bereits frühzeitig Überlegungen zu möglichen Unternehmenskrisen durchführen (Krystek 2002,

S. 101). Dazu gehört die Prognose von potenziellen Gefährdungsbereichen des Unternehmens durch verschiedene Techniken, die unter anderem durch Zukunftsszenarien aufzeigen, wo solche Bereiche im Unternehmen vorhanden sind (Krystek 2002, S. 101). Auch das in Kapitel 2.2.4.2 ausgeführte Risikomanagement ist Teil der Vermeidung von Unternehmenskrisen. Wenn es sich um versicherbare Risiken handelt, dann liegt laut Krystek (2002, S. 101) die versicherungstechnische Absicherung gegen dieses Risiko nahe, um der Unternehmenskrise vorzubeugen. Allen Risiken, gegen die sich das Unternehmen nicht versichern kann, sollten laut Krystek (2002, S. 101) mit einer Alternativplanung begegnet werden, die aus Notfallplänen und einer Eventualplanung besteht.

4. Fazit und Ausblick

Zusammenfassend kann festgehalten werden, dass eine zielgerichtete Datenverarbeitung in einem Unternehmen der heutigen digitalen Gesellschaft als Wettbewerbsvorteil, aber auch als eine Notwendigkeit zum Überleben des Unternehmens gesehen werden kann, da der zunehmende technische Fortschritt den Lebensalltag der Menschen und die Arbeitsweise der Unternehmen verändert hat und weiterhin verändern wird (Wölfl et al. 2019,

S. 227). Etablierte Unternehmen werden in diesem Kontext von neuen aufstrebenden Start-Ups bedroht (Wölfl et al. 2019, S. 227), die auf den in Kapitel 2.3.2 erläuterten datenzentrischen Geschäftsmodellen aufbauen. In der Analyse beschriebene Technologien, die, wie die Open Source Software Hadoop, öffentlich verfügbar sind, können von Unternehmen genutzt werden, um in dieser Wettbewerbslandschaft durch datenbasierte Entscheidungen weiterhin erfolgreich am Markt mitwirken zu können (Wölfl et al. 2019, S. 227). Wie die Beispiele der Analyse gezeigt haben, versuchen die betrachteten E-Commerce Händler ihre Geschäftsaktivitäten durch Datenanalysen effizienter und effektiver zu gestalten. Allerdings müssen für ein Erreichen von mehr Effizienz und Effektivität, die in Kapitel 3.2.4 durch die Analyse herausgestellten Fehlerquellen von dem jeweiligen Unternehmen beachtet werden und die Herausforderung einer erfolgversprechenden Datenverarbeitung im Bereich des E-Commerce angenommen werden. Die zusammenfassende Tabelle in Kapitel 3.2.4 zeigt, dass die Fehlerursachen nicht nur in den Verarbeitungsprozessen der Daten zu finden sind, sondern auch durch mangelndes

Verständnis und fehlenden Kompetenzen und Fähigkeiten der Mitarbeiterinnen und Mitarbeiter, oder auch eine unzureichende strategische Planung der Maßnahme begründet sein können. Hierbei erhebt diese Masterarbeit nicht den Anspruch, dass die ermittelten Fehlerquellen und Präventionsmaßnahmen für alle möglichen Big Data Maßnahmen zutreffend sind. Vielmehr vermittelt diese Arbeit einen Überblick über zentrale Herausforderungen bei der Durchführung von Datenanalysen mit Beispielen aus dem Bereich des E-Commerce, die durch eine fehlende Auseinandersetzung mit den Herausforderungen zu Fehlerquellen für Unternehmen werden können. Hierdurch wird der Aspekt zur Ansiedlung eines Fehlerbewusstseins der Unternehmen für die erläuterten Maßnahmen verdeutlicht, um das maximale Potenzial aus Datenverarbeitungen schöpfen zu können, und die Fehlerrate durch einen optimalen Umgang mit Fehlern auf Grundlage einer Fehlerprävention und eines Fehlermanagements auf das Minimum zu reduzieren. In diesem Zusammenhang soll der Faktor der Transparenz nochmals herausgestellt werden und die Bedeutung davon, dass Unternehmen ihre Datenverarbeitungsprozesse zukünftig transparenter gestalten, um zum Beispiel Datenherkünfte und darauf aufbauende Fehlerquellen besser ausfindig machen zu können.

Die Einführung der neuen DS-GVO im Jahr 2018 (Bundesministerium für Wirtschaft und Energie o.D.) macht deutlich, dass die Gesellschaft in der Zeit der Digitalisierung unter anderem durch zunehmende Datenerhebungen immer mehr nach einem Schutz der individuellen Privatsphäre verlangt. Wie die Betrachtung von Artikel sechs Absatz eins aus der DS-GVO gezeigt hat, ist der Schutz allerdings eher oberflächlich (Europäisches Parlament 04.05.2016, S. 36). Je nach Auslegung der „Wahrung der berechtigten Interessen des Verantwortlichen oder eines Dritten“ (Europäisches Parlament 04.05.2016, S. 36), bietet sich hier den Unternehmen die Möglichkeit der Erläuterung, dass eine Datenverarbeitung zum Zwecke einer verbesserten Lieferung im Interesse des Kunden ist, wie das Beispiel von IKEA zeigt (o.D., Kapitel 4.2). Fraglich bleibt, wie die Gesellschaft auf diese Tatsachen reagiert und ob der Mehrwert durch Datenanalysen auch von den Kunden wahrgenommen wird, oder ob sie diesem eher skeptisch begegnen. In diesem Zusammenhang stellt sich auch die Frage nach der menschlichen Reaktion auf den zunehmenden Einsatz von künstlichen Intelligenzen bei der Datenverarbeitung.

Zudem ist eine Betrachtung der Zukunftstechnologien im Bereich der Datenverarbeitung von Interesse, da der schnelle technische Wandel durch stetige Neuerungen eine aktuelle Betrachtung der technischen Möglichkeiten bedarf, welche in ein paar Jahren nicht mehr durch diese Masterarbeit gegeben wäre. Gentsch (2019, S. 23) merkt mit Blick auf Zukunftstechnologien an, dass momentan weltweit an einer Ablösung der Von-Neumann-Architektur gearbeitet wird, auf der alle modernen Rechner aufbauen. Ziel ist es, durch die Neurowissenschaft inspirierte Strukturen zu erstellen, die auf der Funktionsweise des Gehirns basieren (Gentsch 2019, S. 23). Als Beispiel hierfür nennt er den Bereich der neuromorphen Chips, wobei es sich um elektronische Schaltungen handelt, die von Nervenzellen inspiriert sind (Gentsch 2019, S. 23). Bisher werden diese Chips nur in der Forschung verwendet, haben aber laut Gentsch (2019, S. 23) das Potenzial, den Bereich der AI entscheidend zu beschleunigen. Sie eignen sich zur „Pattern Recognition sowohl bei strukturierten (z. B. Tabellen, Datenbanken) als auch unstrukturierten Daten (z. B. Text-, Bild-, Audio- und Video-Dateien).“ (Gentsch 2019, S. 23) Außerdem erwähnt er die Technologie des Quantencomputers, die auf Quantenbits basiert und dadurch eine gleichzeitige Speicherung von exponentiell anwachsenden Anzahlen digitaler Werte ermöglicht (Gentsch 2019, S. 23). Durch eine zusätzliche vorhandene Verknüpfung der Quantenbits führt diese Struktur dazu, dass diese parallel arbeiten und sich die Rechenleistung dieses Computers exponentiell mit der Anzahl der Quantenbits erhöht (Gentsch 2019, S. 23). Allerdings stehen Forscher hier auch noch vor vielen Herausforderungen, da der Quantencomputer eine ausreichende Kühlung benötigt, um die Rechenleistung zu erbringen, und die Programmierung des Computers ganz anders ist als bei einem herkömmlichen Computer (Gentsch 2019, S. 23-24). Für den Bereich der AI und in diesem Zusammenhang auch für den Bereich der Datenverarbeitungen bergen diese Technologien allerdings mit Blick auf die stetig wachsende Datenmenge zukünftig große Potentiale und eine erste Anwendung des MLs wurde bereits auf einem Quantencomputer getestet (Gentsch 2019, S. 24).

Literaturverzeichnis

- Amazon (2019)*: Amazon Startseite. <https://www.amazon.de/> (Zugriff: 28.08.2019).
- Augsburger Allgemeine (2019)*: Datenschutz als Problem. <https://www.augsburger-allgemeine.de/augsburg/Datenschutz-als-Problemid54665741.html> (Zugriff: 28.08.2019).
- Bannerman, P. L. (2008)*: Risk and risk management in software projects. A reassessment. In: *Journal of Systems and Software* 81, 2008, 12, S. 2118–2133.
- Barney, J. B. (1991)*: Firm Resources and Sustained Competitive Advantage. In: *Journal of Management* 17, 1991, 1, S. 99–120.
- Barney, J. B.; Arikian, A. M. (2001)*: The resource-based view. Origins and implications. In Hitt, M.A.; Freeman, R.E.; Harrison, J.S. (Hrsg.): *The Blackwell handbook of strategic management*. Oxford [u.a.]. S. 124-188.
- Barney, J. B.; Hesterly, W. S. (2008)*: *Strategic management and competitive advantage. Concepts*. 2. ed. Upper Saddle River, NJ.
- Becker, T. (2016)*: Big Data Usage. In: J.M. Cavanillas, E.Curry und W. Wahlster (Hrsg.): *New Horizons for a Data-Driven Economy*. Cham. S. 143–165.
- Becker, W. (2019)*: Digitale Transformation von Geschäftsmodellen – Ein konzeptioneller Bezugsrahmen. In: W. Becker, B. Eierle, A. Fliaster, B. Ivens, A. Leischnig, A. Pflaum und E. Sucky (Hrsg.): *Geschäftsmodelle in der digitalen Welt*. Wiesbaden. S. 15–33.
- Beuth, P. (2014)*: Schwanger ohne digitale Spuren. Eine Soziologin wollte verhindern, dass werbetreibende Firmen von ihrer Schwangerschaft erfahren. Es ähnelte dem Versuch, eine Straftat zu vertuschen. <https://www.zeit.de/digital/datenschutz/2014-04/big-data-schwangerschaft-verheimlichen> (Zugriff: 28.08.2019).
- Beuth, P. (2019)*: Ein Milliarden Mückenstich. US-Handelsbehörde gegen Facebook. <https://www.spiegel.de/netzwelt/web/facebook-und-die-ftc-fuenf-milliarden-dollar-mueckenstich-a-1277192.html> (Zugriff: 28.08.2019).

- bevh* (2019): Pressemitteilung. E-Commerce Umsätze für 1. Quartal 2019 übertreffen Erwartungen - Top-3 Umsatzgewinner sind Lebensmittel, Haushaltswaren sowie Computer. <https://www.bevh.org/presse/pressemitteilungen/details/e-commerce-umsaetze-fuer-1-quartal-2019-uebertreffen-erwartungen-top-3-umsatzgewinner-sind-lebensmit.html> (Zugriff: 28.08.2019).
- Bischof, C.; Gabriel, M.; Rabel, B.; Wilfinger, D.* (2016): Strategic Implications of BIG DATA – A Comprehensive View. In Laporšek, S.; Gomezelj Omerzel, D. (Hrsg.): *Managing Global Changes*. Pula, Kroatien, 1 bis 4 Juni 2016; Proceedings. Koper, Slovenien. S. 124–186.
- BITKOM* (2014): Big-Data-Technologien - Wissen für Entscheider. <https://www.bitkom.org/Bitkom/Publikationen/Big-Data-Technologien-Wissen-fuer-Entscheider.html> (Zugriff: 28.08.2019).
- Braganza, A.* (2018): The Challenges of Artificial Intelligence. In: *International Corporate Rescue* 15, 2018, 2, S. 1–4.
- Bundesministerium für Wirtschaft und Energie (o.D.)*: Europäische Datenschutz-Grundverordnung. <https://www.bmwi.de/Redaktion/DE/Artikel/Digitale-Welt/europaeische-datenschutzgrundverordnung.html> (Zugriff: 28.08.2019).
- Cavanillas, J. M.; Curry, E.; Wahlster, W.* (2016): The Big Data Value Opportunity. In: Cavanillas, J. M.; Curry, E. und Wahlster, W. (Hrsg.): *New Horizons for a Data-Driven Economy*. Cham. S. 3–11.
- Chapanis (1951)*: Theory and methods for analyzing errors in man-machine systems. In: *Annals of the New York Academy of Sciences* 51, 1951, 7, S. 1179–1203.
- Clarke, Roger* (2016): Big data, big risks. In: *Information Systems Journal* 26, 2016, 1, S. 77–90.
- Curry, E.* (2016): The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches. In: Cavanillas, J. M.; Curry, E. und Wahlster, W. (Hrsg.): *New Horizons for a Data-Driven Economy*. Cham. S. 29–38.
- Davenport, T. H.; Patil, D. J.* (2012): Data scientist. The sexiest job of the 21st century. In: *Harvard business review* : HBR 90, 2012, 10, S. 70–76.
- Davenport, T. H.; Ronank, R.* (2018): Artificial intelligence for the real world. Don't start with moon shots. In: *Harvard business review* : HBR 96, 2018, 1, S. 108–116.

- Dionne, G. (2013):* Risk Management. History, Definition, and Critique. In: Risk Management and Insurance Review 16, 2013, 2, S. 147–166.
- Domingue, J.; Lasierra, N.; Fensel, A.; van Kasteren, T.; Strohbach, M.; Thalhammer, A. (2016):* Big Data Analysis. In: Cavanillas, J. M.; Curry, E. und Wahlster, W. (Hrsg.): New Horizons for a Data-Driven Economy. Cham. S. 63–86.
- Dorfer, L. (2016):* Datenzentrische Geschäftsmodelle als neuer Geschäftsmodelltypus in der Electronic-Business-Forschung. Konzeptionelle Bezugspunkte, Klassifikation und Geschäftsmodellarchitektur. In: Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung 68, 2016, 3, S. 307–369.
- Dorfer, L. (2018):* Erfolgsstrategien datenzentrischer Geschäftsmodelle. Wiesbaden.
- Ehmke, E. (2019):* Kommunikation und Vertrauen in betrieblichen Krisensituationen. Eine linguistische Analyse am Beispiel der Bankenkrise. Wiesbaden.
- Europäisches Parlament (04.05.2016):* VERORDNUNG (EU) 2016/679 DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung). <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A32016R0679> (Zugriff: 28.08.2019).
- Fahrrad XXL (o.D.):* Datenschutzerklärung. <https://www.fahrrad-xxl.de/service/datenschutz/> (Zugriff: 28.08.2019).
- Freitas, A.; Curry, E. (2016):* Big Data Curation, In: Cavanillas, J. M.; Curry, E. und Wahlster, W. (Hrsg.): New Horizons for a Data-Driven Economy. Cham. S. 87–118.
- Frese, M.; Keith, N. (2015):* Action errors, error management, and learning in organizations. In: Annual review of psychology 66, 2015, S. 661–687.
- Fuest, B. (2018):* Alexa-Nutzer bekommt fremde Tonaufnahmen zugespielt. <https://www.welt.de/wirtschaft/webwelt/article185843562/Amazon-Alexa-Aufnahmen-an-falschen-Nutzer-verschickt.html> (Zugriff: 28.08.2019).
- Gentsch, P. (2019):* Künstliche Intelligenz für Sales, Marketing und Service. Wiesbaden.
- Große Holtforth, D. (2017):* Schlüsselfaktoren im E-Commerce. Wiesbaden.

- Grottke, M.; Trivedi, K. S. (2007):* Fighting bugs. Remove, retry, replicate, and rejuvenate. In: *Computer* 40, 2007, 2, S. 107–109.
- Hagen, J. U. (2017):* Fehlermanagement. In: Hagen, J. U. (Hrsg.): *Fatale Fehler*. Berlin/Heidelberg. S. 181–191.
- Hofinger, G. (2012):* Fehler und Unfälle. In: Badke-Schaub, P.; Hofinger, G.; Lauche, K. (Hrsg.): *Human Factors*. Berlin/Heidelberg. S. 40–59.
- IKEA (o.D.):* IKEA Datenschutzerklärung. <https://www.ikea.com/de/de/customer-service/privacy-policy/> (Zugriff: 28.08.2019).
- Jagadish, H. V.; Gehrke, J.; Labrinidis, A.; Papakonstantinou, Y.; Patel, J. M.; Ramakrishnan, R.; Shahabi, C. (2014):* Big data and its technical challenges. In: *Communications of the ACM* 57, 2014, 7, S. 86–94.
- Jansen, J. (2018):* Warum sich die Werbung immer stärker personalisiert. Targeting. <https://www.faz.net/aktuell/wirtschaft/diginomics/warum-sichdie-werbung-immer-staerker-personalisiert-15788686.html> (Zugriff: 28.08.2019).
- Klein, D.; Tran-Gia, P.; Hartmann, M. (2013):* Big Data. In: *Informatik Spektrum* 36, 2013, 3, S. 319–323.
- Kollmann, T. (2019):* E-Business. Grundlagen elektronischer Geschäftsprozesse in der Digitalen Wirtschaft. 7. überarbeitete Auflage. Wiesbaden.
- Krcmar, H. (2015):* Einführung in das Informationsmanagement. 2. überarbeitete Auflage. Berlin.
- Krystek, U. (2002):* Unternehmungskrisen: Vermeidung und Bewältigung. In: Pastors, P. M. (Hrsg.): *Risiken des Unternehmens - vorbeugen und meistern*. Mering., S. 87–134.
- Lambrecht, A.; Tucker, C. (2013):* When Does Retargeting Work? Information Specificity in Online Advertising. In: *Journal of Marketing Research* 50, 2013, 5, S. 561–576.
- Liggesmeyer, P.; Dörr, J.; Heidrich, J. (2014):* Big Data in Smart Ecosystems. In: *Informatik Spektrum* 37, 2014, 2, S. 105–111.
- Lyko, K.; Nitzschke, M.; Ngonga Ngomo, A. (2016):* Big Data Acquisition. In: Cavanillas, J. M.; Curry, E. und Wahlster, W. (Hrsg.): *New Horizons for a Data-Driven Economy*. Cham. S. 39–61.

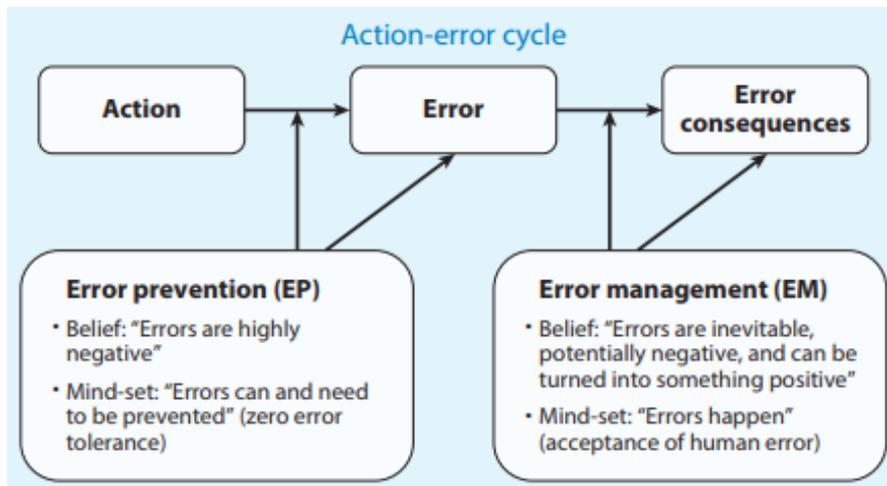
- Marr, B. (2015):* Where Big Data Projects Fail. <https://www.forbes.com/sites/bernardmarr/2015/03/17/where-big-data-projects-fail/#4f328654239f> (Zugriff: 28.08.2019).
- de Mauro, A.; Greco, M.; Grimaldi, M. (2015):* What is big data? A consensual definition and a review of key research topics. In Giannakopoulos, G.; Sakas, D. P.; Kyriaki-Manessi, D. (Hrsg.): International Conference on Integrated Information (IC-ININFO 2014); 4. Internationale Konferenz, Madrid, Spanien, 5. bis 8. September 2014; Proceedings. New York, S. 97–104.
- McAfee, A.; Brynjolfsson, E. (2012):* Big data. The management revolution. In: Harvard business review : HBR 90, 2012, 10, S. 60–68.
- McAfee, A.; Brynjolfsson, E. (2017):* The Business of Artificial Intelligence. What it can - and cannot - do for your organization. <https://hbr.org/cover-story/2017/07/the-business-of-artificial-intelligence> (Zugriff: 28.08.2019).
- Meister, D. (1971):* Human factors. Theory and practice. New York.
- Miller, H. G.; Mork, P. (2013):* From Data to Decisions. A Value Chain for Big Data. In: IT Prof. 15, 2013, 1, S. 57–59.
- Müller-Neuhof, J. (2018):* Barley fordert mehr Meinungen bei Facebook. <https://www.tagesspiegel.de/politik/tagesspiegel-diskussionsveranstaltung-barley-fordert-mehr-meinungen-bei-facebook/21190242.html> (Zugriff: 28.08.2019).
- Najafabadi, M. M.; Villanustre, F.; Khoshgoftaar, T. M.; Seliya, N.; Wald, R.; Muharemagic, E. (2015):* Deep learning applications and challenges in big data analytics. In: Journal of Big Data 2, 2015, 1, S. 1-21
- OASIS (o.D.):* OASIS Advanced Message Queuing Protocol (AMQP) TC. https://www.oasisopen.org/committees/tc_home.php?wg_abbrev=amqp (Zugriff: 28.08.2019).
- Omri, F. (2015):* Big Data-Analysen: Anwendungsszenarien und Trends. In: Dorschel, J. (Hrsg.): Praxishandbuch Big Data. Wirtschaft -- Recht -- Technik. Wiesbaden, S. 104–112.

- Parbel, M. (2019):* Apache Hadoop 3.2.0 vereinfacht Deep-Learning-Projekte. Das Big-Data-Framework bietet in der neuen Version vor allem Verbesserungen für langlebige und Deep-Learning-Anwendungen. <https://www.heise.de/developer/meldung/Apache-Hadoop-3-2-0-vereinfacht-Deep-Learning-Projekte-4285901.html> (Zugriff: 28.08.2019).
- Picot, A.; Heger, D. K.; Neuburger, R. (2001):* Der elektronische Handel - Potentiale für Unternehmen. In: Ifo-Schnelldienst 54, 2001, 6, S. 19–26.
- Porter, M. E. (1998):* Competitive strategy. Techniques for analyzing industries and competitors. New York.
- Porter, M. E. (2008):* The five competitive forces that shape strategy. In: Harvard business review : HBR 86, 2008, 1, S. 78–93.
- Rasmussen, J. (1982):* Human errors. A taxonomy for describing human malfunction in industrial installations. In: Journal of Occupational Accidents 4, 1982, 2-4, S. 311–333.
- Reason, J. T. (1990):* Human Error. Cambridge.
- Reinsel, D.; Gantz, J.; Rydning, J. (2018):* The Digitization of the World. From Edge to Core. <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf> (Zugriff: 28.08.2019).
- Ringel, M.; Zablit, H.; Grassl, F.; Manly, J.; Möller, C. (2018):* The most innovative companies 2018. Innovators go all in on digital: The Boston Consulting Group.
- Scherer, M. U. (2016):* Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies. In: Harvard Journal of Law & Technology 29, 2016, 2, S. 354–398.
- Shapiro, C.; Varian, H. R. (2008):* Information rules. A strategic guide to the network economy. Boston.
- Strohbach, M.; Daubert, J.; Ravkin, H.; Lischka, M. (2016):* Big Data Storage. In: Cavanillas, J. M.; Curry, E. und Wahlster, W. (Hrsg.): New Horizons for a Data-Driven Economy. Cham. S. 119–142.
- Turban, E.; Outland, J.; King, D.; Lee, J. K.; Liang, T.; Turban, D. C. (2018):* Electronic Commerce 2018. Cham.

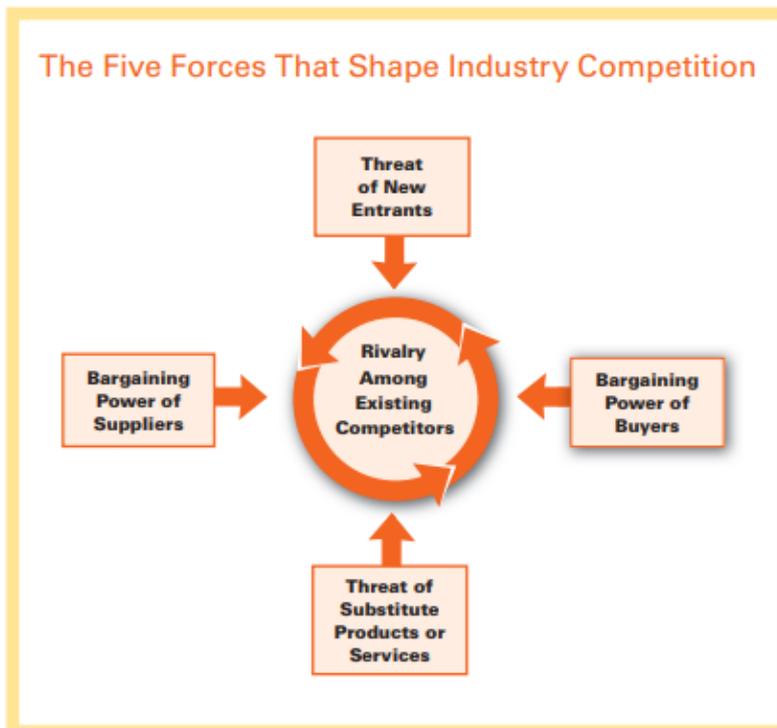
- van Dyck, C.; Frese, M.; Baer, M.; Sonnentag, S. (2005):* Organizational error management culture and its impact on performance. A two-study replication. In: *The Journal of applied psychology* 90, 2005, 6, S. 1228–1240.
- Verhoef, P. C.; Kooge, E.; Walk, N. (2016):* Creating value with big data analytics. Making smarter marketing decisions. London.
- Voß, S.; Gutenschwager, K. (2001):* Informationsmanagement. Berlin/Heidelberg.
- Walsh, G.; Koot, C.; Schmidt, R.; Möhring, M. (2013):* Big Data — neue Möglichkeiten im E-Commerce. In: *Wirtsch Inform Manag* 5, 2013, 2, S. 48– 56.
- Wirtz, B. W.; Becker, D. R. (2002):* Geschäftsmodellansätze und Geschäftsmodellvarianten im Electronic Business. In: *WIST* 31, 2002, 2, S. 85–90.
- Wittmann, W. (1959):* Unternehmung und Unvollkommene Information. Unternehmerische Voraussicht - Ungewißheit und Planung. Wiesbaden.
- Wöfl, S.; Leischnig, A.; Ivens, B.; Hein, D. (2019):* From Big Data to Smart Data – Problemfelder der systematischen Nutzung von Daten in Unternehmen. In: Becker, W.; Eierle, B.; Fliaster, A.; Ivens, B.; Leischnig, A.; Pflaum, A. und Sucky, E. (Hrsg.): *Geschäftsmodelle in der digitalen Welt*. Wiesbaden, S. 213–231.
- Zalando (2019):* Zalando Datenschutzerklärung <https://www.zalando.de/zalando-datenschutz/> (Zugriff: 28.08.2019).

Anhang

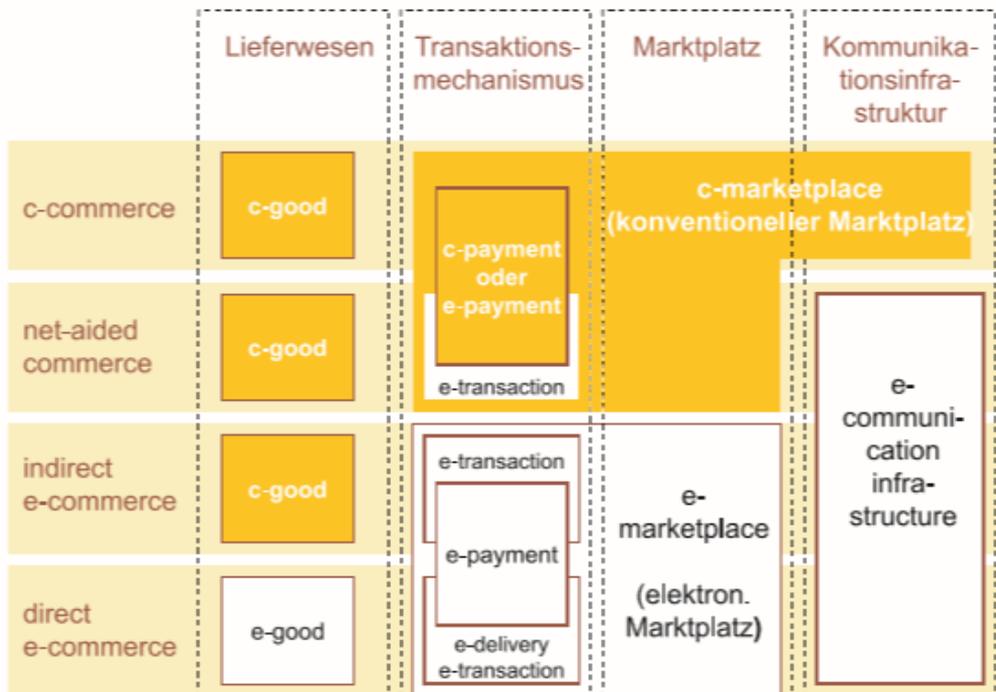
Anhang 1: Fehlerprävention und Fehlermanagement (Frese und Keith 2015, S. 666)



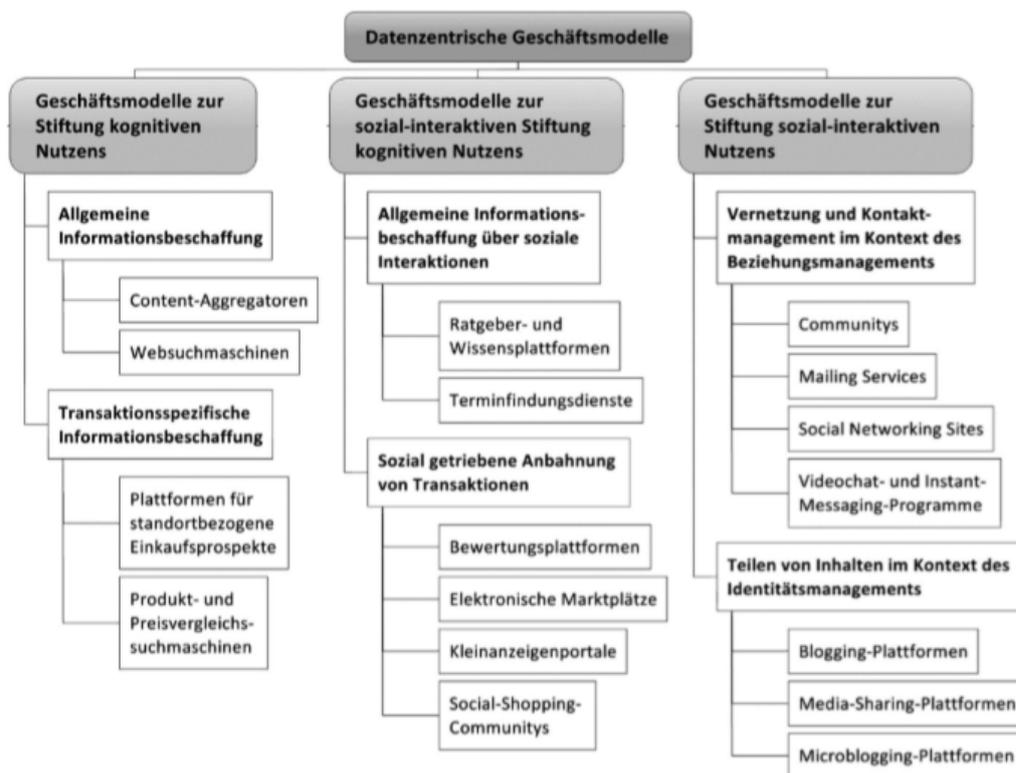
Anhang 2: The Five Forces That Shape Industry Competition (Porter 2008, S. 80)



Anhang 3: Arten und Ebenen von E-Commerce-Realisationsformen nach Picot et al. (2001, S. 21)



Anhang 4: Klassifikation datenzentrischer Geschäftsmodelle (Dorfer 2016, S. 326)



Anhang 5: Elektronischer Wertschöpfungsprozess nach Kollmann (2019, S. 61)

