

Friedrich-Alexander-Universität Erlangen-Nürnberg
Buchwissenschaft
Masterarbeit
Prüferin: Prof. Dr. Svenja Hagenhoff
Wintersemester 2016/17

Big Data in der Buchbranche

Pia Staack
Killingerstr. 46
91056 Erlangen

Studienfach:
Buchwissenschaft

Abstract

Big Data is a relatively new phenomenon that emerged with the rising use of the internet and has been a widely discussed topic in the economy for the last years. More and more companies begin to utilize Big Data to get more insight into the market and the consumers, use that insight for better marketing and service and even adapt their products or predict what consumers might want to buy in the future. This raises the question of how Big Data can be applied to the book industry.

This thesis aims to analyze and evaluate the potential of applying Big Data to the book industry. The first chapter focuses on Big Data in general. What is Big Data? How can Big Data be collected, stored and analyzed? What are possible chances and challenges for the companies and for the consumers? How does Big Data function in certain companies? The second chapter discusses the premises, adaption and perspectives of Big Data in the book industry. First, the specifics of the book market in general and marketing in the book market in particular are outlined. Second, the current usage of data and the current discourse in the book industry are described. Third, the Reader Analytics as a form of Big Data specific to the book industry are analyzed. On the basis of the findings, the final chapter discusses and evaluates the potential of Big Data for the book industry.

Inhaltsverzeichnis

Abbildungsverzeichnis.....	III
Tabellenverzeichnis.....	IV
Abkürzungsverzeichnis.....	V
1. Einleitung.....	1
1.1 Motivation und Problemstellung.....	1
1.2 Erkenntnisinteresse und Forschungsfragen.....	1
1.3 Aufbau und Methodik.....	2
1.4 Forschungsstand.....	2
2. Big Data.....	4
2.1 Einführung und Begriffsbestimmungen.....	4
2.2 Big Data: Funktionen und Funktionsweisen.....	6
2.2.1 Die Entstehung von Big Data.....	6
2.2.2 Datensammlung.....	7
2.2.3 Datenspeicherung.....	10
2.2.4 Datenanalyse.....	11
2.2.5 Big Data im Unternehmen.....	14
2.2.6 Chancen und Risiken von Big Data.....	17
2.3 Fallbeispiele.....	19
2.3.1 Streamingdienste – Netflix und Spotify.....	19
2.3.2 Amazon.....	21
2.4 Fazit.....	23
3. Big Data in der Buchbranche.....	25
3.1 Voraussetzungen für Big Data in der Buchbranche.....	25
3.1.1 Besonderheiten des Buchmarkts.....	25
3.1.2 Marketing im Buchmarkt.....	27
3.2 Anwendung von Big Data in der Buchbranche.....	30
3.2.1 Anwendung von Daten in der Buchbranche.....	30
3.2.2 Diskussion in der Buchbranchenpresse.....	33
3.2.3 Reader Analytics.....	38
3.3 Perspektiven von Big Data in der Buchbranche.....	45
3.3.1 Welches Potenzial haben die Reader Analytics?	46
3.3.2 Lassen sich mit Big Data Bestseller vorhersagen?.....	49
3.3.3 Inwieweit wird sich die Buchbranche insgesamt mehr an den Lesern orientieren?.....	52
3.3.4 Inwieweit können Daten die Intuition ersetzen?	54
4. Synopse.....	58
5. Schluss.....	62
Literaturverzeichnis.....	64

Abbildungsverzeichnis

Abbildung 1: Drei verschiedene Ergebnisse der Reader Analytics von Jellybooks (Alter/Russell 2016).....	42
--	----

Tabellenverzeichnis

Tabelle 1: Die untersuchten Branchenpressenartikel, sortiert nach Bewertung.....37

Abkürzungsverzeichnis

CRM	Customer-Relationship-Management
ID	Identifier
IP	Internet Protocol
IT	Informationstechnik
MVB	Marketing- und Verlagsservice des Buchhandels GmbH
NSA	National Security Agency
VLB	Verzeichnis lieferbarer Bücher
WWW	World Wide Web

1. Einleitung

1.1 Motivation und Problemstellung

Wie sich Big Data, also das Sammeln, Speichern, Verarbeiten und Auswerten von Massendaten (vgl. Bendel 2016), effizient einsetzen lässt und welche Herausforderungen damit verbunden sind, wird derzeit in der Wirtschaft diskutiert. Darauf lassen das große Publikationsvolumen zu dem Thema und die große Medienaufmerksamkeit schließen. Einige Unternehmen, beispielsweise die Automobil-Industrie oder Software-Unternehmen wie Microsoft, setzen Big Data bereits ein (vgl. Vossen 2015, S. 41; vgl. Marr 2016, S. 95–101). Da sich mit Big Data Erkenntnisse zum Kundenverhalten gewinnen lassen und sogar Vorhersagen getroffen werden können und dies die Chance auf Umsatzsteigerungen bietet, lässt sich vermuten, dass sich Big Data zunehmend in der Wirtschaft etablieren wird und immer mehr Unternehmen Big Data nutzen werden.

Im Zusammenhang mit Big Data spricht die Literatur teilweise vom Ende der Intuition oder dem Ende des Zufalls (vgl. Bachmann et al. 2014, S. 340; vgl. Klausnitzer 2013), da durch den Einsatz von Big-Data- und Smart-Data-Konzepten bessere Unternehmensentscheidungen, die allein auf Daten basieren, ermöglicht werden. Entscheidungen für oder gegen die Publikation von Texten werden in Buchverlagen dagegen vermeintlich oft nach »Bauchgefühl« gefällt (vgl. Lichtenberg 2014; vgl. Neary 2016; vgl. Reid 2012; vgl. Anderson 2015). Mit der Nutzung von Daten könnte sich diese Arbeitsweise in der Verlagsbranche grundsätzlich hin zu einem »Data-Smart Publishing« (Rhomberg 2016c) ändern und dazu führen, dass die Verlage wirtschaftlich effektiver arbeiten können. Die Auswertung von großen, nutzerbezogenen Datenmengen könnte demnach in der Zukunft auch für die Buchbranche von großer Bedeutung sein. In der Buchbranche werden Verfahren und Technologien zur Sammlung und Auswertung von Massendaten jedoch noch nicht breit angewandt. Dass das Thema Big Data bei Branchentagungen wie der *future! publish* 2017 auf der Tagesordnung steht (vgl. o. V. 2017), lässt jedoch die Vermutung zu, dass das Thema Big Data auch in der Buchbranche immer mehr diskutiert wird. In dieser Masterarbeit soll daher das Potenzial von Big-Data-Konzepten in der Buchbranche analysiert werden.

1.2 Erkenntnisinteresse und Forschungsfragen

Die Forschungsfrage, der in dieser Arbeit nachgegangen wird, lautet: Welches Potenzial hat Big Data für die Buchbranche? Ziel der Arbeit ist es, die Voraussetzungen, die Anwendung und die Perspektiven von Big Data in der Buchbranche zu untersuchen. Bei den Anwendungsmöglichkeiten soll dabei vertiefend auf die sogenannten Reader Analytics eingegangen werden. Anschließend sollen die gewonnenen Erkenntnisse diskutiert und bewertet werden.

1.3 Aufbau und Methodik

Die Arbeit gliedert sich in drei Teile. Im Zentrum des ersten Teils (Kapitel 2) steht eine umfassende Erläuterung des Konzepts Big Data anhand einer Literaturanalyse. Es geht dabei um die Frage, wie Big Data definiert wird, wie das Datensammeln und -speichern sowie die Datenanalyse funktionieren, wie Big Data im Unternehmen integriert werden kann und welche Chancen und Risiken sich durch Big Data ergeben. Zuletzt wird anhand von zwei Fallbeispiele erläutert, wie Big Data genutzt werden kann und welche Daten dafür gesammelt und analysiert werden. Der zweite Teil (Kapitel 3) analysiert das Verhältnis von Big Data und der Buchbranche in drei Unterkapiteln. Kapitel 3.1 untersucht die Voraussetzungen, die für Big Data in der Buchbranche vorliegen. Die Besonderheiten des Buchmarkts werden herausgearbeitet und es wird die bisherige Funktionsweise des Buchmarkts erläutert. In Kapitel 3.2 wird die Antwort auf die Frage gesucht, welche Datenanwendungen es bereits gibt und wie das Thema Big Data in der Buchbranchenpresse diskutiert wird. Anschließend wird das Thema Reader Analytics vertieft; es wird dargestellt, worum es sich dabei handelt und es werden einige Unternehmen vorgestellt, die mit Reader Analytics arbeiten. Im Kapitel 3.3 werden auf Basis der bisherigen Erkenntnisse das Potenzial der Reader Analytics, die Möglichkeit der Bestseller-Vorhersage, die zunehmende Orientierung am Leser mit Hilfe von Daten und die Kompatibilität von Intuition und Daten untersucht. Im dritten Teil (Kapitel 4), der Synopse, werden die vorher erarbeiteten Erkenntnisse zusammengefasst und im Hinblick auf die Forschungsfrage analysiert, um schließlich das Potenzial von Big Data in der Buchbranche insgesamt bewerten zu können.

1.4 Forschungsstand

In den letzten Jahren, das heißt ungefähr seit dem Jahr 2012 – das als das Jahr des Durchbruchs von Big Data bezeichnet wird (vgl. Lohr 2012) – ist das Publikationsvolumen zum Thema Big Data deutlich angestiegen. Die Literaturanalyse stützt sich im Folgenden vor allem auf Fachliteratur in Form von Monografien oder Sammelwerken. Hinzu kommen Beiträge in Branchenzeitschriften wie das *Börsenblatt* oder das *Buchreport Magazin* und aus Internetdokumenten sowie Graue Literatur wie beispielsweise Essays, die nicht über Verlage veröffentlicht wurden.

Es lassen sich zahlreiche Veröffentlichungen zum Thema Big Data mit jeweils unterschiedlichen Schwerpunkten finden. Zunächst gibt es einige Fachbücher, die erklären und definieren, was Big Data ist, und erläutern, wie es funktioniert oder funktionieren kann, und welche Chancen und Veränderungen Big Data mit sich bringt. Hier seien vor allem das Fachbuch *Creating Value with Big Data Analytics. Making smarter marketing decisions* (Verhoef et al. 2016), der Aufsatz *Big Data. Daten sammeln, aggregieren, analysieren, nutzen* (Vossen 2015) und *Big Data. Potential und Barrieren der Nutzung im Unternehmenskontext* (King 2014) genannt, diese Literatur wurde insbesondere für das Kapitel 2 oft verwendet.

Schwerpunkte sind häufig die ökonomischen Veränderungen, die mit Big Data einhergehen oder spezielle Gebiete, für die Big Data nützlich ist, wie das Marketing. Besonders viele Erkenntnisse zum Marketing mit Big Data und zum Einsatz von Big Data in Unternehmen finden sich in *Big Data – Fluch oder Segen?* (Bachmann et al. 2014) und in dem Sammelwerk *Big Data im Marketing* (Schwarz 2015). Zudem gibt es einige Literatur aus dem Bereich Informatik oder Wirtschaftsinformatik, welche meist eher praxisorientiert ist und Anweisungen gibt, wie Daten gesammelt und analysiert werden können. Dies ist beispielsweise bei *Big Data. Grundlagen, Systeme und Nutzungspotenziale* (Fasel/Meier 2016) der Fall. Außerdem existieren einige Sammelwerke, die versuchen, das Phänomen Big Data aus unterschiedlichen Perspektiven zu beleuchten. Dabei wird unter anderem auch auf Datenschutz und andere Risiken von Big Data eingegangen. Als Beispiele seien hier *Big Data. Analysen zum digitalen Wandel von Wissen, Macht und Ökonomie* (Reichert 2014) und in *Big Data. Das neue Versprechen der Allwissenheit* (Geiselberger/Moorstedt 2013) genannt.

Fachliteratur, die sich dem Verhältnis von Big Data und der Buchbranche widmet, gibt es dagegen bisher nicht. Auch Literatur, die überwiegend erläutert, wie Big Data in der Praxis angewandt wird – wie beispielsweise *Big Data in practice* (Marr 2016) oder *Big Data revolution* (Thomas/McSharry 2015) – hat die Buchbranche nicht im Blick. Berichte, Diskussionen und Analysen aus der Branche lassen sich daher nur über die brancheninternen Organe wie *Buchreport Magazin* und *Börsenblatt* und deren Online-Zweigstellen finden sowie über englischsprachige Pendanten wie *publishersweekly.com* oder Websites, die sich überwiegend mit der digitalen Publishing-Branche auseinandersetzen, etwa *digitalbookworld.com*, finden. Besonders *digitalbookworld.com* publiziert häufig neue Artikel zum Thema Big Data in der Buchbranche.

2. Big Data

2.1 Einführung und Begriffsbestimmungen

Die genaue Bestimmung dessen, was mit dem Begriff Big Data bezeichnet wird, ist nicht einfach: Big Data ist ein recht unpräziser Begriff, der sich in den letzten Jahren allerdings weitgehend in den Medien und in vielen, meist jungen Unternehmen durchgesetzt hat (vgl. Davenport 2014a, S. 6 ff.). Dennoch gibt es einige Versuche, Big Data so zu definieren, dass mit diesem Begriff relativ neue Phänomene möglichst vielschichtig erfasst werden können (vgl. Bachmann et al. 2014, S. 17). Eine kurze Definition lautet: »Big data is defined as the dataset whose size is beyond the processing ability of traditional databases or computers« (Cheng et al. 2016, S. 3). Diese Definition geht lediglich auf die Quantität von Big Data ein, nicht jedoch auf die Qualität. Weitere Definitionen fassen Big Data nicht nur als quantitatives, sondern auch als qualitatives Phänomen auf: Entsprechend erfasst Vossen (2015, S. 37) Big Data nicht nur über das Datenvolumen (*volume*), sondern beschreibt auch den Komplexitätsgrad der Daten als Set von neuen weiteren Merkmalen, den sogenannten Vs. Diese sind: *velocity*, *variety*, *veracity*, *variability*, *venue*, *vocabulary*, *vagueness*, *value*, *validity* und *volume*. Cheng et al. erwähnen in ihrer Definition lediglich das Charakteristikum *volume*.

Laut Cordon et al. (2016, S. 4) prägte das IT-Unternehmen Gartner den Begriff Big Data und die folgende Definition, welche oft in der Fachliteratur verwendet wird:

Big Data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation (Gartner 2016).

Diese Definition nennt drei wichtige Charakteristika von Big Data – die drei Vs *volume*, *velocity* und *variety* – und geht auf ökonomische Funktionen – Big Data ermöglicht erweiterte Erkenntnisse, Entscheidungsprozesse und Prozessautomatisierungen – ein. Die Definition von Gartner entwickelte sich schnell zur Standarddefinition und in späteren Begriffserläuterungen wurden oft noch die ebenfalls sehr wichtigen Vs *veracity* und *volume* hinzugefügt (vgl. Cordon et al. 2016, S. 5). Daher sollen im Folgenden diese fünf V-Begriffe näher erläutert werden.

Volume bezeichnet den Umfang, also die Menge an Daten, die erfasst und verwertet werden müssen (vgl. King 2014, S. 35). Als »big« werden Datenmengen bezeichnet, sobald ihre Größe in den Terabyte- oder sogar Perabyte-Bereichen¹ liegt und sie so groß sind, dass die Speicherkapazität einer einzelnen Organisation sie nicht erfassen kann (vgl. Vossen 2015, S. 36).²

1 1 Terabyte sind 10^{12} Bytes, 1 Petabyte sind 10^{15} Bytes

2 Wie groß das Volumen aller Daten im Internet bisher genau ist, dazu gibt es unterschiedliche Angaben. Das Volumen an Daten in Form von Dokumenten im Internet (vgl. King 2015, S. 37) liegt laut Taylor-Sakya (2016, S. 2) derzeit im Perabyte-Bereich und wird in den nächsten Jahren die Zettabytes-Bereiche erreichen (1 Zettabyte sind 10^{21} Bytes und 1 Mio. Petabytes). Die Menge an

Mit *variety* ist die Vielfalt der Daten gemeint, die in unterschiedlichen Formen und Formaten vorliegen können (vgl. Vossen 2015, S. 37). Die Daten haben dabei auch unterschiedliche Strukturierungsgrade, sie können unstrukturiert, semi-strukturiert und strukturiert vorliegen (vgl. Vossen 2015, S. 37). Strukturierte Daten sind Daten, die in einer gewöhnlichen Datenbankstruktur gespeichert werden, wie beispielsweise CRM-Daten³ (vgl. King 2014, S. 35). Semi-strukturierte Daten sind zum Teil strukturiert, enthalten aber auch unstrukturierte Komponenten, die mit komplexeren Verfahren verarbeitet werden müssen; Beispiele hierfür sind Textverarbeitungsdokumente oder Umfragen (vgl. King 2014, S. 35). Unstrukturierte Daten sind inkonsistent in ihrer Form und Struktur und müssen daher mit neuen Technologien verarbeitet werden (vgl. King 2014, S. 35). Dazu zählen die meisten Internetdaten wie beispielsweise Kundenrezensionen oder Daten aus sozialen Netzwerken (vgl. King 2014, S. 35).

Mit *velocity* ist die Geschwindigkeit gemeint, mit der die Daten erfasst und analysiert werden (vgl. King 2014, S. 35). Daten, die oft in Datenströmen auftreten, können nicht dauerhaft zur Analyse gespeichert werden, und müssen daher unmittelbar, meist in Echtzeit, verarbeitet und analysiert werden (vgl. Vossen 2015, S. 37; vgl. Fasel/Meier 2016, S. 6). Laut Bachmann et al. (2014, S. 24) geht es aber nicht nur um die Geschwindigkeit, in der Daten verarbeitet werden können, sondern auch um die Veränderungsdynamik, also die Geschwindigkeit, mit der sich Daten und auch die Beziehungen zwischen den Daten und deren Bedeutung verändern.

Mit *veracity* wird die Richtigkeit der Daten bezeichnet, also wie präzise, genau und vertrauenswürdig sowie von welcher Qualität die Daten sind und aus welcher Quelle sie stammen (vgl. King 2014, S. 35; vgl. Vossen 2015, S. 37). Außerdem umfasst *veracity*, dass Entscheidungen, die auf Daten basieren, nachvollziehbar und begründbar sein müssen (vgl. King 2014, S. 35).

Value bezeichnet den Wert, der den Daten zugeschrieben wird (vgl. Vossen 2015, S. 37). Damit ist vor allem gemeint, inwieweit Daten wirklich verwendet werden können und damit einen Nutzen für Unternehmen haben (vgl. Iafrate 2015, S. 9).

Neben der umfassenden Definition von Gartner, muss auf die Definition von Taylor-Sakyi hingewiesen werden:

Big Data refers to large sets of complex data, both structured and unstructured which traditional processing techniques and/or algorithms are unable to operate on. It aims to reveal hidden patterns [...] (Taylor-Sakyi 2016, S. 1).

Daten weltweit verdoppelt sich laut Mainzer (2014, S. 232), laut King (2015, S. 37) und laut Chen et al. (2014, S. 1) schätzungsweise alle zwei Jahre. Floridi (2015, S. 31) hingegen gibt an, dass die Menge an Daten bereits zwischen 2006 und 2011 die Zettabyte-Marke erreichte, und da sich die Datenmenge alle drei Jahre vervierfacht, müsste im Jahr 2015 ein Wert von 8 Zettabytes erreicht werden. Schmidt (2014, S. 6) wiederum gibt mit Verweis auf Marktforscher vom IT-Unternehmen EMC an, dass das Datenvolumen im Jahr 2013 bei 4,4 Zettabytes lag und im Jahr 2020 geschätzt das Zehnfache betragen wird.

3 Die Erläuterung von CRM-Daten erfolgt in Kapitel 3.2.1 ab Seite 30.

Diese Definition geht nur auf die technologische Komponente von Big Data ein und macht deutlich, dass es sich bei Big Data sowohl um strukturierte als auch um unstrukturierte Daten handelt, und dass das Erkennen von Mustern eine relevante Funktion von Big Data ist. Diese Informationen fehlen in den Definitionen von Cheng et al. und Gartner, daher ist diese Definition eine gute Ergänzung.

Obwohl die vorgestellten Definitionen große Überschneidungen aufweisen, lassen sie sich doch nach ihren Schwerpunktbildungen unterscheiden. So fokussieren einige Definitionen den technologischen Aspekt von Big Data, wohingegen andere versuchen, auch die wirtschaftliche Komponente von Big Data zu benennen. Die Definition von Gartner hat sich dennoch zu einer Art Standarddefinition entwickelt, welche in der Fachliteratur häufig genannt wird. Dies liegt daran, dass diese Definition die wesentlichen Charakteristika von Big Data benennt und dabei sowohl die technologische als auch die wirtschaftliche Seite von Big Data enthält. Daher kann sie als die vollständigste kompakte Definition bezeichnet werden. Als gute Ergänzung eignet sich die Definition von Taylor-Sakyi, weil sie die Technologie stärker in den Blick nimmt und auf die Strukturierungsgrade und die Mustererkennung bei Big Data eingeht.

2.2 Big Data: Funktionen und Funktionsweisen

2.2.1 Die Entstehung von Big Data

Da es sich bei Big Data um ein recht junges Phänomen handelt, sollen hier kurz die Entstehungsgeschichte und die Ursachen für die zunehmende Durchsetzung von Big Data sowie die öffentliche Wahrnehmung des Phänomens thematisiert werden.

Gewichtige Gründe dafür, dass das Volumen an Daten im Internet so rasant steigt, sind der Übergang vom Web 1.0 zum Web 2.0 und die massenhafte Verbreitung von Smartphones und Tablets in den 2010er Jahren (vgl. Vossen 2015, S. 38; vgl. Iafrate 2015, S. xxii f.). Der Schritt von Web 1.0 zu Web 2.0 wird durch eine Veränderung von Techniken, Technologien und Nutzungsmustern charakterisiert (vgl. Vossen 2015, S. 38). Große Fortschritte in der Technologie, vor allem im Bereich Netzwerktechnik, Rechnerhardware und in der Softwaretechnik haben Dienste und Anwendungen ermöglicht, die leicht zu bedienen sind und heute von den meisten Nutzern verwendet werden (vgl. Vossen 2015, S. 38). Die Nutzer, inzwischen umfassend mit dem Internet sozialisiert, wissen es daher privat und professionell zu nutzen; sie rufen dort Informationen auf und tragen gleichzeitig auch selber welche bei (vgl. Vossen 2015, S. 38). Die massenhafte Nutzung von Smartphones und Tablets führt zu einer hohen Mobilität und Vernetzung der Nutzer und dazu, dass Daten und damit Aspekte des alltäglichen Lebens schneller verfügbar sind (vgl. Iafrate 2015, S. xxii f.; Vossen 2015, S. 35). Dazu gehören beispielsweise Social-Media-Profilen und -Posts, Online-Shopping-Daten und Gesundheitsdaten (vgl. Vossen 2015, S. 35).

Der Beginn der breiten öffentlichen Wahrnehmung des Phänomens Big Data kann auf das Jahr 2013 datiert werden: Im Juni dieses Jahres enthüllte Whistleblower Edward Snowden, dass die NSA in den USA sogenannte Metadaten aller Bürger, also beispielsweise Daten zur Zeit, Dauer und zum Standort eines Handyanrufs, sammelt und speichert (vgl. Boellstorff 2014, S. 106, 111). Dies führte dazu, dass der Datenschutz als kritischer Aspekt von Big Data im Hinblick auf die vielfache Verwendung von Nutzerdaten identifiziert und öffentlich debattiert werden konnte.⁴ Lohr (2012) bezeichnet dennoch bereits das Jahr 2012 als das Jahr des Durchbruchs für Big Data und zählt als Beleg dafür auf, dass Big Data 2012 das Titelthema beim World Economic Forum in Davos war, wo ein Bericht mit dem Titel *Big Data, Big Impact* veröffentlicht wurde. Zudem hat die Bundesregierung in den USA 200 Millionen US-Dollar für Rechercheprogramme zum Thema »Big Data computing« in Aussicht gestellt (vgl. Lohr 2012). Auch die New York Times veröffentlichte in dem Jahr bereits mehrere Artikel zu Big Data (vgl. Lohr 2012).

Das Phänomen Big Data ist also noch recht jung und die Durchsetzung von mobilen Geräten wie Smartphones und Tablets hat die Entwicklung deutlich beschleunigt, da sich hier viele verschiedene Daten sammeln lassen. Zugleich handelt es sich um ein in der Öffentlichkeit zunehmend kontrovers diskutiertes Thema. In den folgenden Kapiteln wird erläutert, was Big Data ist und wie es funktioniert. Zunächst geht es darum, was für Daten wie gesammelt werden und woher sie kommen. Danach wird auf Datenspeicherungsmöglichkeiten eingegangen. Anschließend wird dargestellt, wie die Big-Data-Analyse funktioniert. Danach wird erläutert, wie Big Data in einem Unternehmen integriert und angewandt werden und in welchen Bereichen eines Unternehmens Big Data nützlich sein kann. Zum Schluss werden die Chancen und Risiken von Big Data auf den Punkt gebracht.

2.2.2 Datensammlung

Es wurde bereits erwähnt, dass die Daten strukturiert oder unstrukturiert sein können. In diesem Abschnitt soll der Unterschied zwischen Daten aus internen und externen Datenquellen, sowie zwischen strukturierten und unstrukturierten Daten erläutert werden. Danach wird auf die spezifischen Unterschiede zwischen Markt-, Marken- und Kundendaten eingegangen. Dies ist wichtig, um zu verstehen, welche Daten mit dem Begriff Big Data bezeichnet werden können und woher sie kommen.

Verhoef et al. (2016, S. 76) unterscheiden interne und externe Datenquellen. Daten aus externen Quellen sind Daten, die ein Unternehmen nicht selbst sammeln kann, sondern käuflich erwerben muss; ein Beispiel hierfür sind Postleitzahlen- und Haushaltsdaten (vgl. Verhoef et al. 2016, S. 77). Diese enthalten Informationen über die Charakteristika eines Haushalts in einer Postleitzahlen-Region wie beispielsweise Einkommen, Bildung oder den durchschnittlichen Preis für ein Haus (vgl. Verhoef et al. 2016, S. 77). Auch Daten zur Kreditwürdigkeit einer Person und Marktforschungs-

⁴ Im Kapitel 2.2.6 ab Seite 17 wird auf diesen Kritikpunkt noch einmal genauer eingegangen.

daten gehören zu solchen Daten (vgl. Verhoef et al. 2016, S. 77). Daten aus sozialen Netzwerken wie beispielsweise Facebook oder Twitter sind ebenfalls eine rasant wachsende Datenquelle für externe Daten; diese werden aber sehr unstrukturiert gesammelt, sodass es schwierig ist, sie zu interpretieren und für kommerzielle Zwecke zu nutzen (vgl. Verhoef et al. 2016, S. 78).

Daten aus internen Quellen dagegen stammen aus dem Unternehmen selbst, dies sind beispielsweise Rechnungsdaten, Transaktionsdaten, Kontaktdaten, Nutzungsdaten etc. (vgl. Verhoef et al. 2016, S. 78). Diese Daten können sehr nützlich sein, um Kundenverhalten zu beschreiben, zu verstehen und vorherzusagen; dafür müssen diese Daten aber erst aufbereitet werden (vgl. Verhoef et al. 2016, S. 78).

Nachdem die Quellen der Datengenerierung erläutert wurden, geht es im Folgenden um die Qualität der Daten, also um den Strukturierungsgrad.

Strukturierte Daten sind Daten, welche ein fixes Format und eine hohe Qualität haben (vgl. Verhoef et al. 2016, S. 79). Strukturierte Daten können in einer traditionellen Datenbank gespeichert werden (vgl. King 2014, S. 35). Postleitzahlendaten, Rechnungsdaten oder auch CRM-Daten und Transaktionsdaten sind gute Beispiele für solche sehr strukturierten Daten (vgl. Verhoef et al. 2016, S. 79; vgl. King 2014, S. 35).

Unstrukturierte Daten hingegen sind meist von sehr umfangreicher Größe, haben kein fixes Format, enthalten oft viel unformatierten Text und müssen erst interpretiert und reduziert werden, um daraus nutzbare Informationen zu erhalten (vgl. Verhoef et al. 2016, S. 79). Wegen ihrer inkonsistenten Struktur werden neue Technologien benötigt, um sie verarbeiten zu können (vgl. King 2014, S. 35). Zu diesen Daten zählen Daten, die zum großen Teil aus dem Internet stammen, wie beispielsweise Kundenrezensionen, Daten aus sozialen Netzwerken, Wetterdaten und Videos (vgl. King 2014, S. 35).

Eine weitere Form sind semi-strukturierte Daten, welche sowohl strukturierte als auch unstrukturierte Elemente enthalten und daher nicht standardisiert verarbeitet werden können (vgl. King 2014, S. 35). Beispiele hierfür sind Textverarbeitungsdokumente, Umfragen und Weblogs (vgl. King 2014, S. 35).

Mobile Daten stellen eine spezielle Form an Daten dar. Mobile Daten sind ein recht neues Phänomen, das mit der Verbreitung von Smartphones und Tablets auftritt (vgl. Verhoef et al. 2016, S. 79). Über installierte Apps auf den Smartphones und Tablets der Nutzer bieten mobile Daten die Möglichkeit, zu überwachen, was Nutzer gerade tun und wo sie es tun (vgl. Verhoef et al. 2016, S. 79). Dies wiederum bietet die Gelegenheit, ortsbasierte Dienste zur Verfügung zu stellen (vgl. Verhoef et al. 2016, S. 79). Da diese Daten meist noch nicht in die üblichen Datenumgebungen, meist die Infrastruktur eines Unternehmens, integriert werden können und diese Daten oft noch unstrukturiert sind, ist hier noch einiges an Aufbereitungsaufwand nötig, um diese Daten sinnvoll verwenden zu können (vgl. Verhoef et al. 2016, S. 79 f.).

Verhoef et al. (2016, S. 80) unterscheiden desweiteren zwischen Marktdaten, Produkt- und Markendaten und Kundendaten. Dies soll die ökonomische Komponente von Daten deutlich machen und die jeweilige Stärke der Daten aus verschiedenen Datenquellen, welche speziell den Markt, die Marke und die Produkte oder die Kunden betreffen, hervorheben.

Marktdaten umfassen zum einen Daten auf der Angebotsseite, die Maße wie Marktgröße, Marktvolumen, Marktanteile, Marktentwicklungen etc. für alle Marktteilnehmer beschreiben und erklären, zum anderen Daten auf der Nachfrageseite, die die Konsumenten im Markt beschreiben, also wie und was sie kaufen, wieviel Geld sie wofür ausgeben, ihre soziodemografischen Daten wie Alter, Einkommen etc. und ihre Bedürfnisse (vgl. Verhoef et al. 2016, S. 80). Die Marktdaten der Angebotsseite kommen oft von Agenturen, die beispielsweise Daten aus Verkaufszahlen erzeugen, um eine Marktübersicht darstellen zu können (vgl. Verhoef et al. 2016, S. 80). Marktdaten der Nachfrageseite werden hauptsächlich von Marktforschungsagenturen gesammelt (vgl. Verhoef et al. 2016, S. 80).

Auch die Markendaten werden in eine Angebotsseite und eine Nachfrageseite eingeteilt; Markendaten der Angebotsseite sind Daten, die das Volumen und den Marktanteil von bestimmten Marken oder Produkten beschreiben (vgl. Verhoef et al. 2016, S. 82). Markendaten der Nachfrageseite sind Daten, die beschreiben und erklären, wie Kunden oder potenzielle Kunden eine bestimmte Marke bewerten; diese Art an Daten zu erhalten ist oft kostspielig (vgl. Verhoef et al. 2016, S. 82). Markendaten der Angebotsseite werden oft aus Marktdaten kreiert oder sie kommen aus Systemen, welche die Verkaufszahlen pro Marke speichern (vgl. Verhoef et al. 2016, S. 82). Markendaten auf der Nachfrageseite werden durch Marktforschung gesammelt, indem die Markenwahrnehmung, die Markenerwägung und die Markenpräferenz von Kunden gemessen wird (vgl. Verhoef et al. 2016, S. 82).

Die Kundendaten werden ebenfalls in Angebotsseite und Nachfrageseite unterteilt. Kundendaten auf der Angebotsseite sind Daten, die die Produkte und Dienstleistungen, die ein Kunde in der Beziehung mit einer Organisation genutzt hat, beschreiben (vgl. Verhoef et al. 2016, S. 84). Kundendaten auf der Nachfrageseite sind Daten, die die Erwartungen, die Zufriedenheit und die Interaktionen der Kunden mit einer Organisation beschreiben; diese sind meist sehr detailliert und genau (vgl. Verhoef et al. 2016, S. 84). Kundendaten der Angebotsseite werden meist in CRM-Systemen gespeichert, Kundendaten der Nachfrageseite kommen aus Marktforschungen, die beispielsweise die Kundenzufriedenheit messen (vgl. Verhoef et al. 2016, S. 84).

Es hat sich gezeigt, dass es sich bei Big Data um Daten von unterschiedlichem Strukturierungsgrad und aus unterschiedlichen Datenquellen handelt. Gerade unstrukturierte Daten stammen oft aus dem Internet und sind daher erst im Zuge der zunehmenden Digitalisierung zugänglich geworden. Die meisten Unternehmen haben bisher überwiegend mit strukturierten Daten gearbeitet, welche gut mit der

herkömmlichen IT-Infrastruktur eines Unternehmens verarbeitet werden können, beispielsweise mit Hilfe von Kundendatenbanken. Für die Verarbeitung von unstrukturierten Daten sind hingegen neue Technologien nötig.⁵ Die Unterscheidung zwischen Marktdaten, Produkt- und Markendaten und Kundendaten ist wichtig, um einordnen zu können, welche Daten welche Bereiche eines Unternehmens betreffen, aus welchen Quellen sie jeweils stammen, wie die Daten zugänglich sind, was sie abbilden und wie sie gespeichert werden können.

2.2.3 Datenspeicherung

Strukturierte Daten, die schon länger in Unternehmen verwendet werden, beispielsweise relevante kommerzielle Daten wie Verkaufszahlen, Produkte, Rechnungen, Kunden etc., werden in der Regel in Datenbanken gespeichert (vgl. Verhoef et al. 2016, S. 88). Wegen des großen Umfangs an Daten sind traditionelle Datenspeicherungs-Einrichtungen und -Managementmethoden allerdings für Big Data nicht geeignet, da die Serverleistung begrenzt ist (vgl. Chen et al. 2014, S. 33). Daher werden neue Datenspeichersysteme gebraucht, welche die Anforderungen von Big Data erfüllen können (vgl. Chen et al. 2014, S. 34). Zudem ist mit Datenbanken keine Echtzeitanalyse von Daten möglich, da diese nur auf die reine Speicherung von Daten angelegt sind.

Eine Lösung für die Speicherung von großen Datenmengen sind sogenannte Data Warehouses. Sie bereiten die Daten auf eine Speicherung vor und speichern, beschreiben, verwalten und kontrollieren die Daten (vgl. Verhoef et al. 2016, S. 87). Hauptzweck ist aber die dauerhafte Speicherung von großen Datenmengen, damit die Basis von Analysen, Planungsläufen oder Simulationen widerspruchsfreie historische Daten sind (vgl. Piller 2014, S. 93). Ein Data Warehouse stellt damit Datenkonsistenz für Geschäftsprozesse und Organisationen her und sichert die Stabilität von Analyseergebnissen (vgl. Piller 2014, S. 93). Ein Data Warehouse kann um softwarebasierte Informationslieferungs-Systeme erweitert werden, welche Daten in Informationen verwandeln (vgl. Verhoef et al. 2016, S. 87). Diese Informationen sind meist rein deskriptiv und beinhalten Metriken, beispielsweise die Zahl an aktiven Kunden, die Kundenabwanderungszahlen des letzten Monats oder den durchschnittlichen Absatz pro Kunde (vgl. Verhoef et al. 2016, S. 87 f.).

Neben der Speicherung der Daten ist ein Anspruch an Big Data, die Daten während der Speicherung bereits möglichst schnell analysieren zu können, bestenfalls in Echtzeit (vgl. Kraus 2013, S. 4). Datenbanken und Data Warehouses können allerdings keine Echtzeitanalyse bieten, weshalb hier die Streaming Analytics sowie das In-Memory-Datenmanagement zum Einsatz kommen können.

Streaming Analytics sollen Echtzeit-Daten so schnell bearbeiten, dass die Auswertung ebenfalls in Echtzeit geschieht (vgl. Ortega 2015, S. 55). Data Warehouses sind hier nicht geeignet, da die Datenströme ein sehr hohes Volumen und dabei eine

⁵ Dies wird im Kapitel 2.2.4 ab Seite 11 weiter erläutert.

kurze Verzögerungszeit haben und die Sammlung, Speicherung und Bearbeitung dieser Daten in Datenlagern zu zeitaufwendig wäre (vgl. Ortega 2015, S. 55). Um eine schnelle Auswertung gewährleisten zu können, verfügen Streaming-Analytics-Systeme über eine hohe Durchlaufgeschwindigkeit der Daten, da die Daten oft unstrukturiert sind (vgl. Ortega 2015, S. 56). Da sie Daten von großem Volumen in großer Geschwindigkeit analysieren können, sind Streaming Analytics gut für Big Data geeignet (vgl. Ortega 2015, S. 56).

Das In-Memory-Datenmanagement ermöglicht ebenfalls die Echtzeitanalyse von großen Datenvolumen und stellt damit eine Art Gegenentwurf zu den traditionellen Data Warehouses dar (vgl. Bachmann et al. 2014, S. 190). Laut Piller (2015, S. 83) ist mit In-Memory-Datenmanagement ein angestrebter Idealzustand in Unternehmen möglich, da sie jederzeit Zugriff auf aktuelle und historische Daten aus der gesamten Wertschöpfungskette gewähren und die Analyse dieser Daten in Echtzeit nach verschiedenen Aspekten ermöglichen. In-Memory-Datenbanken speichern die Daten – im Gegensatz zu traditionellen Data Warehouses, die die Daten auf einer Festplatte speichern – im Hauptspeicher (vgl. Bachmann et al. 2014, S. 191). Die großen Hauptspeicher und leistungsfähige Prozessoren bei In-Memory-Datenbanken ermöglichen das Belassen von großen Datenvolumen im Arbeitsspeicher und die dortige Bearbeitung der Daten mit hoher Geschwindigkeit (vgl. Piller 2015, S. 83). Zudem kann die Datenanalyse nun im Hauptspeicher ausgeführt werden, wo sich auch die Daten befinden, wohingegen die Daten in traditionellen Data Warehouses zur Analyse aus der Datenbank in eine Anwendungsschicht transportiert werden müssen (vgl. Bachmann et al. 2014, S. 193). Damit ist ein schnellerer Zugriff auf die Daten möglich, sie können besser durchsucht werden und die Performance ist gegenüber Data Warehouses deutlich besser (vgl. Bachmann et al. 2014, S. 192).

Wegen des stetig und schnell wachsenden Volumens an Daten sind neue Speichermöglichkeiten nötig geworden, da herkömmliche Datenbanken nicht die nötigen Kapazitäten für große Datenmengen besitzen. Für dieses Problem stellen Data Warehouses eine Lösung dar. Spezielle Datenspeicherungs- und -verarbeitungstechnologien wie Streaming Analytics und In-Memory-Datenbanken lösen das Problem des Anspruchs, Daten sehr schnell oder sogar in Echtzeit analysieren zu können.

2.2.4 Datenanalyse

Der wichtigste Bestandteil von Big Data ist die Analyse der Daten. Mit ihr sollen nützliche Informationen aus einer großen Menge an Daten herausgefiltert werden (Cheng et al. 2016, S. 4). Die analytische Funktion von Big Data ist zentral für Unternehmen und kann unterschiedliche Bereiche eines Unternehmens unterstützen (vgl. Verhoef et al. 2016, S. 119). Im Folgenden sollen daher verschiedene Arten der Analyse von großen Datenmengen erläutert werden.

Laut Verhoef et al. (2016, S. 123) gibt es vier grundlegende Analyse-Strategien für Big Data: Problemlösung, Datenmodellierung, Data Mining und Collateral Catch. Hierbei ist zu unterscheiden, ob die Analyse aufgrund eines vordefinierten Problems durchgeführt wird oder ob einfach verfügbare Daten analysiert und kombiniert werden (vgl. Verhoef et al. 2016, S. 123). Solche Probleme können durch Marketing-Herausforderungen wie abnehmende Kundenloyalität, sinkende Preise oder durch Wachstumsziele im Marketing definiert werden (vgl. Verhoef et al. 2016, S. 123).

Eine problemlösungsorientierte Analyse-Strategie ist deduktiv und soll ein Problem auf Management-Ebene lösen, beispielsweise, welche Preisstrategie die beste ist, um mehr Kunden anzulocken (vgl. Verhoef et al. 2016, S. 124). Nachdem das Problem hypothetisch definiert wurde, können explizite oder implizite Mutmaßungen darüber angestellt werden, wie das Problem gelöst werden könnte (vgl. Verhoef et al. 2016, S. 124). Big Data kann dabei helfen, das Problem zu lösen, ist aber keine Notwendigkeit dafür, denn die Probleme können beispielsweise auch mit limitierten, schon vorhandenen Daten, wie CRM-Datenbanken, oder mit anderen Mitteln der Datensammlung, wie einer Studie, gelöst werden (vgl. Verhoef et al. 2016, S. 124). Zudem beginnt die Problemlösung bei dieser Strategie nicht mit Datenanalyse wie beim Data Mining, sondern mit der Problemdefinition (vgl. Verhoef et al. 2016, S. 124).

Bei der Datenmodellierung steht ebenfalls immer ein zu lösendes Problem voran, beispielsweise um Anzeichen für eine eventuelle Kundenabwanderung zu finden (vgl. Verhoef et al. 2016, S. 124). Der Unterschied zur Problemlösungs-Analyse liegt darin, dass der Fokus hauptsächlich darauf liegt, das Problem mit Hilfe der Nutzung von neuen Datenquellen zu lösen (vgl. Verhoef et al. 2016, S. 124). Bei dieser Strategie ist darauf zu achten, dass Analysten sich nicht nur auf Daten an sich konzentrieren, sondern auch darauf, warum bestimmte Beziehungen in den Daten zu finden sind, um Scheinkorrelationen zu vermeiden (vgl. Verhoef et al. 2016, S. 124). Diese Strategie kann dennoch unerwünschte Ergebnisse produzieren und zu sehr datenfokussiert statt problemlösungsorientiert sein; der Vorteil liegt darin, dass beim Datenmodellierungs-Ansatz Daten flexibler zur Lösung des Problems genutzt werden können als beim Problemlösungs-Ansatz (vgl. Verhoef et al. 2016, S. 124 f.).

Data Mining ist im Gegensatz zu den beiden vorangegangenen Analyse-Strategien induktiv und sehr viel forschungsorientierter (vgl. Verhoef et al. 2016, S. 125 f.). Hier sind die Daten nicht vorgegeben und es steht auch kein Problem oder keine Hypothese voran (vgl. Verhoef et al. 2016, S. 125 f.). Bei dieser Strategie wird davon ausgegangen, dass verfügbare Daten allein wertvolle Einsichten bieten können, indem man in den Daten »wühlt« (vgl. Verhoef et al. 2016, S. 126). Dabei können Muster aufgedeckt werden – zum Beispiel dass, wenn ein Kunde ein Produkt der Kategorie A kauft, oft auch Produkte der Kategorien B und C kauft – und beispielsweise für gezielte Werbung verwendet werden (vgl. Verhoef et al. 2016, S. 126). Die Entdeckung solcher Muster kann Innovationen inspirieren (vgl. Verhoef et al. 2016,

S. 126). Ein Nachteil dieser Strategie ist, dass diese Analysen ungesteuert ausgeführt werden und daher viele Assoziationen aufdecken können, die schwer zu interpretieren sind (vgl. Verhoef et al. 2016, S. 126). Da kein konkretes Problem gelöst wird, sind laut Verhoef et al. (2016, S. 126) viele dieser Analysen von geringem Nutzen und daher wenig effektiv.

Collateral Catch ist keine explizite Strategie, sondern ein Nebenprodukt einer problemlösungsorientierten Analyse (vgl. Verhoef et al. 2016, S. 126). Wenn vordefinierte Daten analysiert werden, werden manchmal neue Beziehungen entdeckt, welche sehr wertvoll sein können und dann weiter untersucht werden (vgl. Verhoef et al. 2016, S. 126). Dies wird Collateral Catch genannt, da ursprünglich nicht danach gesucht wurde, und nur wegen ungewöhnlicher Merkmale in der Suche oder durch eine weiterforschende Analyse neben der Problemlösungs-Analyse entdeckt werden (vgl. Verhoef et al. 2016, S. 126). Datenanalysten sollten also offen für solche Collateral Catches sein, da diese neue Lösungen für ein Problem bieten oder Innovationen fördern können (vgl. Verhoef et al. 2016, S. 126).

Bei allen Analyse-Strategien geht es letztendlich darum, Muster zu erkennen. Muster können entscheidende Kenntnisse bieten, beispielsweise um das Kaufverhalten von Kunden vorherzusagen und ihnen entsprechende maßgeschneiderte Werbung liefern zu können (vgl. Floridi 2015, S. 34 f.). Thomas/McSharry (2015, S. 153) weisen allerdings darauf hin, dass die »leaders of the data revolution« (Thomas/McSharry 2015, S. 153) in der Lage sein müssen, Muster zu erkennen, zu verstehen, warum die Muster da sind, sie zu analysieren und sie anzuwenden, um Mehrwert zu erzeugen. Die Unterschiede in den verschiedenen Analyse-Strategien liegen darin, ob und inwieweit vorher schon ein Problem definiert ist, das mit Hilfe von Daten gelöst werden soll. Gibt es bereits ein vorher definiertes Problem, so kann nach spezifischen Mustern gesucht werden, die dieses Problem lösen können. Bei den Strategien ohne vordefiniertes Problem, besonders beim Data Mining, geht es hingegen darum, Muster zu entdecken, ohne vorher zu wissen, wonach genau gesucht wird.

Insgesamt verändert Big Data die Art, wie Datenanalysen ausgeführt werden, da die Daten nun aus vielen verschiedene Datenquellen stammen und es deshalb viel mehr unstrukturierte und oft komplexe Daten gibt und die Daten daher auf sinnvolle Weise verknüpft werden müssen (vgl. Verhoef et al. 2016, S. 127 f.). Dies betrifft Marktanalysen, Produkt- und Markenanalysen sowie Kundenanalysen (vgl. Verhoef et al. 2016, S. 128). So können Veränderungen auf dem Markt vorhergesagt, der Einfluss von Marketing auf Markenwahrnehmung und Verkaufszahlen bestimmt und Kundendaten mit Daten aus neuen Datenquellen – beispielsweise aus sozialen Netzwerken – angereichert und so das Kundenmanagement verbessert werden (vgl. Verhoef et al. 2016, S. 128–132). Wie die Analyse von Daten genau Unternehmensentscheidungen und -aktivitäten verbessern kann, in welchen Bereichen sie eingesetzt

werden kann und welche Veränderungen im Unternehmen dies mit sich bringt, soll im folgenden Kapitel erläutert werden.

2.2.5 Big Data im Unternehmen

Nachdem in den vorherigen Kapiteln erläutert wurde, worum es sich bei Big Data handelt, woher die Daten kommen und wie sie analysiert werden können, stellt sich nun die Frage, wie der Einsatz von Big Data im Unternehmen gehandhabt werden kann. Die Anwendung von Big Data im Unternehmen wird als Business Intelligence oder auch als Smart Data bezeichnet. Was genau unter diesen Begriffen zu verstehen ist, soll im Folgenden erläutert werden. Danach wird erklärt, in welchen Unternehmensbereichen Datenanalyse eingesetzt und in einem Unternehmen integriert werden kann.

Der Begriff Business Intelligence kommt ursprünglich aus der Wirtschaftsinformatik (vgl. Schön 2016, S. 291). Ebenso wie beim Begriff Big Data ist es schwierig, eine allgemeingültige Definition zu finden. Business Intelligence kann folgendermaßen definiert werden: »Business Intelligence beschreibt die geschäftsumfassende und übergreifende daten-gestützte Informationsbeschaffung, -analyse und -darstellung« (Haberich 2015, S. 69). King definiert Business Intelligence ähnlich, nämlich als eine »Managementstrategie [...], die eine strukturierte und effektive Entscheidungsfindung auf der Basis von Fakten ermöglichen soll« (King 2015, S. 37). Die ausführlichste und umfassendste Definition, welche die beiden vorherigen Definitionen zusammenfasst und erläutert, liefert Schön:

Business Intelligence ist die Integration von fachlichen Management-Methoden, IT-Verfahren und analytischen Prozessen, die sowohl die Aufbereitung und Bereitstellung von Daten als auch die Aufdeckung relevanter Zusammenhänge sowie die Kommunikation der gewonnenen Erkenntnisse zur Entscheidungsunterstützung für das Management umfassen, und hierzu insbesondere ausgewählte neue IT wie Data-Warehouse- und Big-Data-Technologie einsetzen (Schön 2016, S. 295).

Die Definitionen verdeutlichen, dass es bei Business Intelligence darum geht, Daten für das Management eines Unternehmens zu nutzen bzw. Daten systematisch in entscheidungsrelevante Informationen umzuwandeln (vgl. Ortega 2015, S. 55). Die Technologien für den Zugang, die Analyse und die Präsentation der relevanten Informationen bilden die Grundlage für diesen Ansatz (vgl. King 2015, S. 37). Business Intelligence beinhaltet das Datensammeln aus operativen Anwendungen, die Analyse der Daten und die Prozessoptimierung auf der Grundlage der Analyseergebnisse (vgl. King 2015, S. 38). Die Fundamente sind das Datenmanagement, das festlegt, wie Daten akquiriert und verwaltet werden, und die Data Warehouses, in denen Daten aus verschiedenen Quellen integriert werden (vgl. King 2015, S. 38). Die Daten sind hochaktuell und werden idealerweise sogar in Echtzeit gesammelt und analysiert (vgl. King 2015, S. 38). Inzwischen werden zunehmend Daten aus dem WWW für Business Intelligence gewonnen (vgl. King 2015, S. 39). Mit Business

Intelligence wird also die Anwendung von Daten-Strategien im Unternehmen bezeichnet. Damit bezeichnet Business Intelligence den Anschluss von Big Data als Technologie im ökonomischen Kontext mit dem Ziel, wirtschaftliche Vorteile zu generieren. Einsatzfelder für Business Intelligence sind beispielsweise Reporting, CRM, Planungs- und Budgetierungs-Systeme und Risiko-Management (vgl. Schön 2016, S. 295).

Für den Einsatz von Big Data in Unternehmen hat sich ebenfalls der Begriff Smart Data etabliert (vgl. z. B. Bloching et al. 2015, S. 10), welcher eine Weiterentwicklung des Begriffs Big Data darstellt. Als Smart Data werden Daten bezeichnet, die bereits Handlungsanweisungen beinhalten (vgl. Klausnitzer 2013, S. 96). Mit Smart Data wird also ebenfalls – ähnlich wie bei Business Intelligence – der Vorgang bezeichnet, aus der Analyse von Daten einen Nutzen zu gewinnen. Sowohl bei Smart Data als auch bei Business Intelligence geht es also vor allem um das Big-Data-Charakteristikum *value* von Big Data.

Es ist bereits deutlich geworden, dass Big Data für Unternehmen eine große Herausforderung darstellen kann, da die Infrastruktur, um Business Intelligence und Smart Data anzuwenden, im Unternehmen gegeben sein muss. Daher soll im Folgenden die Frage beantwortet werden, wie sich Big Data und speziell die Datenanalyse in ein Unternehmen integrieren lässt. Dafür muss zunächst herausgearbeitet werden, welche Bereiche in einem Unternehmen besonders von Big Data beeinflusst werden. Danach geht es darum, wie und unter welchen Voraussetzungen die Integration von Datenanalyse im Unternehmen gelingen kann.

Besonders das Marketing und der Kundenservice können mit Datenanalysen verbessert werden (vgl. Verhoef et al. 2016, S. 16 f.). Die Herausforderungen und Chancen im Marketing bestehen darin, mit der zunehmenden Menge an Daten bessere Einsicht in Nutzerverhalten zu erhalten (vgl. Verhoef et al. 2016, S. 1). Mit Datenanalysen können damit Marketing-Entscheidungen verbessert werden, beispielsweise bei der effektiven Aufteilung von Marketing-Budgets (vgl. Verhoef et al. 2016, S. 16). Außerdem können auf der Grundlage von Datenanalysen effektivere Marketing-Kampagnen durchgeführt werden, die spezifisch die richtige Zielgruppe ansprechen (vgl. Verhoef et al. 2016, S. 16).

Im Kundenservice können Daten dabei helfen, dem Kunden Kaufentscheidungen mit guten Algorithmen zu erleichtern (vgl. Verhoef et al. 2016, S. 17). Zudem verhindert das Fehlen von genutzten Daten die Personalisierung von Kundenservice, was dazu beiträgt, dass die Erwartungen an Kundenservice oft nicht erfüllt werden (vgl. Thomas/McSharry 2015, S. 75). Leggett (2012, zitiert nach Thomas/McSharry 2015, S. 75) nennt einige Bereiche, die besonders auf die erfolgreiche Nutzung von Big Data angewiesen sind. Diese sind das Prüfen von Kundenservice-Systemen, das Nutzen von Studien, um die Bedürfnisse von Kunden besser zu verstehen, und das Integrieren von Feedback-Schleifen, indem der Erfolg von Kundenservice-Inter-

aktionen mit den Kosten und der Erfüllung von Zielvorgaben gemessen wird (vgl. Leggett, zitiert nach Thomas/McSharry 2015, S. 75).

Nachdem mit Marketing und Kundenservice die Bereiche im Unternehmen identifiziert wurden, die besonders von Big Data profitieren können, soll im Folgenden eruiert werden, wie die Analyse von Big Data gezielt im Unternehmen integriert werden kann. Dabei muss beachtet werden, dass aus der Perspektive des Unternehmens dessen wirtschaftliche Ziele, wie etwa Ergebnisverbesserung und Wachstum, immer im Fokus sind; demnach sollten Big-Data-Aktivitäten die mittel- und langfristige Erreichung der Unternehmensziele unterstützen (vgl. Bachmann et al. 2014, S. 45).

Big Data soll durch geeignete Analysen neue Erkenntnisse generieren, indem Daten in entscheidungsrelevantes Wissen mit fachspezifischer Relevanz transformiert werden (vgl. Bachmann et al. 2014, S. 45 f.). Die Integration von Big Data im Unternehmen muss auf fachlicher und technologischer Ebene erfolgen, genauer in der internen Organisationsstruktur, in den Unternehmensprozessen, in der bestehenden IT-Architektur und in der Unternehmenskultur (vgl. Bachmann et al. 2014, S. 58). Damit Big Data in Unternehmen effektiv eingesetzt werden kann, müssen die neuen Technologien, wie beispielsweise In-Memory-Datenbanken, in die bereits bestehenden IT-Strukturen integriert werden (Buhl et al. 2013, S. 65). Außerdem müssen die passenden Analyse-Werkzeuge zur Verfügung stehen, um die unterschiedlichen Daten aus meist mehreren Quellen auswerten zu können (Buhl et al. 2013, S. 65).

Ein strategisches Ziel von Unternehmen im Zusammenhang mit Big Data sollte es sein, sich zu einem analytischen Wettbewerber zu entwickeln, denn mit Datenanalysen lassen sich Wettbewerbsvorteile erzeugen, die für Unternehmen zu einem Erfolgsfaktor werden können (vgl. Bachmann et al. 2014, S. 48). Dies gilt insbesondere für Unternehmen, deren Geschäftsmodell stark auf Daten gestützt ist, und welche daher in einem Wettbewerb um die wirkungsvollsten Datenanalysen stehen (vgl. Bachmann et al. 2014, S. 48). Laut Bachmann et al. (2014, S. 48) werden die Märkte in der Zukunft von Unternehmen beherrscht werden, welche ihre Datenanalysen am effizientesten für ihre Unternehmensstrategie einsetzen können und dabei neue Wertschöpfungen, Geschäftsmodelle und Marketingmethoden entwickeln können. Um mit Hilfe von Datenanalysen zu neuen Erkenntnissen zu gelangen, müssen in einem Unternehmen die internen Rahmenbedingungen und Strukturen stimmen oder geschaffen werden; das bedeutet, dass eine große Bereitschaft zu substantiellen Veränderungen in der Unternehmenskultur vorhanden sein muss: Eine hohe Fehlertoleranz der Unternehmenskultur erlaubt allen Beteiligten, sich für neue Verhaltens- und Denkmuster zu öffnen und vorhandene Strukturen in Frage zu stellen (vgl. Bachmann et al. 2014, S. 49, 51). Neben Änderungen in der allgemeinen Unternehmenskultur können außerdem personelle Veränderungen oder die Schaffung von neuen Positionen nötig sein; ein Unternehmen benötigt beispiels-

weise oft neue Experten der Datenanalyse, sogenannte Data Scientists, welche es verstehen, Problemlösungen anhand von Daten kreativ und engagiert anzugehen (vgl. Bachmann et al. 2014, S. 50).

Haben Unternehmen es geschafft, Datenanalysen in die bestehende Unternehmensstruktur zu integrieren, stehen sie mit anderen Unternehmen im Wettbewerb um die Informationsvorherrschaft und die Deutungshoheit über Daten und Informationen (vgl. Bachmann et al. 2014, S. 52). Dabei geht es darum, welches Unternehmen die relevanten Daten für ein Thema besitzt und welches diese in valide Informationen und Wissen transformieren und dieses Wissen richtig deuten kann (vgl. Bachmann et al. 2014, S. 53).

Zusammenfassend lässt sich sagen, dass die Analyse von Big Data einem Unternehmen die Möglichkeit bietet, besonders in den Bereichen Marketing und Kundenservice, einen Mehrwert zu generieren. Dabei müssen die Integration von Datenanalysen im Unternehmen und der strategisch richtige Umgang mit Big Data sichergestellt sein, sodass Business Intelligence und Smart-Data-Konzepte eingesetzt werden können.

2.2.6 Chancen und Risiken von Big Data

Es hat sich gezeigt, dass der Einsatz von Big Data – genauer der Einsatz von Datenanalysen – einem Unternehmen einige Chancen bieten kann. Dies betrifft vor allem die Unterstützung von Entscheidungsfindungen in einem Unternehmen. Das ist dadurch möglich, dass durch Datenanalyse neue Blickwinkel auf die Umwelt, vor allem auf die Marktsituationen und die Kunden, ermöglicht werden (vgl. Fasel 2014, S. 399). Darüber hinaus sollen mit Big Data auch Vorhersagen getroffen werden können, welche dabei helfen können, Produkte besser an Kundenbedürfnisse anzupassen und ein gezielteres und effektiveres Produktmarketing und Reputationsmanagement zu betreiben (vgl. Fasel 2014, S. 399).

Big Data bringt aber auch einige Risiken mit sich. Die bedeutendsten sind die Möglichkeit von Scheinkorrelationen und Datenschutzbedenken. Diese sollen an dieser Stelle erörtert werden.

Bei der Datenanalyse muss das Risiko beachtet werden, dass mehr Scheinkorrelationen entstehen können, wenn viele Daten ausgewertet werden (vgl. Fasel 2014, S. 399). Gerade bei großen Datenbeständen aus unterschiedlichen Quellen kann es schwierig sein, einen effektiven Mehrwert für ein Unternehmen zu erhalten (vgl. Fasel 2014, S. 399). Daher werden immer Data Scientists oder auch Data Science-Teams gebraucht werden, also Experten, die das technische, mathematische und wirtschaftliche Know-how haben, um aus Big Data für ein Unternehmen Nutzen generieren zu können (vgl. Fasel 2014, S. 399).

Das größte Risiko beim Einsatz von Big Data ergibt sich aber aus Kundenperspektive: Ansatzpunkt für massive Kritik ist der Datenschutz. Big Data hat dazu geführt, dass aus persönlichen Daten von Personen ein Wirtschaftsgut geworden ist,

mit dem Geld verdient werden kann (vgl. Haustein 2015, S. 254). Weichert (2013, S. 133) stellt fest, dass Big Data daher mit Risiken verbunden ist, sobald es sich um personenbezogene Daten handelt. Weltweit sind Datenschutzbehörden besorgt und verfolgen die zunehmende personalisierte Werbung mittels Anwendung von Big Data und thematisieren dies auf internationalen Konferenzen (vgl. Scheuing 2015, S. 116). Im Jahr 2013 führte die US-Verbraucherschutzbehörde Federal Trade Commission (FTC) eine Untersuchung durch, um die Geschäftsmodelle von Datendienstleistern zu verstehen (vgl. Scheuing 2015, S. 116). Sie fand heraus, dass das Risiko besteht, dass durch Kategorisierung von Kunden mit Hilfe von datenbasiertem Marketing einzelne Personen diskriminiert werden könnten, beispielsweise, wenn ihnen aufgrund von Daten ein Vertragsabschluss verwehrt würde (vgl. Scheuing 2015, S. 116; vgl. Weichert 2013, S. 134). Außerdem ist Big Data sehr kompliziert, sodass ein durchschnittlicher Verbraucher nicht verstehen kann, was dabei stattfindet (vgl. Scheuing 2015, S. 116). Ein wichtiges Konzept für den Datenschutz ist der Schutz des informationellen Selbstbestimmungsrechts, welches sowohl im Bundesdatenschutzgesetz als auch in der EU-Datenschutzrichtlinie festgeschrieben ist (vgl. Scheuing 2015, S. 117). Diese Gesetze legen eine Transparenzpflicht für Werbetreibende fest, was bedeutet, dass die Verbraucher darüber informiert werden müssen, was Werbetreibende mit den Daten machen möchten; dies wird meist über Datenschutzerklärungen erfüllt (vgl. Scheuing 2015, S. 117). Weiterhin gilt das Zweckbindungsprinzip, wonach personenbezogene Daten nur für die Zwecke verwendet werden dürfen, für die sie gespeichert wurden (vgl. Weichert 2013, S. 135). Der Verbraucher wiederum kann von einem Widerspruchsrecht Gebrauch machen und damit festlegen, ob er die Datennutzung erlaubt oder nicht (vgl. Scheuing 2015, S. 117). Außerdem haben Verbraucher ein Auskunftsrecht, und können daher herausfinden, welche Daten über sie bei einem bestimmten Unternehmen gespeichert sind (vgl. Scheuing 2015, S. 117). Scheuing (2015, S. 124) gibt allerdings zu Bedenken, dass Verbraucher meist den Datenschutzerklärungen zustimmen, ohne sie gelesen zu haben, und Unternehmen damit die personenbezogenen Daten nutzen können, ohne dass die Verbraucher wissen, welchen Bedingungen sie zugestimmt haben, oder die Risiken einschätzen können. Seemann (2014, S. 117) spricht sogar von einem Kontrollverlust für die Nutzer, da sie nicht mehr wissen können, welche Daten erhoben werden und vor allem, welche Aussagen die Analyse der eigenen Daten zulassen. Die Rechtsprechung fordert daher technisch-organisatorische und verfahrensrechtliche Vorkehrungen, um das Recht auf informationelle Selbstbestimmung zu schützen (vgl. Weichert 2013, S. 135 f.).

Es ist deutlich geworden, dass Big Data und damit vor allem die Analyse der Daten Unternehmen viele Chancen bieten können. Der Unterschied zwischen Big Data und konventionellen Datensammlungen (etwa in Kundendatenbanken), besteht darin, dass mit den neuen massenhaften Daten aus vielen verschiedenen Datenquellen Vorhersagen getroffen werden können und sollen. Zu berücksichtigen sind

jedoch die damit verbundenen Risiken. Kunden sollten nicht das Gefühl haben, dass sie bei der Sammlung ihrer Daten durch Unternehmen kein Mitspracherecht haben.

2.3 Fallbeispiele

In einigen Branchen und großen Unternehmen wird Big Data schon erfolgversprechend eingesetzt. Im Folgenden sollen zwei Fallbeispiele erläutern, wie der Einsatz von Big-Data-Konzepten in der Praxis aussehen kann. Als Beispiele dienen dafür die Streamingdienste Netflix und Spotify und Amazon. Streamingdienste wurden als Beispiel gewählt, da sie neue Modelle des Vertriebs von digitalen Medien entwickelt haben, die ihnen die Möglichkeit gewähren, aus den hierbei gewonnenen Daten weit mehr über ihre Nutzer zu erfahren. Auch in der Buchbranche findet Digitalisierung statt, beispielsweise in Form von E-Books. Auch hier sind neue Vertriebsmodelle entstanden, wie beispielsweise Selfpublishing-Plattformen oder E-Book-Flatrate-Anbieter. Daher können die Streamingdienste im Bezug auf die Nutzung von Nutzerdaten als Beispiel für die Buchbranche dienen. Amazon wurde gewählt, da der Einfluss dieses Unternehmen seit Jahren in der Buchbranche diskutiert wird. Dabei geht es unter anderem auch darum, dass Amazon viele Nutzerdaten sammelt.

2.3.1 Streamingdienste – Netflix und Spotify

Netflix

Netflix ist ein Service, bei dem man für die Zahlung eines festen monatlichen Betrags Zugriff auf viele Serien und Filme hat, die gestreamt und teilweise sogar offline gespeichert werden können. Netflix ist in fast allen Ländern der Welt verfügbar⁶ und hatte im Oktober 2016 86 Millionen Nutzer (vgl. Fiegermann 2016).

Netflix erfasst nach eigenen Angaben zunächst strukturierte Kundendaten wie den Namen, die E-Mail-Adresse, die Zahlungsart etc. der Nutzer (vgl. Netflix 2016). Darüber hinaus werden auch sowohl strukturierte als auch unstrukturierte Daten aus interner Quelle gespeichert, beispielsweise wie die Nutzer mit Netflix interagieren und welche Geräte die Nutzer verwenden, um auf Netflix zuzugreifen (vgl. Netflix 2016). Diese beinhalten im Detail die Nutzeraktivitäten wie Titelwahl, Sehverlauf und Suchanfragen, Daten zu Interaktionen mit dem Kundendienst, Geräte-IDs, Geräte- und Software-Eigenschaften, Verbindungsinformationen, Statistiken über Seitenabrufe, IP-Adressen, Werbedaten etc. (vgl. Netflix 2016). Zudem werden diese Informationen mit Daten aus externen Quellen ergänzt, wie beispielsweise demografische Daten, interessenbasierte Daten und Daten zum Internet-Nutzungsverhalten (vgl. Netflix 2016). Laut Marr (2016, S. 20) erhebt Netflix zudem Daten

⁶ Lediglich in China, Nordkorea, Syrien und auf der Krim ist Netflix nach eigenen Angaben nicht verfügbar (vgl. <https://media.netflix.com/en/press-releases/netflix-is-now-available-around-the-world> [06.01.2016 / 24.10.2016])

darüber, zu welcher Tageszeit Filme und Serien gesehen werden, wie viel Zeit zur Auswahl eines Films verwendet wird und wie oft ein laufender Film pausiert wird.

Netflix setzt zur Datenanalyse Algorithmen ein, die vorhersagen können, wie ein Nutzer einen Film auf Grundlage seiner vorherigen Bewertungen von Filmen bewerten würde (vgl. Marr 2016, S. 18). Zunächst standen dafür nur die Nutzer-ID, die Film-ID, die Bewertung und das Datum, an dem der Film gesehen wurde, zur Verfügung, doch mit dem Ausbau des Streaming-Angebots ergaben sich die oben genannten, neuen Datenquellen (vgl. Marr 2016, S. 18). Außerdem verwendet Netflix sogenanntes Tagging: Filme werden von professionellen Bewertern mit verschiedenen Kennzeichen (tags) versehen; sieht ein Nutzer einen Film bei Netflix, werden ihm daraufhin Filme vorgeschlagen, die ähnliche Kennzeichen aufweisen (vgl. Marr 2016, S. 18 f.). Dies hat den Zweck, die überwiegend unstrukturierten Daten in Form von Video- und Audiodateien in strukturierte und analysefähige Daten zu transformieren, in dem Fall in über 80.000 Mikro-Genres (vgl. Marr 2016, S. 21 f.). Es wird vermutet, dass Netflix versucht, diesen Vorgang zu automatisieren (vgl. Marr 2016, S. 22).

Netflix produziert zudem auch eigene Dokumentationen oder ganze Serien. Dies geschieht auf der Grundlage von Data Mining und Datenmodellierung-Analyse der Daten, die von den Nutzern von Netflix gesammelt werden, um herauszufinden, was die Nutzer gerne sehen möchten und vorhersagen zu können, was den Nutzern gefallen wird, um auf dieser Grundlage eigene Produktionen in Auftrag geben zu können (vgl. Marr 2016, S. 17 f.).⁷

Ziel der Big-Data-Strategien von Netflix ist es, neue Abonnenten durch das Anbieten von beliebten und qualitativ hochwertigen Filmen und Serien zu gewinnen, und die Zeit, die ein Nutzer auf Netflix verbringt, zu erhöhen, damit er sein Abonnement nicht kündigt (vgl. Marr 2016, S. 19). Die Big-Data-Aktivitäten von Netflix können als Erfolg gewertet werden, da die Zahl der Abonnenten stetig ansteigt und die Nutzer immer mehr Filme und Serien, speziell auch die eigenen Produktionen, auf Netflix streamen (vgl. Marr 2016, S. 20). Die Möglichkeit vorhersagen zu können, was die Nutzer gerne sehen wollen, ist ein großer Teil des Erfolgs von Netflix (vgl. Marr 2016, S. 20).

Spotify

Spotify ist ein Streaming-Dienst für Musik. Dieser wird ebenfalls entweder für zahlende Kunden über ein Abo-Modell oder für diejenigen, die Spotify kostenlos nutzen, über Werbung finanziert. Spotify hat weltweit 100 Millionen aktive Nutzer

⁷ Die Netflix-Serie *House of Cards* entstand so auf der Basis von Daten, die die Nutzerwünsche abbilden. Die Datenanalyse ergab, dass die Nutzer gerne Inhalte sehen, bei denen David Fincher Regie geführt hat und bei denen der Schauspieler Kevin Spacey mitspielt. Die Hauptperson in *House of Cards* wird also von Kevin Spacey gespielt und David Fincher führt bei der Serie Regie. Selbst die Farben für das Titelbild der Serien wurden auf Basis von Daten bestimmt. Diese Serie ist tatsächlich ein großer Erfolg auf Netflix (vgl. Marr 2016, S. 19).

(Stand Juni 2016), davon sind 40 Millionen zahlende Abonnenten (Stand September 2016) (vgl. Statista 2016a, vgl. Statista 2016b).

Ähnlich wie Netflix verwendet Spotify die Daten seiner Nutzer aus interner Quelle, um Angebot und Empfehlungen zu verbessern. Dazu sammelt und speichert Spotify Daten, wenn ein Nutzer Lieder über Spotify anhört, und erstellt ein Nutzerprofil, das den Musikgeschmack eines Nutzers wiedergibt (vgl. Schlüter 2016). Außerdem sind Playlists für Spotify eine gute Datenquelle. Playlists sind Listen mit Liedern, beispielsweise zu bestimmten Themen, Stimmungen oder historischen Zeiträumen. Diese können von Spotify zur Verfügung gestellt werden, aber auch Nutzer können selbst Playlists erstellen. Zudem wird für einen Nutzer auf Basis von Algorithmen jeden Montag eine Playlist erstellt mit Liedern, die ihm vermeintlich gefallen könnten (vgl. Schlüter 2016). Empfehlungen und die wöchentliche, persönliche Playlist basieren dann zum einen auf dem Musikgeschmack des Nutzers, den Spotify gespeichert hat, und zum anderen auf Playlists, die Nutzer erstellt oder gehört haben, die ein ähnliches Nutzerprofil aufweisen (vgl. Schlüter 2016). Damit ist ein Teil der Empfehlungen von Spotify der Kreativität von vielen Nutzern zu verdanken, welche Playlists erstellen (vgl. Schlüter 2016).

Ein Beispiel für weitere Daten, die Spotify sammelt und verwertet, ist die sogenannte Skipping-Rate, die angibt, wie oft ein Lied übersprungen wurde (vgl. Güßgen 2016). Lieder mit einer hohen Skipping-Rate werden aus einer von Spotify veröffentlichten Playlist wieder entfernt (vgl. Güßgen 2016).

Auch bei Spotify hat das Sammeln, Speichern und Analysieren von Daten einen großen Beitrag am Erfolg: Je mehr Daten Spotify sammeln kann, desto besser werden die Empfehlungen, was wiederum mehr Nutzer anzieht, welche mehr Einnahmen generieren (vgl. Rijmenam 2016). Mit diesem Konzept hat Spotify laut Rijmenam (2016) die Musikindustrie nachhaltig verändert.

Spotify verwendet aber die Daten nicht nur intern, sondern erfasst über die mobilen Endgeräte ihrer Nutzer auch Kontakte, Fotos und Mediendateien sowie Ortungsdaten; Spotify ist zudem mit Facebook verknüpft und erhält dadurch auch die Facebook-Daten seiner Nutzer (vgl. Zeit Online 2015). Dies wiederum dient zwar der Verbesserung des Nutzerservices, soll aber auch bessere Werbemöglichkeiten generieren (vgl. Zeit Online 2015). Dafür teilt Spotify gesammelte Daten mit Werbekunden (vgl. Zeit Online 2015). Dies wiederum ist ein gutes Beispiel dafür, dass der Datenschutz bei Big Data ein kritisches Thema ist und die Kunden und Nutzer schnell die Hoheit über ihre eigenen Daten verlieren können.

2.3.2 Amazon

Amazon ist einer der größten Online-Händler der Welt. Das Unternehmen verkauft neben physischen auch virtuelle Güter wie E-Books und Videostreams (vgl. Marr 2016, S. 287). Der große Erfolg von Amazon basiert unter anderem auf der Technologie, Kunden Empfehlungen zu geben auf Basis dessen, was sie bisher

gekauft oder angesehen haben, und damit vorherzusagen, was Kunden möchten und wann sie es möchten (vgl. Marr 2016, S. 287). Inzwischen verkauft Amazon aber nicht nur Güter, sondern stellt auch selbst welche her; dazu gehören – wie bei Netflix – eigene Filme und Serien sowie Elektroartikel wie Tablets, TV-Boxen und Streaming-Hardware (vgl. Marr 2016, S. 287).

Amazon verwendet strukturierte, interne Daten von Kunden, die gesammelt werden, während sie auf der Website stöbern, um die Empfehlungen zu verbessern, beispielsweise wie lang ein Kunde auf einer Seite verweilt und welche Produkte er anklickt (vgl. Marr 2016, S. 288, 290). Dahinter steht die theoretische Annahme, dass steigendes Wissen über das Kaufverhalten eines potenziellen Kunden zu besseren Vorhersagen über seine zukünftigen Kaufabsichten führt (vgl. Marr 2016, S. 288). Der Kaufprozess kann dem Kunden dann insofern erleichtert werden, als er das gewünschte Produkt gar nicht mehr suchen muss, sondern es ihm gleich vorgeschlagen wird (vgl. Marr 2016, S. 288). Die Empfehlungen gelingen über kollaboratives Filtern: Es wird ein Profil über eine Person erstellt, auf dessen Basis Produkte vorgeschlagen werden, welche von Kunden mit einem ähnlichen Profil erworben wurden (vgl. Marr 2016, S. 289). Dies gelingt überwiegend über strukturierte Daten wie Metadaten; dazu gehören der Name eines Produkts, wie viel das Produkt kostet, wer es gekauft hat etc. (vgl. Marr 2016, S. 289). Amazon sammelt zudem sowohl strukturierte als auch unstrukturierte Daten, die entstehen, während Kunden die verschiedenen Services von Amazon verwenden (vgl. Marr 2016, S. 289). Dazu gehören Daten darüber, was ein Kunde kauft, was für Produkte er sich ansieht, seine Lieferadresse – um demografische Daten zu erhalten – und die Tatsache, ob ein Kunde Rezensionen oder Feedback abgegeben hat (vgl. Marr 2016, S. 289). Außerdem wird aufgezeichnet, zu welcher Tageszeit ein Nutzer auf der Website stöbert, um Gewohnheiten herauszufinden und diese Daten mit Daten anderer Kunden abzugleichen, die ähnliche Gewohnheiten haben (vgl. Marr 2016, S. 289). Beim Videostream-Angebot von Amazon werden Daten dazu gesammelt, wie viel Zeit ein Kunde dafür verwendet, Videos zu schauen, und wann, wo und wie er das tut (vgl. Marr 2016, S. 289 f.). Bei den E-Books wird das Leseverhalten der Kunden aufgezeichnet, beispielsweise die Leseposition – also an welcher Stelle des Buches sich der Kunde gerade befindet – und die Lesedauer und ob ein Kunde das Buch zu Ende liest oder nicht (vgl. Leisegang 2014, S. 97). Auch unstrukturierte Daten wie Hervorhebungen und Anmerkungen der Leser im E-Book werden gespeichert (vgl. Leisegang 2014, S. 97). Mit all diesen Daten soll laut Marr (2016, S. 289) eine 360-Grad-Sicht auf jeden individuellen Kunden entstehen. Im Jahr 2013 begann Amazon, diese Daten auch an Werbeagenturen zu verkaufen, damit diese eigene Big-Data-basierte Marketing-Kampagnen initiieren können (vgl. Marr 2016, S. 289). Auch hier lässt sich anmerken, dass Kunden die Kontrolle über die Nutzung ihrer Daten schnell verlieren können und die Datenschutzbedenken bei Big Data daher berechtigt sind.

Zusammenfassend lässt sich sagen, dass Amazon Big Data hauptsächlich dafür verwendet, Profile über jeden einzelnen Kunden oder potentiellen Kunden zu erstellen, um diese Profile mit anderen Profilen abgleichen zu können. Auf Basis von ähnlichen Profilen können einem Kunden dann Empfehlungen präsentiert werden. Diese Empfehlungen sind wichtig, da das riesige und vielfältige Angebot von Amazon die Kunden sonst überfordern könnte und sie damit davon abgehalten würden, etwas zu bestellen (vgl. Marr 2016, S. 291). Marr fasst die Philosophie von Amazon so zusammen: »The more a business knows about a customer, the better it can sell to them« (Marr 2016, S. 291).

2.4 Fazit

Daten, beispielsweise in Kundendatenbanken, sind schon lange in den meisten Unternehmen vorhanden. Wegen der zunehmenden Digitalisierung stehen nun aber immer mehr Daten und vor allem Daten aus ganz neuen Datenquellen zur Verfügung. Diese neuen Quellen können beispielsweise soziale Netzwerke oder Apps auf Smartphones sein. So können Kundendaten mit neuen Daten angereichert werden und es kann ein umfassender Einblick in die Kundenbedürfnisse entstehen. Diese neuen Daten zeichnen sich dadurch aus, dass sie ein großes Volumen haben und daher schnell verarbeitet werden müssen und sehr vielfältig sind. Vor allem wegen des großen Datenvolumens hat sich der Begriff »Big Data« durchgesetzt.

Im besten Falle verschaffen diese Daten einem Unternehmen einen Mehrwert. Die Herausforderung besteht darin, diese große Menge an Daten zu sammeln, zu speichern und so zu analysieren, dass sie einen Erkenntnisgewinn generieren und damit die Entscheidungsfindungen in einem Unternehmen unterstützen können. Vor allem im Marketing können die Daten dabei helfen, das Budget sinnvoll einzusetzen, indem die Kunden zielgerichtet mit Werbebotschaften angesprochen werden können, da nun eine ganz neue, umfassendere Einsicht in die Kunden möglich ist.

Der Einsatz von Big Data in einem Unternehmen ist zunächst mit Investitionen in neue Technologien wie beispielsweise Analyse-Software und einer Umstrukturierung und Neuorientierung innerhalb des Unternehmens verbunden. Wenn Big Data dann im Unternehmen mit einem guten Konzept effektiv und sinnvoll eingesetzt wird, kann aus Big Data Smart Data werden, was letztendlich den Mehrwertgewinn für ein Unternehmen bedeutet. Dabei müssen aber immer zum einen die Gesetze und Richtlinien zur Bewahrung des Datenschutzes beachtet werden, zum anderen ist – vor allem aus der Perspektive des Kunden – auch ein freiwilliger, verantwortungsvoller Umgang mit den Kundendaten wünschenswert.

Da Big Data seit Jahren ein immer stärker diskutiertes Thema in der Wirtschaft ist und die Datenquellen weiter zunehmen und immer besser ausgeschöpft werden können, ist davon auszugehen, dass sich Big Data immer weiter durchsetzen wird und immer mehr Branchen und Unternehmen Big Data bzw. Smart Data einsetzen werden. Unternehmen, die sich dem verweigern, werden daher wahrscheinlich

Nachteile im Wettbewerb mit anderen Unternehmen in Kauf nehmen müssen. Auf der anderen Seite ist zu bedenken, dass viele Kunden Datenschutzbedenken haben, daher könnten Unternehmen, die bewusst auf das Sammeln von Daten verzichten und dies kommunizieren, das Vertrauen der Kunden gewinnen.

Auch die Buchbranche ist von diesem Umbruch betroffen und muss sich jetzt oder spätestens in naher Zukunft mit dem Konzept Big Data beschäftigen und entscheiden, ob und inwieweit Big Data beispielsweise in Verlagen sinnvoll eingesetzt werden kann. Im folgenden Kapitel wird daher das Potenzial von Big Data in der Buchbranche beleuchtet.

3. Big Data in der Buchbranche

Im folgenden Kapitel soll es um Big Data in der Buchbranche gehen, um am Ende die Forschungsfrage – welches Potenzial hat Big Data in der Buchbranche – beantworten zu können. Dafür ist das Kapitel in drei Unterkapitel unterteilt. Im ersten Unterkapitel (Kapitel 3.1) geht es um die Buchbranche, insbesondere darum, warum der Buchmarkt ein spezieller Markt ist und wie das Marketing im Buchmarkt bisher funktioniert hat. Im zweiten Unterkapitel (3.2) soll der Status quo in der Buchbranche in Bezug auf den Einsatz von Daten und Big Data beschrieben werden. Zunächst wird erläutert, welche Daten in der Buchbranche schon verwendet werden. Danach wird dargelegt, wie die Diskussion um Big Data in der Buchbranchenpresse bisher geführt wurde. Zuletzt soll in diesem Unterkapitel vertiefend auf das Thema Reader Analytics eingegangen werden. Im dritten Unterkapitel (Kapitel 3.3.) wird auf Basis der vorher gewonnenen Erkenntnisse das Potenzial von Big Data für die Buchbranche anhand von vier Fragen analysiert.

3.1 Voraussetzungen für Big Data in der Buchbranche

Damit sich Big Data innerhalb der Buchbranche etablieren kann, müssen einige Voraussetzungen berücksichtigt werden. Zwei Aspekte sind hier besonders bedeutungsvoll, weshalb im folgenden Kapitel näher darauf eingegangen wird: Zum einen ist die Besonderheit des Buchmarkts aufgrund von unterschiedlichen Funktionen des Buchs relevant und zum anderen muss das Marketing im Buchmarkt, wie es bisher praktiziert wurde, untersucht werden, da Big Data insbesondere von Bedeutung für das Marketing von Unternehmen sein kann.

3.1.1 Besonderheiten des Buchmarkts

Das Besondere am Buchmarkt ist, dass Bücher, wie die meisten Medien, keine Waren oder Güter wie andere sind, da sie einen Doppelcharakter haben: Sie sind zum einen Wirtschaftsgüter und haben daher Eigenschaften, welche sie ökonomisch bedeutend machen, und zum anderen sind sie Kulturgüter, was bedeutet, dass sie Eigenschaften haben, welche von gesellschaftlicher Relevanz sind (vgl. Rimscha/Siegert 2015, S. 23). Der Buchmarkt ist somit sowohl einem wirtschaftlichen als auch einem kulturellen System zugehörig (vgl. Meyer-Dohm 1974, S. 433). Diese beiden Systeme müssen daher miteinander kooperieren, obwohl sie unterschiedlichen Leitwerten und Steuerungsmechanismen unterliegen, was zu Zielkonflikten führen kann (vgl. Beyer/Carl 2008, S. 12). Dieser Zusammenhang der Eigenschaften von Medien, die zwei verschiedenen Systemen zuzuordnen sind, soll im Folgenden dargelegt werden. Die spezielle Charakteristik von Medienmärkten muss auch beim Einsatz von Big Data berücksichtigt werden.

In seiner Funktion als Wirtschaftsgut gilt das Buch als ein privates Gut, da es die beiden Kriterien Ausschlussprinzip – Personen, die für das Gut nicht bezahlen, sind

von dessen Konsum und Nutzung ausgeschlossen – und Konsumrivalität – ein Gut wird beim Konsum oder bei der Nutzung verbraucht – erfüllt, und kann daher unproblematisch über den Markt gehandelt werden (vgl. Rimscha/Siegert 2015, S. 25). Dies gilt jedoch nur für das Buch als Träger eines Inhalts, da sich der Inhalt selbst nicht verbraucht (vgl. Rimscha/Siegert 2015, S. 25). Der Druck eines Buches oder die Produktion eines E-Books allerdings machen einen Inhalt zu einem privaten Gut, und damit konsumierbar und marktfähig (vgl. Keuschnigg 2012, S. 51).

Auf der anderen Seite gelten Bücher als Kulturgut, da sie als Medien der Wissensvermittlung und des kulturellen Austausches gelten (vgl. Weuster 2007, S. 62). An die Bestimmung des Buches als Kulturgut sind kulturpolitische Zielsetzungen geknüpft, wie jene, dass eine hohe Vielfalt an Büchern und ein flächendeckender Zugang zu Büchern wünschenswert ist (vgl. Keuschnigg 2012, S. 51). Daher werden Bücher auch als meritorische Güter bezeichnet. Meritorische Güter sind Güter, deren Nachfrage nach Ansicht des Staates zu gering ist, da die Bürger die gesellschaftliche Bedeutung des Gutes nicht genügend anerkennen (vgl. Beck 2011, S. 10). Der Konsum von Büchern ist gesellschaftlich erwünscht, weil damit die Hoffnung auf positive externe Effekte verknüpft ist (vgl. Rimscha/Siegert 2015, S. 26). Externe Effekte sind Auswirkungen, die nicht im Preis des Gutes enthalten sind; mit positiven externen Effekte ist meist die Erfüllung gesellschaftlicher Funktionen gemeint (vgl. Rimscha/Siegert 2015, S. 26). Beispielsweise sollen mit Medienangeboten die Bürger politisch umfassend informiert werden, damit diese überlegte Wahlentscheidungen treffen können und »staatsbürgerliche Handlungskompetenz« (Rimscha/Siegert 2015, S. 26) erhalten. Medien sind also relevant für den Meinungsbildungsprozess und daher wesentlicher Bestandteil der Demokratie und stellen eine Selbstbeobachtungsinstanz der Gesellschaft dar (vgl. Rimscha/Siegert 2015, S. 36). Der leichte Zugang zu Medien, beispielsweise über Bibliotheken, sowie eine Vielfalt an Medieninhalten, die Themen aus unterschiedlichen Blickwinkel beleuchten, sind daher von hoher gesellschaftlicher Bedeutung.

Durch die Einordnung des Buches als Kulturgut und meritorisches Gut wird die Produktion von Büchern vom Staat mit Beschränkungen der Marktgesetze geschützt, weil befürchtet wird, dass die Nachfrage allein nicht ausreichend ist, um die gesellschaftlichen Funktionen zu erfüllen (vgl. Rimscha/Siegert 2015, S. 26 f.; vgl. Bramann/Cremer 2014, S. 158). Das wichtigste Instrument hierfür ist das Buchpreisbindungsgesetz, welches seit 2002 bundesweit gesetzlich verankert ist (vgl. Keuschnigg 2012, S. 54 f.).⁸ Aufgrund des Buchpreisbindungsgesetzes sind Verlage dazu verpflichtet, verbindliche Preise für die Endkunden festzulegen, und der Einzelhandel ist verpflichtet, Bücher generell zu dem festgesetzten Endverkaufspreis abzugeben (vgl. Bramann/ Cremer 2014, S. 158; vgl. Janello 2010, S. 66). Der Zweck des Gesetzes ist, dass eine große Vielfalt an Titeln vorhanden ist, kulturell wertvolle

8 Die Buchpreisbindung gibt es bereits seit 1887 und wurde durch Einzelabsprachen geregelt, ab 1966 dann durch Multilateralverträge (vgl. Keuschnigg 2012, S. 54).

Bücher verlegt werden⁹, die flächendeckende Versorgung der Bevölkerung mit Büchern durch Buchhandlungen gesichert ist und dass sowohl kleine als auch große Verlage existieren können, die vielen Autoren das Publizieren ermöglichen (vgl. Weuster 2007, S. 62 f.).¹⁰

Eine weitere Unterstützung des Buches als Kulturgut bietet der Staat, indem für Bücher der reduzierte Umsatzsteuersatz¹¹ und reduzierte Kosten für den Postversand von Büchern gelten¹² (vgl. Bramann/Cremer 2014, S. 94).

Die Doppelrolle des Buches als Wirtschafts- und als Kulturgut muss in der Arbeit der Verlage berücksichtigt werden. Ein Verlag muss einerseits wirtschaftlich arbeiten, damit seine Existenz gesichert ist (vgl. Bramann 2014, S. 16). Damit ist er an marktwirtschaftliche Gesetzmäßigkeiten gebunden (vgl. Bramann 2014, S. 16). Andererseits produziert er oft Bücher, mit denen er wahrscheinlich keinen Gewinn erwirtschaften wird, um seiner Rolle als Kulturvermittler gerecht zu werden (vgl. Bramann 2014, S. 16).

Zusammenfassend lässt sich sagen, dass die Rollen des Buches als Kulturgut und als Wirtschaftsgut untrennbar sind. Dies müssen Unternehmen der Buchbranche, vor allem Verlage, beachten. Als Kulturgut stehen dem Buch einige Schutzmaßnahmen des Staates zu, in Form der Buchpreisbindung und des ermäßigten Umsatzsteuersatzes. Dadurch schützt der Staat das Buch vor gewissen Marktgesetzmäßigkeiten.

3.1.2 Marketing im Buchmarkt

An dieser Stelle soll auf das Marketing als einen spezifischen Teil der Wertschöpfungsstufe im Buchmarkt, der hauptsächlich im Verlag verortet ist, eingegangen werden. Dies ist relevant, weil, wie in Kapitel 2 bereits dargelegt, Big Data besonders für das Marketing gut einsetzbar ist. Daher werden an dieser Stelle der Arbeit die Funktionen und Funktionsweisen von Marketing in der Buchbranche erklärt, da die Analyse von vielfältigen Daten, wie Kundendaten, vor allem Einfluss auf das Marketing hat.¹³ Dies dient dem Zweck, später feststellen zu können, inwieweit Big Data das Marketing in der Branche verändern könnte.

Beim Marketing geht es zum einen um die Analyse der Marktchancen von Produkten und zum anderen um die Förderung des Verkaufs eines schon fertigen

9 Das bedeutet, dass das Veröffentlichen von insbesondere kulturell wertvollen Büchern auch bei kleiner Auflage gefördert werden soll (vgl. Weuster 2007, S. 62).

10 Der Zweck laut Buchpreisbindungsgesetz lautet: »§ 1 BuchPrG: Das Gesetz dient dem Schutz des Kulturgutes Buch. Die Festsetzung verbindlicher Preise beim Verkauf an Letztabnehmer sichert den Erhalt eines breiten Buchangebots. Das Gesetz gewährleistet zugleich, dass dieses Angebot für eine breite Öffentlichkeit zugänglich ist, indem es die Existenz einer großen Zahl von Verkaufsstellen fördert.« (Bundesministerium für Justiz und Verbraucherschutz 2016)

11 Für Bücher gilt der ermäßigte Steuersatz von 7 Prozent statt 19 Prozent; damit haben Bücher steuerlich gesehen den gleichen Stellenwert wie etwa viele Lebensmittel (vgl. Titel 2007, S. 217). Dies soll die Nachfrage stärken und die Leserschaft erhöhen (vgl. Janello 2010, S. 69).

12 Der Versand von Büchern mit einem Gewicht von bis zu 1000 Gramm gilt als Büchersendung und kostet weniger als ein herkömmlicher Versand (vgl. Bramann/Cremer 2014, S. 94).

13 Wie bereits in Kapitel 2.2.5 ab Seite 14 erläutert.

Produktes (vgl. Breyer-Mayländer 2014, S. 128). Dafür müssen Kundenbedürfnisse erfasst und erfüllt werden oder auch spezifische Bedürfnisse im Kunden geweckt werden, um ihn zum Kauf zu animieren und am Ende den Umsatz zu steigern

Das Marketing ist ein wesentlicher Bestandteil des Buchhandels und gehört in der Regel zu den Kompetenzen und Aufgaben eines Verlages. Im Verlagswesen wird überwiegend eine schon produzierte Auflage vermarktet, das bedeutet, dass für einen nicht näher bestimmten Markt produziert wird, bei dem von einer grundsätzlichen Nachfrage ausgegangen wird (vgl. Rautenberg/Wetzel 2001, S. 96). Meist wird der Markt dann mit einer Startauflage geprüft und gegebenenfalls bei hoher Nachfrage mit weiteren Auflagen reagiert (vgl. Rautenberg/Wetzel 2001, S. 96). Die Preisgestaltung ist dabei sehr homogen, die Preise werden selten der Marktnachfrage angepasst (vgl. Keuschnigg 2012, S. 56). Dies liegt daran, dass der Preis überwiegend anhand der Produktionskosten des Buches, also des Seitenumfangs und der Bindungsqualität, festgelegt wird (vgl. Keuschnigg 2012, S. 56). Bücher von Bestseller-Autoren kosten daher oft etwa genau so viel wie Bücher von unbekanntem Autoren (vgl. Keuschnigg 2012, S. 56). Auch die Buchpreisbindung und die Existenz von Preisschwellen¹⁴ tragen dazu bei, dass keine großen Spielräume bei der Preisgestaltung vorhanden sind (vgl. Rautenberg/Wetzel 2001, S. 97). Wegen meist geringer Auflagen sind die Marketing-Budgets oft nicht sehr hoch, sodass damit vor allem Kataloge zu den Neuerscheinungen und Werbematerialien in Buchhandlungen davon bezahlt werden (vgl. Rautenberg/Wetzel 2001, S. 97).

Ein Bestandteil des Marketings ist die Marktorientierung. Eine Marktorientierung ist zunehmend von Bedeutung, wenn es auf dem Markt viele Konkurrenzprodukte gibt, sodass die Käufer zwischen vergleichbaren Alternativangeboten auswählen können (vgl. Breyer-Mayländer 2014, S. 128 f.). Wegen der großen Zahl an Neuerscheinungen auf dem Buchmarkt – Breyer-Mayländer (2014, S. 129) spricht von ca. 82.000 Erstauflagen pro Jahr – ist diese Konkurrenz durchaus vorhanden. Verlage müssen sich daher an den Kunden orientieren, um möglichst viel über sie zu erfahren, wie beispielsweise Freizeitbeschäftigungen und die vorhandene Kaufkraft (vgl. Breyer-Mayländer 2014, S. 129). Schon 2001 erkannten Rautenberg/Wetzel (2001, S. 97), dass sich Verlage daher zunehmend an den Wünschen der Käufer orientierten und Titel nach Trends konzipierten.

Um Daten zu erhalten, war bisher die Marktforschung für Verlage die hauptsächliche Datenquelle (vgl. Bramann/Cremer 2014, S. 383 f.). Marktforschungsinstitute liefern die reinen Abverkaufdaten eines Buches aus dem verbreitenden Buchhandel und erstellen Bestsellerlisten (vgl. Roesler-Graichen 2015, S. 11; vgl. Hauck 2016, S. 16). Weitere Daten stammen beispielsweise aus Befragungen, Beobachtungen und Panels (vgl. Breyer-Mayländer 2014, S. 136–140). Die wenigsten Verlage allerdings können eigene Marktstudien finanzieren; dies können sich nur sehr

¹⁴ Preisschwellen sind Schwellen, ab denen ein Kunde ein Produkt als zu teuer ablehnt; im Buchmarkt soll die Schwelle bei 10 Euro für ein Taschenbuch und 20 Euro für ein Hardcover-Buch liegen (vgl. Dieckmann 2015).

große Verlage oder der *Börsenverein des Deutschen Buchhandels* als Dachverband leisten (vgl. Bramann/Cremer 2014, S. 383 f.). Der Börsenverein gibt beispielsweise jährlich *Buch und Buchhandel in Zahlen* heraus, in dem relevante Studien und Quellen ausgewertet werden (vgl. Bramann/Cremer 2014, S. 384). Daher stammen die Daten für kleine und mittelgroße Verlage meist aus Sekundärquellen. Konzerne, die Buchhandelsketten oder einen Buchversand unter ihrem Dach vereint haben, haben zudem oft statistische Verkaufsdaten zur Verfügung, die Trends schneller erkennbar machen können (vgl. Rautenberg/Wetzel 2001, S. 97). Im literarischen Bereich lässt sich ein Bestseller dennoch nur schwer vorhersagen, da dies von schwer kalkulierbaren Faktoren abhängt; hier ist laut Rautenberg/Wetzel (2001, S. 97) »das intuitive Einschätzungsvermögen und Erfahrungswissen der Lektoren« wichtig. Hat sich ein Autor zum Bestseller-Autor entwickelt, kann er dann allerdings wie eine Produktmarke gepflegt werden; die Bücher des Autors erzielen in dem Fall genug Gewinn, um auch kostenintensivere Marketingkampagnen durchzuführen (vgl. Rautenberg/Wetzel 2001, S. 98).

Das Marketing, wie es derzeit auf dem Buchmarkt praktiziert wird, weist einige Probleme auf. Wie bereits erwähnt, ist das Budget für Marketing in Verlagen meist recht klein. Daher ist es in der Regel nicht möglich, alle Neuerscheinungen mit Marketingkampagnen zu bewerben. Das Marketing konzentriert sich also meist auf diejenigen Bücher, bei denen der Verlag davon ausgeht, dass sie Bestseller werden könnten, oder die von Bestseller-Autoren stammen. Wie aber von Rautenberg und Wetzel beschrieben, sind Bestseller besonders im literarischen Bereich schwer zu kalkulieren. Dies wiederum liegt unter anderem daran, dass den Verlagen bisher recht wenige Daten zu den Lesern zur Verfügung standen. Die hauptsächliche Datenquelle sind die Verkaufszahlen eines Buches. Verkaufszahlen können zwar abbilden, welche Bücher sich gut verkauft haben und sogar zu Bestsellern geworden sind, aber hieraus lassen sich wenig Informationen darüber gewinnen, warum sich diese Bücher gut verkauft haben und wie die Leserschaft dieser Bücher gestaltet ist. Dies sollen Befragungen abdecken, welche recht aufwendig sind und von den Verlagen nicht selbst gesteuert werden können, da sie diese Daten in der Regel von den Marktforschungsinstituten kaufen müssen und die Daten somit aus sekundären Quellen stammen. So lassen sich zwar einige Erkenntnisse zu grundsätzlichen Interessen und Gewohnheiten von Lesern gewinnen, aber nicht dazu, wie Leser spezielle Bücher lesen oder wie die Zielgruppe für ein spezifisches Buch aussehen könnte. Da die Leserorientierung, wie oben beschrieben, bei der Programmgestaltung und beim Marketing zunehmend bedeutender wird, ist dies problematisch, da die bisher zugänglichen Daten zu den Leserbedürfnissen und Lesegehnheiten nicht ausreichend detailliert sind. Dieses Problem wird dadurch verstärkt, dass sich das Marketing meist auf potenzielle Bestseller konzentriert. Aus den Verkaufsdaten lässt sich allein ablesen, ob ein Buch ein Bestseller ist oder nicht. Bei Befragungen geht es meist darum, die Meinungen und Wünsche der Mehrheit zu erhalten. Daten zu

Nischenlektüre sind daher besonders unzureichend. Es sind also neue Datenquellen nötig, um die Leserwünsche breitflächiger zu erfahren und das Marketing dahingehend gezielter anpassen zu können. Diese Möglichkeit bietet Big Data, da hier meist verschiedene Datenquellen verknüpft werden, um umfassendere Informationen zu erhalten. Eine Quelle wären beispielsweise soziale Netzwerke. Eine weitere Möglichkeit speziell in der Buchbranche sind die Reader Analytics, auf die im Kapitel 3.2.3 näher eingegangen wird.

3.2 Anwendung von Big Data in der Buchbranche

Um herauszufinden, welches Potenzial Big Data für die Buchbranche hat, ist es wichtig, den Status quo zu definieren. Daher soll im folgenden Kapitel einerseits beleuchtet werden, welche Daten in der Buchbranche schon wie verwendet werden. Andererseits soll hier auch dargestellt werden, ob und wie sich die Buchbranche bisher zum Thema Big Data positioniert hat. Zuletzt soll vertiefend auf die Reader Analytics, also die Leserdaten-Analyse, eingegangen werden.

3.2.1 Anwendung von Daten in der Buchbranche

Wie im vorherigen Kapitel bereits dargelegt, werden im Buchmarkt bestimmte Daten bereits verwendet, beispielsweise um Verkaufszahlen auswerten zu können. Um den Status quo in der Buchbranche umfassender darzustellen, ist es daher wichtig, zu erläutern, welche Daten in der Praxis schon verwendet werden.

Ein Datenkonzept, das häufig auch in der Buchbranche, vor allem in Verlagen, verwendet wird, ist das Customer-Relationship-Management (CRM) zur Verwaltung von Kundendaten. Laut Leußner et al. (2011, S. 18) ist CRM eine kundenorientierte Unternehmensstrategie, welche mit Hilfe von integrierten CRM-Informationssystemen eine ganzheitliche Sicht auf den Kunden bietet und somit eine gezielte Kundenansprache ermöglicht. Die Strategie dahinter ist, dass eine gute Kundenbeziehung aufgebaut wird, damit die Kundenloyalität und letztlich Kundenbindung gestärkt wird (vgl. Homburg/Bruhn 2013, S. 9). Diese ist ein relevanter Faktor dafür, dass ein Kunde ein Produkt wieder kauft (vgl. Kreutzer 2016, S. 159). Profitable Kunden sollen an das Unternehmen gebunden werden, wofür ein umfangreiches Wissen über die Struktur, das Verhalten und die Bedürfnisse der Kunden nötig ist, welches mit CRM aggregiert werden kann (vgl. Hippner/Wilde 2013, S. 181).

Für die Durchführung von CRM kommen meist spezielle CRM-Systeme zum Einsatz, deren Aufgaben die Synchronisation und Unterstützung von Marketing, Vertrieb und Service, die Einbindung aller Kommunikationskanäle zwischen Kunden und Unternehmen sowie die Zusammenführung und Auswertung aller Kundeninformationen sind (vgl. Rentzmann et al. 2011, S. 131). Dafür werden beispielsweise alle Kundenkontakte und Kundenreaktionen in den Bereichen Marketing, Vertrieb und Service systematisch aufgezeichnet (vgl. Hippner/Wilde 2013, S. 182). Dadurch

entsteht eine einheitliche Sicht auf den Kunden und die Mitarbeiter können auf alle relevanten Kundeninformationen zurückgreifen (vgl. Hippner/Wilde 2013, S. 182).

CRM kann auch für Big-Data-Konzepte eingesetzt werden. So kann insbesondere das Data Mining Bestandteil einer CRM-Strategie sein. Dabei werden die im CRM-System gespeicherten Daten nach dem Prinzip des Data Mining analysiert; die dadurch gewonnenen Erkenntnisse können dann wiederum Ausgangspunkt für weitere Aktionen oder Optimierungen sein (vgl. Hippner/Wilde 2013, S. 187). Data Mining kann damit in verschiedenen Phasen des Kundenbeziehungslebenszyklus¹⁵ die phasenspezifischen Aufgaben unterstützen und so am Ende zur Optimierung der Kundenbeziehung beitragen (vgl. Rentzmann et al. 2011, S. 136; vgl. Hippner et al. 2011, S. 799). So können beispielsweise die Zielgruppen bei Akquise-, Kundenbindungs-, oder Kundenrückgewinnungs-Kampagnen gezielt angesprochen werden (vgl. Hippner et al. 2011, S. 801–804). Desweiteren bieten CRM-Systeme, die prinzipiell strukturierte Kundendaten aus internen Quellen enthalten, die Möglichkeit, diese mit unstrukturierten Daten aus externen Quellen, wie beispielsweise Daten aus sozialen Netzwerken, anzureichern; ein Merkmal von Big Data ist ja, dass sich die Datenquellen vervielfältigt haben, was dem CRM zugute kommt und sich in CRM-Systemen gut umsetzen lässt (vgl. Verhoef et al. 2016, S. 131).

Insbesondere in Verlagen gewinnt das Thema CRM an Bedeutung (vgl. Spiller 2015, S. 12). Bereits im Jahr 2005 berichtet Deininger, dass viele Verlage in CRM investieren und beispielsweise der Marktführer von CRM-Software für Verlage 350 Kunden zählt (vgl. Deininger 2005). Auch hier ist das Wissen über die Kundenbedürfnisse und eine gute Kommunikation mit den Kunden elementar (vgl. Spiller 2015, S. 12). CRM-Zielsetzungen von Verlagen können beispielsweise die Früherkennung von Trends und eine höhere Kundenzufriedenheit sowie besseres Erfassen und Austauschen von Daten und die Verbesserung der Datenqualität durch Aufbau einer umfassenden Kundendatenbank sein (vgl. Spiller 2015, S. 12 f.). Besonders in Marketing und Vertrieb eines Verlags kann CRM einige Erkenntnisse bieten, wie beispielsweise den Ausbau von Informationen über die Kunden und ein besseres Verständnis für den Markt (vgl. Spiller 2015, S. 17). Mit CRM können Verlage beispielsweise gezielt Kunden, die in einer speziellen Branche arbeiten, und Unternehmen, die ein spezielles Produkt anbieten, ein Abonnement für eine entsprechende Fachzeitschrift oder bestimmte Fachbücher anbieten. CRM unterstützt damit insbesondere den Direktverkauf aus dem Verlag (vgl. Kappes 2016). Der Direktvertrieb der Verlage hat in den letzten Jahren dementsprechend kontinuierlich zugenommen (vgl. Weise 2016, S. 8).

Die Buchbranche arbeitet darüber hinaus schon länger mit Marktdaten, die sie aus Marktforschungen erhält.¹⁶ So sind die Abverkaufdaten eines Buches, die über die Scannerkassen in den Buchhandlungen gespeichert werden, sowohl für die Buch-

¹⁵ Der Kundenbeziehungslebenszyklus beinhaltet potenzielle Kunden, aktive Kunden, gefährdete Kunden, verlorene Kunden und reaktivierte Kunden (vgl. Hippner et al. 2011, S. 800).

¹⁶ Darauf wurde in Kapitel 3.1.2 ab Seite 27 bereits kurz eingegangen.

handlungen als auch für die Verlage wichtig (vgl. Roesler-Graichen 2015, S. 11). Marktforschungsinstitute wie Media Control und GfK Entertainment sammeln diese Daten bei verschiedenen Partnerbuchhandlungen, liefern diese an die Verlage und erstellen Bestsellerlisten (vgl. Roesler-Graichen 2015, S. 11; vgl. Hauck 2016, S. 16). Weiterhin werden Lagerbestandsdaten bei Zwischenbuchhändlern, wie Zugänge, Abgänge und Lagerabflussgeschwindigkeiten, erfasst, aus denen sich wiederum Verkaufszahlen errechnen lassen (vgl. Roesler-Graichen 2015, S. 13). Solche Daten können die Verlage mit eigenen Daten ergänzen und dann Trendanalysen erstellen (vgl. Roesler-Graichen 2015, S. 13). Darüber hinaus bietet Media Control nach eigenen Angaben im Bereich Buch beispielsweise auch Daten aus sozialen Netzwerken und Konsumentenpanels an (vgl. media control GmbH 2015). Diese Daten stellen, wie zuvor bereits erwähnt, momentan noch die Hauptquelle an Daten für Marketingzwecke in Verlagen dar.

Ein weiteres essentielles Datenmanagement-Werkzeug bei Büchern sind die Metadaten. Das VLB¹⁷, das Verzeichnis lieferbarer Bücher, wurde in eine große Metadatenbank umgewandelt, in dem sich Ende 2015 bereits zwei Millionen Titel von mehr als 21.000 Verlagen befanden (vgl. Heimann 2015). Die reinen Titeldaten werden dabei mit verschiedenen Daten wie Schlagworten, Coverbildern und Zusatztexten angereichert. Laut einer Broschüre des MVB zum neuen VLB+ können solche Metadaten, wenn sie von hoher Qualität sind, dazu beitragen, den Verkauf eines Titels teilweise deutlich zu fördern (vgl. MVB o. J., S. 4–6). Auf die Metadaten sollen sowohl Buchhändler, Bibliotheken und Presse als auch Endkunden Zugriff haben (vgl. MVB o. J., S. 16). Aufgrund des hohen Datenvolumens der Metadatenbank wird diese von Ronald Schild, dem MVB-Geschäftsführer, als Big Data bezeichnet (vgl. Heimann 2015). Ob dies zutreffend ist, wird im nächsten Schritt dargelegt.

Es hat sich gezeigt, dass bereits einige Daten, die zu Big Data gezählt werden können, in der Buchbranche verwendet werden. Insbesondere CRM-Daten können zu Big Data gezählt werden¹⁸; besonders, wenn die Daten hierfür aus mehreren, verschiedenen Quellen stammen. Hier haben Verlage also bereits Berührung mit dem Thema Big Data. Leider lassen sich keine genauen Angaben dazu finden, wie viele Verlage CRM bereits verwenden und in welchem Umfang. Vermutlich werden aber zumindest die größeren Verlage dementsprechende Software einsetzen.¹⁹ Bei CRM-Daten handelt es sich um Kundendaten der Angebotsseite.

Um an Marktdaten zu gelangen, verlässt sich die Buchbranche bisher überwiegend auf Daten von Marktforschungsinstituten. Dabei handelt es sich um strukturierte Daten der Angebotsseite wie Abverkaufszahlen und Daten zu Lagerbeständen. Da

17 Eine Datenbank des deutschen Buchhandels mit Titeldaten zu allen lieferbaren Büchern im deutschsprachigen Raum, herausgegeben vom MVB des Börsenvereins des Deutschen Buchhandels (vgl. MVB 2016, S. 3ff.)

18 wie bereits in Kapitel 2.2.2 ab Seite 7 erwähnt.

19 Thomas Krimmer vom Thieme Verlag berichtet beispielsweise in einem Beitrag für das Buchreport Magazin vom Einsatz von CRM in seinem Verlag (vgl. Krimmer 2015, S. 16).

diese Daten aufgrund ihrer Strukturiertheit leicht verarbeitet und analysiert werden können, handelt es sich hierbei nicht um Big Data. Allerdings bietet Media Control auch Marktdaten der Nachfrageseite, also detailliertere Informationen zu den Käufern aus sozialen Netzwerken und Konsumentenpanels, an, wobei es sich um viele, unstrukturierte Daten aus verschiedenen Datenquellen und damit um Big Data handelt.

Ein spezieller Fall ist die für die gesamte Branche gedachte Metadatenbank, die den Verlagen ein Werkzeug bieten möchte, Daten von hoher Qualität bereitzustellen, um einen verkaufsfördernden Effekt zu erhalten. Die Daten der Metadatenbank können als Big Data bezeichnet werden, weil sie ein hohes Datenvolumen haben, der Wert der Daten eindeutig ist, nämlich Verkaufsförderung von Titeln, und die Daten sowohl strukturiert – die reinen Metadaten wie Titeldaten und Autorennamen – als auch unstrukturiert – zusätzliche Fließtexte, beispielsweise bei Inhaltsbeschreibungen – sein können. Damit sind drei wichtige Vs – nämlich *volume*, *value* und *variety* – erfüllt. Die Qualität der Daten allerdings ist recht unterschiedlich, je nachdem, wie ausführlich die Verlage ihre Titel in der Datenbank mit Metadaten bestücken. Zudem findet hier keine Datenanalyse oder gar Echtzeitanalyse statt, sondern es ist nur eine Datenbank, welche hauptsächlich den Zweck hat, Daten zu speichern.

3.2.2 Diskussion in der Buchbranchenpresse

Zuletzt soll an dieser Stelle auf den Stand der Diskussion in der Buchbranche eingegangen werden. Die Analyse der Buchbranchenpresse im Bezug darauf, inwieweit hier bereits das Phänomen Big Data diskutiert wird und wie es bewertet wird, lässt Rückschlüsse darauf zu, wie offen der Markt bereits für die neuen Technologien sein könnte und wie die Branche das Potenzial bewertet.

Um herauszufinden, wie Big Data in der Buchbranchenpresse diskutiert wurde und diskutiert wird, wurden die zwei bedeutsamsten Organe der Buchbranche, das *Börsenblatt* sowie das *Buchreport Magazin*, jeweils von 2012 bis 2016 analysiert. Das Jahr 2012 wurde als Startjahr gewählt, weil es als das Jahr des Durchbruchs von Big Data gilt, und daher davon ausgegangen werden kann, dass das Thema in der Buchbranche vorher nicht aufgegriffen wurde.²⁰ Zudem soll hier nur die zeitlich aktuelle Diskussion untersucht werden, weswegen der Analysezeitraum die vergangenen fünf Jahre umfasst. Zusätzlich wurde die aktuelle Branchendiskussion auf den jeweiligen Websites von *Börsenblatt* und *Buchreport* sowie den dazugehörigen Blogs – *bookbytes* vom *Börsenblatt* und *pubiz* vom *Buchreport* – verfolgt, teilweise mit Hilfe von Newslettern. So konnte ein Eindruck von der Diskussion von Big Data in der Buch- und Publishingbranche im deutschsprachigen Raum gewonnen werden.

Die untersuchten Artikel werden danach eingeteilt, ob sie überwiegend negativ oder positiv über Big Data berichten oder ob es Artikel mit abwägender Bewertung sind oder reine Berichte ohne Bewertung.

²⁰ wie bereits in Kapitel 2.2.1 ab Seite 6 erwähnt.

Insgesamt gab es meist Artikel mit positiver Bewertung, einige neutrale Berichte und relativ wenig Texte mit negativer Bewertung (s. auch Tabelle 1 auf Seite 37). Einen kritischen Unterton hatten zwei kleine Notizen im *Börsenblatt*, nämlich die Meldung, dass Amazon, Barnes & Nobles, Google und Kobo deutlich mehr Daten aus E-Books sammeln als das Software-Unternehmen Adobe (vgl. o. V. 2012, S. 10) oder im »Aufreger der Woche«, in dem von einem Programm berichtet wird, das Daten aus E-Lehrbüchern sammelt (vgl. o. V. 2013, S. 6). Im Jahr 2016 druckte das *Buchreport Magazin* ein Interview mit dem dänischen Markenexperten Martin Lindstrom, der sich kritisch zu Big Data äußert: Er plädiert für Small Data, was aus »Beobachtungen, Erkenntnissen und Hinweisen, die zunächst unbedeutend erscheinen« (Lindstrom 2016, S. 33) besteht, was für ihn bedeutet, dass das Marketing sich nicht auf Datenanalysen, sondern wieder mehr auf Verbraucherbeobachtungen verlassen soll. Er kritisiert zudem, dass mit Big Data nur Korrelationen aufgedeckt werden können, aber keine Kausalitäten (vgl. Lindstrom 2016, S. 32).

In allen untersuchten Publikationsformen wird überwiegend in neutraler Art und Weise über neue Entwicklungen im Bereich Daten oder über neue Technologien wie die Reader Analytics berichtet. Im *Börsenblatt* erschien im Jahr 2013 ein Artikel zum Thema Big Data, in dem das Phänomen Big Data kurz beschrieben wird und Buchtipps zum Thema präsentiert werden, allerdings wird hier nicht Bezug auf den Einsatz von Big Data in der Buchbranche genommen (vgl. Schmidt 2013, S. 32 f.). Im Jahr 2016 findet sich im *Börsenblatt* außerdem ein Hinweis auf die Veranstaltung *MVB Data Summit* am 21. November 2016, bei dem es um Datenmanagement für Verlage und Buchhandlungen geht (vgl. o. V. 2016c, S. 14). Auf *boersenblatt.net* wird im Oktober 2013 das erste Mal von einer Experten-Diskussionsrunde auf der Frankfurter Buchmesse zum Thema *Wissen oder Algorithmus: Woher kommt das Know-how der Zukunft?* berichtet, bei der das Potenzial von Big Data diskutiert wurde (vgl. Busse 2013). Weiterhin berichtet Selle (2016) bei *bookbytes* über eine Reform der EU-Datenschutzrichtlinien, welche die Nutzung von Big Data einschränkt. Im Jahr 2015 erläutert Scherer erstmals ausführlich im *Buchreport Magazin* die Methode der Reader Analytics; der Artikel beinhaltet außerdem ein Interview mit dem Gründer des Reader-Analytics-Startups Jellybooks Andrew Rhomberg (vgl. Scherer 2015, S. 16–19). Ende des Jahres 2015 griff Steffen Meier im Blog *bookbytes* das Thema Reader Analytics auf und versuchte zu sondieren, ob es sich hierbei nur um einen Hype handelt, oder ob wirklich relevante Erkenntnisse gewonnen werden können (vgl. Meier 2015a, 2015b).

Besonders im *Buchreport Magazin* sowie auf *buchreport.de* und dem zu Buchreport gehörenden Blog *pubix* wird häufig positiv über Big Data berichtet. Bereits im Jahr 2012 erschien im *Buchreport Magazin* der erste Artikel *Wie Verlage mit Daten bessere Bücher machen*, der darlegt, was Big Data ist, und welche Werkzeuge für die Datenanalyse es bereits gibt, die in der Buchbranche angewandt werden können, und berichtet vom US-amerikanischen Verlag Sourcebooks, der Big Data bereits vorbild-

lich einsetzt (vgl. Lenz 2012, S. 98–101). 2014 veröffentlichte das Buchreport Magazin ein Interview mit dem Datenanalyse-Experten Thomas Davenport; darin beschreibt Davenport, dass Big Data auch für die Buchbranche großes Potenzial hat, die Entscheidungsfindung innerhalb des Unternehmens zu verbessern sowie neue Produkte und Serviceangebote zu entwickeln (vgl. Davenport 2014c, S. 12). Hier wird im Bezug auf Big Data für Verlage vor allem auf die Reader Analytics verwiesen (vgl. Davenport 2014c, S. 12 f.). Kleine Verlage, die keine Mittel für solche Maßnahmen haben, könnten laut Davenport (2014c, S. 13) zumindest die Stimmung der Leser in sozialen Netzwerken beobachten, was aber selten gemacht werde. Im Jahr 2015 erschien im *Buchreport Magazin* ein Artikel unter der Rubrik *Meinung*, der ebenfalls die Chancen für die Buchbranche, welche die Reader Analytics – also die Auswertung von Leserdaten – bieten, thematisiert und positiv bewertet (vgl. Lenz 2015, S. 8).

Parallel zu Artikeln im *Buchreport Magazin* finden sich auch auf *buchreport.de* viele Artikel und Interviews zum Thema Big Data in der Buchbranche. Die Suche für das Stichwort »Big Data« in Online-Artikeln des Buchreports bietet ca. 80 Ergebnisse bis zurück ins Jahr 2012. Daher seien hier nur exemplarisch einige erwähnt. Bereits 2013 wurde beispielsweise ein Interview mit dem E-Book-Experten Sebastian Posth geführt, in dem es um die Perspektiven des Einsatzes von Daten in der Buchbranche geht (vgl. Posth 2013). 2015 berichtet *buchreport.de* von einer Studie, nach der nur wenige Medienhäuser und Verlage bereits Daten einsetzen und Big-Data-Tools kaum genutzt werden (vgl. o. V. 2015). Im Jahr 2014 erschien ein Interview mit dem Big-Data-Experten Viktor Mayer-Schönberger, in dem er den Verlagen empfiehlt, vermehrt auf Basis von Daten Entscheidungen zu treffen, um nicht von Unternehmen wie Amazon, die durch Datenanalyse viel mehr über ihre Kunden wissen, verdrängt zu werden (vgl. Mayer-Schönberger 2014). Auf *pubiz.de* betont Heiko Beier, dass in der Buchbranche Smart-Data-Konzepte eingesetzt werden sollten und dass diese gerade für Marketing und Kundenorientierung wichtig sein können (vgl. Beier 2016a, 2016b).

Auch im Blog *bookbytes* vom *Börsenblatt* und auf *boersenblatt.net* gibt es einige Artikel mit positiver Bewertung. Im Jahr 2016 veröffentlichte *boersenblatt.net* einige Artikel im Bezug auf den *MVB Data Summit*, wie beispielsweise ein Interview mit Mark Hoenke, einem Spezialisten für CRM-Systeme, in dem er Big Data große Chancen in der Buchbranche attestiert (vgl. Hoenke 2016), sowie einen Nachbericht des *MVB Data Summit* (vgl. o. V. 2016h). Meier veröffentlichte 2016 bei *bookbytes* einen Artikel über das sogenannte Data Driven Publishing, in dem er die Meinung vertritt, dass es für Verlage nicht verwerflich ist, Datenanalysen für die Programmplanung zu verwenden (vgl. Meier 2016).

Zusammenfassend lässt sich beobachten, dass das *Börsenblatt* bisher insgesamt sehr wenig über Big Data berichtet hat. In der Printausgabe des *Börsenblatts* ist im untersuchten Zeitraum kein einziger großer Artikel erschienen, der sich mit Big Data in der Buchbranche auseinandersetzt. Eine Diskussion fand hier also – zumindest in

der Printzeitschrift – bisher nicht statt. Hier wurde Big Data anscheinend noch nicht als relevant für die Buchbranche klassifiziert. Ausführlichere Artikel und Meinungen wurden meist über den Blog *bookbytes* veröffentlicht (s. auch Tabelle 1 auf Seite 37). Beim *Buchreport Magazin* hingegen sind schon früh erste Artikel zu Big Data erschienen und es wird immer wieder in Artikeln hinterfragt, wie groß das Potenzial von Big Data für die Buchbranche ist, und es werden Interviews mit Experten zu dem Thema geführt. Beide Branchenorgane berichten insgesamt überwiegend positiv über Big Data, das Potenzial für die Buchbranche wird meist als gut eingeschätzt, auch im Hinblick auf die immerwährende Konkurrenz der Branche mit Amazon (s. Tabelle 1 auf Seite 37).

Es ist weiterhin deutlich geworden, dass bei Anwendungsbeispielen von Big Data in der Buchbranche, vor allem in Verlagen, fast immer die Reader Analytics genannt wurden. Das Potenzial dieser Reader Analytics wurde zudem in der Branchenpresse überwiegend positiv bewertet (vgl. z. B. Lenz 2015, S. 8). Diese Art von Daten scheinen bisher das vielversprechendste Big-Data-Konzept für die Buchbranche zu sein, da mit ihnen ein Data Driven Publishing möglich ist. In diesem Bereich haben sich zudem bereits einige Unternehmen gegründet, die sich auf Reader Analytics spezialisiert haben. Daher sollen im nächsten Schritt die Reader Analytics als Fallbeispiel für teilweise schon verwendete Big-Data-Konzepte mit großem Potenzial für die Buchbranche ausführlich erläutert werden.

	positive Bewertung	negative Bewertung	abwägende/ neutrale Bewertung oder Berichte ohne Bewertung
Börsenblatt		<ul style="list-style-type: none"> • Datenkrake E-Reader (o. V. 2012) • Digitaler Spion (o. V. 2013) 	<ul style="list-style-type: none"> • Erster MVB Data Summit (o. V. 2016 c) • ohne Ende Daten – aber was bedeuten sie? (Schmidt 2013)
börsenblatt.net, bookbytes	<ul style="list-style-type: none"> • Wissen macht was (o. V. 2016h) • Do Androids Dream of Electric Literature? (Meier 2016) • »Big Data bietet enorme Chancen« (Hoenke 2016) 		<ul style="list-style-type: none"> • Big Data: Überrollt die riesige Datenmenge die Fachverlage? (Busse 2013) • »Big Data« werden Grenzen gesetzt (Selle 2016)
Buchreport Magazin	<ul style="list-style-type: none"> • Wie Verlage mit Daten bessere Bücher machen (Lenz 2012) • »Daten sind effektiver als die Geschäftsessen-Kultur« (Davenport 2014c) • Mehr als Kribbeln im Bauch (Lenz 2015) 	<ul style="list-style-type: none"> • »Big Data ist ein Ausdruck von Unsicherheit« (Lindstrom 2016) 	<ul style="list-style-type: none"> • Reader Analytics: So tickt das Publikum (Scherer 2015)
buchreport.de, pubiz.de	<ul style="list-style-type: none"> • Mit Daten statt Intuition entscheiden (Posth 2013) • Angst vor Big Data ist kein gutes Konzept (Mayer-Schönberger 2014) • Smart ist smarter als Big (Meier 2016a) • Mit Smart Data gezielter planen, publizieren, werben und kommunizieren (Beier 2016b) 		<ul style="list-style-type: none"> • Heiliger Gral oder Schierlingsbecher? / Am Anfang war die Datenbasis (Meier 2015a,b)

Tabelle 1: Die untersuchten Branchenpressenartikel, sortiert nach Bewertung

3.2.3 Reader Analytics

Reader Analytics, also die Möglichkeit, Leserdaten zu sammeln und auszuwerten, sind erst durch den Markteinstieg von E-Books als digitale Produkte möglich geworden. Neben CRM-Daten können vor allem diese Leserdaten für die Buchbranche, beispielsweise für das Marketing, einen Zugewinn bedeuten, da bisher Marktdaten hauptsächlich aus den Verkaufszahlen gewonnen wurden. Sie stellen also eine neue Datenquelle sowie eine vollkommen neue Datenart dar. Wie bereits ausgeführt²¹, ist der Zugang zu neuen Datenquellen eine Komponente und eine Chance des Einsatzes von Big Data. In der Buchbranchenpresse wird außerdem überwiegend über die Reader Analytics berichtet, wenn es um das Thema Big Data in der Buchbranche geht. Aus diesen Gründen werden sie hier als Fallbeispiel vertieft behandelt. Dazu soll zunächst dargestellt werden, was Reader Analytics sind, wie sie funktionieren und welche Daten und Erkenntnisse sie liefern. Im Anschluss daran werden fünf Unternehmen exemplarisch vorgestellt, die mit Reader Analytics arbeiten.

Reader Analytics sind Leserdaten, die entstehen, wenn ein Leser ein E-Book liest (vgl. Scherer 2015, S. 16). Diese Daten lassen sich relativ leicht aus E-Books, die im EPUB3-Format vorliegen, auslesen, da E-Books technisch gesehen wie Websites aufgebaut sind, welche sich ebenfalls gut tracken lassen (vgl. Gröger 2016). Reader Analytics können beispielsweise zeigen, ob ein E-Book überhaupt geöffnet und gelesen wurde, ob die Lektüre nach einigen Seiten abgebrochen wurde und wenn dies zutrifft, an welcher Stelle sie abgebrochen wurde (vgl. Scherer 2015, S. 16). Geöffnete Kapitel, Lektüreabbruch, Lesetempo, Leseпаusen, Lesezeiten und -rhythmus, Lese-dauer, Geräteinformationen und der geografische Zugriffsort sind Daten, die beispielsweise über E-Book-Apps auslesbar sind (vgl. Scherer 2015, S. 18). Laut Meier (2015b) handelt es sich bei Reader Analytics sowohl um quantitative Nutzungsdaten, als auch um qualitative Daten aus Fokusgruppen, beispielsweise durch Befragungen der Leser oder Social Reading²², welche die Nutzungsdaten anreichern.

An dieser Stelle muss anhand der Terminologie aus Kapitel 2.2 erläutert werden, warum es sich bei den Reader Analytics um Big Data handelt. Zunächst einmal stammen Reader Analytics aus einer Datenquelle, nämlich den E-Books; manchmal werden aber auch Daten aus weiteren Quellen (beispielsweise Leserbefragungen, Social Reading) hinzugefügt. Die Datenquelle ist in der Regel extern, da Verlage selbst keinen Zugang zu den Leserdaten haben, sondern sie nur von anderen Unternehmen erhalten können. Es handelt sich zum einen um Marktdaten der Nachfrageseite, da hierbei die Konsumenten im Markt beschrieben werden, und zum

21 s. Kapitel 2.2.5 ab Seite 14.

22 Social Reading ist ein »online geführter, intensiver und dauerhafter Austausch über Texte« (Pleimling 2012), der beispielsweise in Foren, Blogs oder sozialen Netzwerken stattfindet; auch die Dokumentation des individuellen Leserverhalten wird darunter gefasst, beispielsweise das Teilen von Notizen, Markierungen und Bewertungen (vgl. Pleimling 2012). Es werden hier also überwiegend unstrukturierte Daten aus den Diskussionen und Bewertungen der Leser gesammelt und ausgewertet.

anderen auch um Kundendaten der Nachfrageseite, weil mit den Leserdaten die Zufriedenheit der Kunden mit einem Buch analysierbar wird. Leserdaten beinhalten strukturierte Daten (beispielsweise die Anzahl der beendeten Bücher in Prozent) und unstrukturierte Daten (beispielsweise Anmerkungen der Leser). Die Daten sind außerdem vielfältig, da mehrere verschiedene Faktoren mit den Leserdaten analysiert werden können. Aufgrund der Vielfalt der Daten kann davon ausgegangen werden, dass das Datenvolumen recht groß ist. Da die Daten von allen Lesern zu einem Buch erstmal gesammelt und dann ausgewertet werden, findet hier keine Echtzeitanalyse statt, es ist aber denkbar, dass dies theoretisch auch möglich wäre. Die Qualität der Daten kann als gut bezeichnet werden, da die Daten recht eindeutig interpretiert werden können. Auch der Nutzen der Daten kann benannt werden, beispielsweise gezielteres Marketing für bestimmte Bücher. Die 5 Vs von Big Data werden also überwiegend erfüllt.

Zuletzt ist es wichtig, dass Leserdaten eindeutig Muster darlegen können, da dies ein Kriterium von Big Data ist. Ein vorher definierter Sachverhalt wäre hier beispielsweise die Frage, wie viele Leser das Buch zu Ende lesen und wenn sie es nicht zu Ende lesen, an welcher Stelle sie im Buch aussteigen. Die Analyse der Leserdaten legt dazu genaue Muster offen, in welchen Kapiteln wie viele Leser aussteigen und wie viele Leser das Buch beendet haben (s. hierzu auch Abbildung 1 auf Seite 42). Es handelt sich hierbei also um eine Datenanalyse der Kategorie Datenmodellierung. Zuletzt sollen Reader Analytics Daten zu Entscheidungsfindungen in Verlagen, beispielsweise im Marketing, liefern, was ein wichtiger Aspekt und ein wichtiges Ziel von Business Intelligence ist.

Die Chance der Reader Analytics liegt darin, dass Verkaufszahlen als alleinige Marktdaten der Nachfrageseite allein oft nicht genug Aussagekraft für den Erfolg eines Buches haben und daher als Basis für Marketingmaßnahmen nur bedingt taugen (vgl. Scherer 2015, S. 16). Dank des digitalen Lesens können mit Reader Analytics viel umfassendere Daten zum Leserverhalten gesammelt und damit Trends erkannt werden (vgl. Albazaz 2016a). Mit dem dadurch zugänglichen Wissen über die Leser kann sich die Art und Weise verändern, wie Verlage Bücher akquirieren, bearbeiten und vermarkten (vgl. Alter/Russell 2016). Die Daten darüber, wie viele Leser ein Buch bis zum Ende gelesen haben, und Empfehlungen, die in einer frühen Phase nach der Veröffentlichung gesammelt werden, können beispielsweise Indikatoren dafür liefern, ob das Buch sich gut verkaufen wird oder nicht (vgl. Kirkbride 2016). In einer späteren Phase können die Reader Analytics dem Marketing wichtige Daten darüber liefern, wie die Zielgruppe beschaffen ist, wann und wo sie liest und damit, wie sie angesprochen werden sollte (vgl. Kirkbride 2016). Solche Daten helfen insbesondere dabei, das Marketing-Budget effektiv einzusetzen (vgl. Kirkbride 2016). Es kann beispielsweise sein, dass der Bestseller eines Autors den meisten Lesern nicht gefällt und daher ein Folgebuch nur noch wenig Käufer finden würde, was aber der Verlag in jenem Fall, dass ihm nur Verkaufszahlen zur Verfügung ständen, nicht

weiß (vgl. Scherer 2015, S. 16). Oder ein eigentlich gut bewertetes Buch hat ein Cover, das nicht dem Inhalt entspricht oder insgesamt nicht ansprechend ist (vgl. Rhombert 2016d).

Zudem bieten Reader Analytics viele weitere chancenreiche Funktionen. So können Autoren mit Bestsellerpotenzial und erfolgreiche Inhalte für Publikationen entdeckt werden und das Marktpotenzial für Reihen kann bewertet werden (vgl. Scherer 2015, S. 18). Weiterhin können Stellen innerhalb eines Buches ausgebessert werden, an denen Leser häufig die Lektüre abbrechen, Marktprognosen für Serien erstellt werden und es kann eruiert werden, ob Fußnoten und Anhänge genutzt werden (vgl. Scherer 2015, S. 18). Reader Analytics können damit eine datenbasierte Beurteilung darüber, wie ein Buch am Markt angenommen werden wird, abgeben (vgl. Kirkbride 2016).

Reader Analytics können darüber hinaus auch übergreifende Erkenntnisse über das Leseverhalten der Leser generieren. So hat beispielsweise der E-Reader-Hersteller Kobo herausgefunden, dass 15 bis 20 Prozent der Leser auf mehreren Geräten lesen (vgl. Nawotka 2015). Die Chance, dass der Leser ein Buch beendet, steigt mit dem Anstieg der Gesamtlesezeit (vgl. Nawotka 2015). Im Sommer werden außerdem mehr Liebesgeschichten gelesen, während in der Zeit um Silvester mehr Sachbücher gelesen werden (vgl. Nawotka 2015). Dies sind alles verschiedene Analyseergebnisse von Leserdaten. Albazaz (2016a) legt zudem dar, dass auch der einzelne Leser besser verstanden werden kann: Beispielsweise lässt sich mit den Daten erkennen, ob das Lieblingsgenre eines Lesers Fantasyromane ist und er am liebsten die Nächte durchliest. Dies kann für individualisierte Buchempfehlungen wichtig sein.

Problematisch ist, dass die Verlage in der Regel nicht die Akteure sind, die Leserdaten über E-Books erhalten, sondern die Händler, die E-Books vertreiben, wie beispielsweise Amazon oder Apple (vgl. Davenport 2014b). Dies ist ein Grund, warum Verlage die Reader Analytics bisher nur selten verwenden, nämlich weil sie vorerst keinen Zugriff auf die von E-Book-Händlern erhobenen Daten haben. Ein weiterer Grund ist, dass viele Verlage zwar grundsätzlich an Reader Analytics interessiert sind, aber nicht wissen, wie sie mit den Daten umgehen sollen (vgl. Scherer 2015, S. 17). Diese Daten sind allerdings inzwischen vermehrt über externe Unternehmen verfügbar, sodass die Verlage entsprechende Daten zwar bezahlen müssen, aber nicht im eigenen Haus große Investitionen, beispielsweise in Big-Data-Softwares, tätigen müssen (vgl. Kirkbride 2016). Diese Reader-Analytics-Unternehmen werden im nächsten Kapitel vorgestellt.

Weiterhin muss hier erneut das grundsätzliche Big-Data-Problem des Datenschutzes angemerkt werden. Es werden recht umfassende Daten der Leser gesammelt und damit können viel präzisere Angaben zu gewissen Lebensgewohnheiten von bestimmten Lesern analysiert werden, beispielsweise wann, wo und wie viel sie lesen. Es ist also wichtig, dass mit den Daten erstens in verantwortungsvoller Weise umgegangen wird und den Lesern transparent kommuniziert wird, welche Daten für

welchen Zweck in welcher Weise gesammelt und analysiert werden, und zweitens nur so ausgewertet werden, dass keine Rückschlüsse auf einzelne Leser möglich sind.

Ein weiterer kritischer Aspekt ist, dass Lektoren sich sorgen, dass die Leserdaten dazu führen können, dass auch Inhalte in den Büchern zunehmend verändert werden (vgl. Rhomberg 2016b). Dies liegt daran, dass Lektoren oft ein Manuskript nicht ausschließlich mit einer rein ökonomischen Perspektive lesen, sondern auch beachten, welchen kulturellen, inhaltlichen, sprachlichen oder stilistischen Mehrwert es bietet. Dies ist von Bedeutung, da das Buch, wie bereits ausgeführt²³, auch ein Kulturgut darstellt. Eine Änderung des Inhalts von Büchern, um es wettbewerbsfähiger für den Markt zu machen, widerstrebt daher den meisten Lektoren.

Wie schon erwähnt²⁴, sammeln große Konzerne wie Amazon bereits Leserdaten aus E-Books, doch Verlage haben auf diese Daten keinen Zugriff. Es gibt bereits einige Unternehmen, die auf Reader Analytics spezialisiert sind und seit einigen Jahren im Rahmen der gesetzlichen Datenschutzbestimmungen Daten zum Leserverhalten sammeln und Verlagen, die daran interessiert sind, schon testweise zur Verfügung stellen (vgl. Scherer 2015, S. 18). Desweiteren gibt es E-Book-Plattformen mit unterschiedlichen Konzepten, welche ebenfalls Reader Analytics einsetzen. Welche Unternehmen in welcher Form bereits Daten sammeln und zur Verfügung stellen, wird nun erläutert. Dadurch kann auch eingehender erklärt werden, wie das Sammeln und Analysieren der Leserdaten funktioniert. Das Unternehmen, das sehr häufig, auch in der Buchbranchenpresse, in Verbindung mit Reader Analytics genannt wird, ist Jellybooks. Als weitere Beispiele werden im Folgenden noch die Self-publishing-Plattform Inkitt, der E-Book-Flatrate-Anbieter Skoobe, Readfy, ein Anbieter von werbefinanzierten E-Books und eines E-Book-Verleihsystems, und die gerade entstehende E-Book-Serien-Plattform Oolipo genannt.

Jellybooks

Das Startup Jellybooks wurde 2011 in London gegründet und war ursprünglich eine Plattform für E-Books; im Jahr 2015 veröffentlichte Jellybooks dann seine Reader-Analytics-Plattform (vgl. LinkedIn 2016; vgl. Hoffelder 2015). Damit will Jellybooks hauptsächlich für Verlage Daten aus E-Books sammeln und analysieren und ihnen damit Leserdaten zugänglich zu machen, die Amazon und andere große E-Book-Händler bereits sammeln, beispielsweise Daten darüber, ob die Leser das E-Book überhaupt öffnen, wann sie es lesen, wie schnell sie es lesen, in welchen Kapiteln sie Schwierigkeiten haben weiterzulesen sowie soziodemografische Daten der Leser (vgl. Rhomberg 2015; vgl. Gröger 2016). Im deutschsprachigen Raum haben unter anderem bereits die Verlage Ullstein und Piper sowie die Verlagsgruppe Random House mit Jellybooks zusammengearbeitet (vgl. Lenz 2016). In Abbildung 1 sind Ergebnisse der Analyse von Leserdaten von Jellybooks dargestellt. Das erste Buch

²³ s. Kapitel 3.3.1 ab Seite 25.

²⁴ s. Kapitel 2.3.2 ab Seite 21.

wurde von 62 Prozent der Leser beendet, wohingegen bei Buch 2 viele Leser das Buch spätestens nach den ersten 100 Seiten zur Seite legten. Das Buch 3 wurde sogar nur von 10 Prozent der Leser beendet.

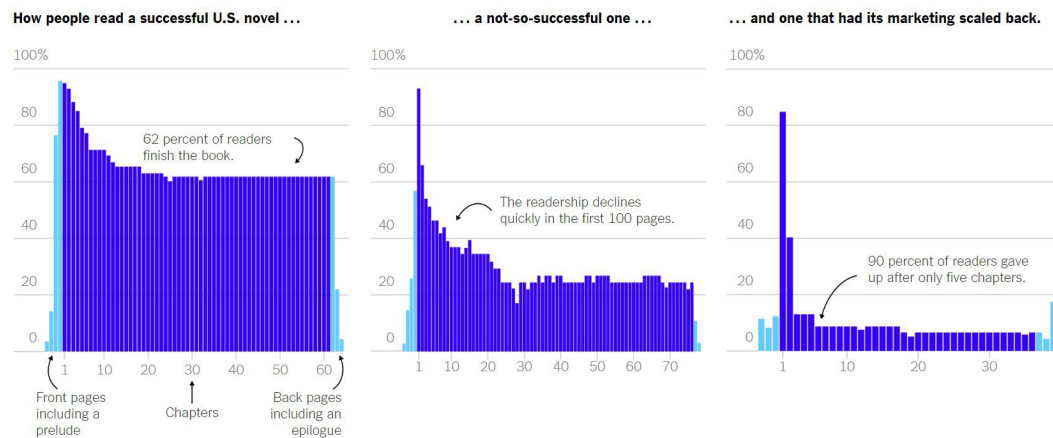


Abbildung 1: Drei verschiedene Ergebnisse der Reader Analytics von Jellybooks (Alter/Russell 2016)

Um an die Leserdaten zu gelangen, schenkt Jellybooks Lesern E-Books, teilweise vor der offiziellen Veröffentlichung (vgl. Alter/Russell 2016). Dafür werden vor allem Fokusgruppen verwendet (vgl. Scherer 2015, S. 19). In den verschenkten E-Books befindet sich eine Software namens *candy.js*, welche die Daten sammelt und lokal speichert. Am Ende stimmt der Leser mit einem Klick zu, dass seine gesammelten Daten an Jellybooks vermittelt und dort weiterverarbeitet werden (vgl. Rhomberg 2015).

Darüber hinaus verwendet Jellybooks den sogenannten Net Promoter Score. Dieser soll messbar machen, wie wahrscheinlich Leser ein Buch weiterempfehlen (vgl. Rhomberg 2016e). Der Net Promoter Score basiert auf der Frage an die Leser: »Würden Sie das Buch einem Freund empfehlen?« (vgl. Rhomberg 2016e). Die Antwort soll dabei auf einer Skala von 0–10 gegeben werden, 0 bedeutet dabei »auf keinen Fall«, 5 ist neutral und 10 bedeutet »auf jeden Fall« (vgl. Rhomberg 2016e). Diejenigen, die 9 oder 10 wählen, gelten als Promoter, also starke Fürsprecher des Buches, diejenigen, die 7 oder 8 wählen, gelten als Neutrale und diejenigen, die eine Zahl zwischen 0 und 6 gewählt haben, werden als Kritiker bewertet (vgl. Rhomberg 2016e). Der Prozentsatz der Kritiker wird von dem Prozentsatz der Fürsprecher abgezogen, das Ergebnis ist der Net Promoter Score (vgl. Rhomberg 2016e).

Ein positiver Aspekt von Jellybooks ist der für die Leser transparente Umgang mit ihren Daten. Die Leser wissen zum einen, dass sie das E-Book geschenkt bekommen, damit ihre Leserdaten gesammelt und analysiert werden können, und stimmen zum anderen bei Abschluss des Buches auch nochmal gezielt der Datenverwertung zu (vgl. Jellybooks 2017). Darüber hinaus können sie sogar ihre eigenen Leserdaten einsehen (vgl. Jellybooks 2017). Hier wird der Datenschutz also ernst genommen. Zudem werden die Leser sogar für die Freigabe ihrer Daten belohnt, indem sie das entsprechende E-Book umsonst bekommen.

Inkitt

Inkitt wurde im Jahr 2015 von Ali Albazaz und Linda Gavin gegründet und wird als »the world's first data-driven publisher« (Callari 2016) bezeichnet. Inkitt ist eine Selfpublishing-Plattform und konnte im Jahr 2016 bereits eine halbe Million Nutzer verzeichnen (vgl. Callari 2016; vgl. Shaffi 2016).

Inkitt hat einen intelligenten Algorithmus entwickelt, der das Leserverhalten der Nutzer analysiert und damit zukünftige Bestseller vorhersagen soll (vgl. Shaffi 2016). Wurde solch ein zukünftiger Bestseller entdeckt, sichert Inkitt sich die Rechte am Manuskript, veröffentlicht es auf E-Publishing-Plattformen wie Amazon und arbeitet auch mit Verlagen zusammen, um das Manuskript als Print-Buch herauszubringen, oder veröffentlicht es im Eigenverlag (vgl. Shaffi 2016; vgl. Albazaz 2016b, S. 16; vgl. Richters 2015). Der Algorithmus nimmt damit die Stelle des Lektors ein, die Entscheidung zur Veröffentlichung eines Buches über einen Verlag werden auf Grundlage der Daten der Leser mit Hilfe von etwa 1200 Leseverhaltensmustern getroffen (vgl. Albazaz 2016b, S. 16). Im April 2016 wurde das erste Buch, das anhand des Algorithmus ausgesucht wurde, bei Tor Books veröffentlicht (vgl. o. V. 2016b)

Die Vorteile dieses datenbasierten Vorgehens liegen laut Gründer Albazaz darin, dass potenzielle Bestseller-Autoren entdeckt werden können, die eventuell von Verlagen, deren Lektoren oft nach Bauchgefühl entscheiden, abgewiesen worden wären (vgl. Richters 2015). Durch die reine Nutzung von Daten treffen im Prinzip die Leser der Plattform die Entscheidung, welche Bücher erfolgreich werden und damit zur Veröffentlichung ausgewählt werden.

Die Aussage, dass Bestseller vorhergesagt werden können, laut Inkitt-Gründer Albazaz »mit einer statistischen Signifikanz von 99,99 Prozent und ohne Fehler zu machen« (Albazaz 2015, zitiert nach Richters 2015), wird aber auch kritisch gesehen. Auf diese Diskussion wird im Kapitel 3.3.2 ab Seite 49 näher eingegangen.

Zudem muss beachtet werden, dass die Zielgruppen der Nutzer der Plattform Inkitt nicht mit der Zielgruppe der Käufer eines Buches im Handel übereinstimmen muss. Es ist also fraglich, ob auf Grund der Leservorlieben auf der Plattform mit einer solch starken Sicherheit vorhergesagt werden kann, dass sich das Buch auch sehr gut verkaufen wird.

Readfy und Skoobe

Auch Anbieter, die kostenlose E-Books oder E-Book-Flatrates anbieten, haben die Möglichkeit, Leserdaten zu sammeln und zu analysieren und sie gegebenenfalls Verlagen zur Verfügung zu stellen. Daher sollen diese hier auch erwähnt werden.

Das Startup Readfy, das 2013 gegründet wurde, bietet den Lesern kostenlose E-Books, was mit gelegentlichen Werbeeinblendungen finanziert wird (vgl. Hüsing 2016). Außerdem gibt es auch die Möglichkeit, sich E-Books für 30 Tage mit einem Pay-per-Rent-System auszuleihen (vgl. o. V. 2016d, S. 35). Angeboten werden ca. 50.000 E-Books in deutscher Sprache, 120.000 monatliche Smartphone- und Tablet-

Nutzer verzeichnete Readfy im Jahr 2016 (vgl. o. V. 2016d, S. 35). Readfy kann anonymisierte Standardreportings, die Lesergruppen und Toptitel inklusive Detaildaten enthalten, und auf Wunsch auch Sonderauswertungen der Leserdaten erstellen und den Verlagen, die im Gegenzug auch Inhalte für Readfy bereitstellen, zur Verfügung stellen (vgl. Scherer 2015, S. 19). Verlage fragen diese Daten bei Readfy vermehrt nach (vgl. Gröger 2016). Nach eigenen Angaben bietet Readfy den Verlagen dabei demografische Daten wie Alter und Geschlecht sowie das Nutzerverhalten, wie beispielsweise den Lesefortschritt, das Abbruchverhalten während der Lektüre eines E-Books und die durchschnittliche Bewertung bei Rezensionen; dabei werden die Titel seitengenau analysiert (vgl. Readfy 2016a, S. 1). Jeder Readfy-Leser ist Teil der analysierten Datenmenge (vgl. Readfy 2016a, S. 1). Die Daten werden den Verlagen monatlich aktualisiert zur Verfügung gestellt (vgl. Readfy 2016a, S. 1).

Der E-Book-Flatrate-Anbieter Skoobe wurde 2010 gegründet und bietet drei Abo-Modelle mit Kosten zwischen 9,99 Euro und 19,99 Euro monatlich, mit denen E-Books aus dem Angebot von Skoobe ausgeliehen werden können (vgl. Sachs 2012; vgl. o. V. 2016f). Skoobe hat ca. 160.000 deutschsprachige E-Books im Angebot (vgl. o. V. 2016e, S. 35). Skoobe kann für Verlage monatlich rückblickend eine Übersicht über die am meisten gelesenen Titel und deren Entwicklung erstellen, damit Verlage eine Entscheidungsbasis dafür haben, welche Bücher sie in welcher Menge über Skoobe im Abonnement zu Verfügung stellen sollten (vgl. Scherer 2015, S. 19). Außerdem kann Skoobe die verlagseigenen Titel auswerten, die Nachfrage danach ist laut Skoobe aber sehr gering (vgl. Scherer 2015, S. 19).

Auch bei den Flatrate-Anbietern muss bedacht werden, dass die Lesergruppen der Abokunden nicht deckungsgleich mit den Endkunden im Handel sind und daher die Analyseergebnisse der Daten nur eingeschränkt auf den Handel übertragen werden können (vgl. Scherer 2015, S. 19).

Oolipo

Zuletzt soll hier Oolipo erwähnt werden, da es eine Art Streaming-Plattform für E-Books werden soll, was wiederum im Hinblick auf die hier erhobenen Daten interessant werden könnte. Oolipo ist ein Projekt des Verlags *Bastei Lübbe* und soll vor allem E-Book-Serien anbieten, die mit Bildern, Animationen, Musik und Geräuschen angereichert sind (vgl. o. V. 2016g). Das Modell hierfür soll ein Streamingdienst nach dem Vorbild von Spotify sein, bei dem die Nutzer einen monatlichen Betrag für die Nutzung bezahlen (vgl. o. V. 2016a; vgl. o. V. 2016g). Auch bei Oolipo sollen – ähnlich wie bei Inkitt – erfolgreiche Texte als E-Books und Print-Bücher veröffentlicht werden (vgl. o. V. 2016a). Oolipo soll datengetrieben sein, damit ist hier vor allem datengesteuertes Suchen und Entdecken gemeint (vgl. o. V. 2016a). Es ist aber vorstellbar, dass Oolipo – ähnlich wie die Streamingdienst-Vorbilder Spotify oder auch Netflix – die Leserdaten auch dazu verwenden wird, das Programm nach den Leserwünschen auszurichten. Leser sollen beispielsweise Serien folgen und liken

können (vgl. o. V. 2016g), was die Gelegenheit bieten würde, diese Daten zu sammeln und zu analysieren.

3.3 Perspektiven von Big Data in der Buchbranche

Nachdem in den vorigen Abschnitten die Doppelfunktion des Buches als Wirtschafts- und Kulturgut beschrieben, die Funktionsweise und Probleme des Marketings im Buchmarkt dargelegt sowie konventionelle und relativ neue Methoden der Datenanalyse vorgestellt wurden (CRM, Reader Analytics), sollen die gewonnenen Erkenntnisse in diesem Kapitel zur Beantwortung der Forschungsfrage zusammengefasst und untersucht werden. Dafür werden, abgeleitet aus den vorherigen Darlegungen und Erkenntnissen, vier Fragen formuliert, die hier beantwortet werden.

Zwei Fragen leiten sich aus dem Kapitel zu den Reader Analytics ab. Es wurde bereits erläutert²⁵, welche Erkenntnisse die Analyse von Leserdaten generieren kann und wie sie bereits in der Buchbranche eingesetzt werden. Daher soll in diesem Kapitel folgende Frage beantwortet werden: Welches Potenzial hat der Einsatz von Reader Analytics für die Buchbranche? Aus den Ausführungen wurde zweitens deutlich, dass einige Akteure im Bereich Reader Analytics, beispielsweise Inkitt-Gründer Ali Albazaz, der Meinung sind, dass mit Reader Analytics Bestseller vorhergesagt werden können. Hieraus ergibt sich die zweite Frage nach der Möglichkeit solcher Voraussagen.

Die dritte Frage leitet sich aus dem Kapitel über die Fallbeispiele ab. Hier wurden die Streamingdienste mit Flatrate-Bezahlmodell, Netflix und Spotify, als Fallbeispiele für die Verwendung von Nutzerdaten zur Anpassung des Angebots und für individuelle Empfehlungen für die Nutzer erläutert.²⁶ Ein Streamingdienst namens Oolipo ist auch für E-Books geplant.²⁷ Daher soll das Potenzial von Streamingdiensten und den Daten, die hierbei gesammelt und ausgewertet werden können, für die Buchbranche analysiert werden. Die dritte Frage lautet also: Inwieweit wird sich die Buchbranche insgesamt mehr an den Lesern orientieren?

Schließlich lautet die letzte Frage: Inwieweit können Daten die Intuition ersetzen? Wie schon in Kapitel 2 dargelegt, ist ein wichtiger Aspekt von Big Data, und damit auch eine Chance, Entscheidungsfindungen auf Basis von Daten zu verbessern. In Bezug auf den Einsatz von Daten in der Verlagsbranche wird von verschiedenen Autoren immer wieder erwähnt, dass Entscheidungen über die Veröffentlichung eines Buches in Verlagen oft intuitiv getroffen werden (vgl. Neary 2016; vgl. Anderson 2015; vgl. Lichtenberg 2014; vgl. Reid 2012). Wie bereits dargestellt²⁸, galt in der Buchbranche bisher die Annahme, dass sich Bestseller nur schwierig vorher-sagen lassen, da ein solcher Erfolg von schwer kalkulierbaren Faktoren abhängt, und daher die Intuition von erfahrenen Lektoren zur Einschätzung des Erfolgs von

25 s. Kapitel 3.2.3 ab Seite 38.

26 s. Kapitel 2.3.1 ab Seite 19.

27 s. Kapitel 3.2.3 ab Seite 44.

28 s. Kapitel 3.1.2 ab Seite 27.

Büchern maßgeblich ist (vgl. Rautenberg/Wetzel 2001, S. 97). Der Einsatz von Daten wiederum wird oft als das Ende dieser Intuitionsentscheidungen gesehen (vgl. z. B. Posth 2013). Dies ist problematisch, weil in der Buchbranche die Intuition eine wichtige Entscheidungsgrundlage ist. Es konnte auch in dieser Arbeit bereits dargelegt werden, dass die Reader Analytics von Startups wie Inkitt dazu verwendet werden, Bestseller auf Grund der Leserdaten zu identifizieren und sie daraufhin erst zu veröffentlichen. Diese Vorgehensweise benötigt also keinen Lektor mehr, der eine intuitions- und erfahrungsbasierte Einschätzung abgibt. Daher soll untersucht werden, ob Daten wirklich die Intuition in den Verlagen vollständig ablösen werden.

3.3.1 Welches Potenzial haben die Reader Analytics?

Die Reader Analytics bieten erstmals die Chance, das »Phänomen des ungelesenen Buches« (Zeckert 2007, S. 153) zu ergründen: Bücher, die gekauft werden, werden nicht unbedingt gelesen. Das hat verschiedene Gründe, beispielsweise werden Bücher manchmal nur gekauft, um sie ins Regal zu stellen, damit die Käufer sich etwa vor Besuchern als belesen darstellen können, oder der Leser findet keine Zeit, das gekaufte Buch wirklich zu lesen (vgl. Zeckert 2007, S. 153). Der Buchkauf ist also nicht unbedingt ein Indikator für die Rezeption eines Buches (vgl. Zeckert 2007, S. 153). Die Reader Analytics können nun einerseits nachweisen, ob ein Buch überhaupt gelesen und beendet wurde oder nicht. Andererseits basieren die Analysen von Reader Analytics eben nur auf tatsächlichen Lesern. Buchkäufer, die ein Buch nicht lesen, können hier nicht abgebildet werden.

Wenn das bedacht wird, haben die Reader Analytics dennoch Potenzial dafür, Verlagen interessante Erkenntnisse bieten zu können. Wie bereits dargestellt²⁹, können Zielgruppen, Leserbedürfnisse und Lesegewohnheiten identifiziert werden, was den Verlagen die Chance bietet, Marketing oder auch das Programm mehr an die Leserwünsche anzupassen. Mittels zusätzlichen Befragungen, wie sie beispielsweise Jellybooks durchführt, können Leser zusätzlich Kritikpunkte und Wünsche äußern. Außerdem erfahren Verlage durch Reader Analytics, wie häufig das Buch überhaupt gelesen wurde, und können damit den Verkaufserfolg besser einschätzen. Wenn Bücher sogar schon vor Veröffentlichung analysiert werden, wie es beispielsweise bei Inkitt geschieht, können zudem Bücher mit großem Erfolgspotenzial entdeckt werden.

Auffallend ist, dass es bei einigen Big-Data-Unternehmen der Buchbranche, beispielsweise bei Inkitt, letztendlich darum geht, Bücher für die klassische Veröffentlichung als Printbuch oder E-Books zu finden, dabei aber den konventionellen Weg der Auswahl durch Lektoren in den Verlagen, die oft intuitiv entscheiden, zu umgehen. Die Leserdaten sollen hier also die Entscheidung treffen, welche Bücher veröffentlicht werden. Dies würde bedeuten, dass die Verlage die Entscheidungshoheit darüber, welche Bücher veröffentlicht werden, teilweise aufgeben müssen, um

²⁹ s. Kapitel 3.2.3 ab Seite 38.

quasi die Leser entscheiden zu lassen, welche Bücher es auch im Handel geben soll. Skoobe und Readfy orientieren sich an Streaming-Diensten aus anderen Bereichen wie Musik oder Serien, um E-Book-Flatrates anzubieten und mit Hilfe der Analyse der Leserdaten das Angebot den Leserwünschen anzupassen. Hier haben die Leser also indirekt Einfluss darauf, welche Bücher aus den Verlagen es auch im Flatrate-Angebot gibt. Jellybooks kann vor allem als Daten-Dienstleister für Verlage gesehen werden, welche damit die Gelegenheit bekommen, Leserdaten zu erhalten, auf die sie sonst keinen Zugriff haben. Hier haben die Leser erstmals die Chance, Kritik oder Wünsche zu äußern, durch zusätzliche Befragungen und durch die Verwendung des Net Promoter Score. Ein interessanter Aspekt der Reader Analytics ist es daher, dass indirekt nun auch die Leser und damit die Endkunden bei Entscheidungsfindungen beteiligt sind und damit das Angebot an veröffentlichten Büchern beeinflussen. Das bedeutet auch, dass die Leser stärker und unmittelbarer in den Gesamtprozess der Entstehung eines Buches integriert sind als vorher, da ihre Rezeption nun auch direkt zu Entscheidungen für eine Veröffentlichung beitragen kann.

Wie bereits erwähnt³⁰, werden die Reader Analytics in der Buchbranche bisher positiv bewertet und erste deutsche Verlage haben bereits begonnen, sie zu nutzen. Interessant wird dabei sein, wie sich der Einfluss von Leserdaten entwickeln wird. Im Moment werden diese hauptsächlich dafür verwendet, schon fertige Bücher zu veröffentlichen oder zu vermarkten oder bei Lesern beliebte Bücher aus dem Selfpublishing-Bereich zu entdecken. Grundsätzlich wäre es aber auch möglich, Bücher inhaltlich zu verändern und den Leserwünschen anzupassen. Darauf wird unter anderem im folgenden Kapitel näher eingegangen.

Wie schon dargelegt³¹, sind die Marktforschungsdaten, die bisher überwiegend als Datengrundlage für das Marketing in Verlagen und im Buchhandel verwendet wurden, nur von eingeschränkter Aussagekraft. Es hat sich gezeigt, dass die Reader Analytics umfangreichere und detailliertere Daten liefern können, die insbesondere für das Marketing relevant sind, wie beispielsweise Daten darüber, welche Lese-gewohnheiten die Zielgruppe für ein Buch hat. Damit ist es möglich, das Marketing-budget im Verlag effektiver einzusetzen, indem die Zielgruppen gezielter ange-sprochen werden kann. Grundsätzlich ist das Potenzial der Reader Analytics daher als positiv zu bewerten, da sie neue Einsichten in die Leser bieten, die bisher nicht möglich und zudem nur den Konkurrenten des klassischen Buchmarkts wie Amazon zugänglich waren. Die Reader Analytics können daher auch als wichtiges Werkzeug im Konkurrenzettbewerb mit Online-Händlern wie Amazon gesehen werden, da hierbei nun Daten zugänglich sind, die Amazon bereits seit Jahren von allen Lesern, die E-Books auf Amazons E-Reader Kindle oder über die Kindle-App lesen, spei-chert. Das zeigt auch die überwiegend positive Resonanz der Buchbranchenpresse. Beachtet werden sollte aber, dass die Testleser nicht alle Leser- und damit spätere

30 s. Kapitel 3.2.2 ab Seite 33.

31 s. Kapitel 3.1.2 ab Seite 27.

Käufergruppen repräsentieren. Daher sollten sich Verlage nicht allein auf die Reader Analytics verlassen, sondern sie als eine zusätzliche Quelle an interessanten Daten betrachten. Zudem kann die Problematik auch verbessert werden, indem die Testleser gezielt nach Kriterien der Repräsentativität ausgewählt werden. Bisher nutzen allerdings nur wenige Verlage in Deutschland die Reader Analytics. Die überwiegend positive Resonanz auf Reader Analytics von Verlagen, die Reader Analytics getestet haben, sowie von der Buchbranchenpresse könnte dafür sprechen, dass zukünftig immer mehr Verlage die Reader Analytics verwenden werden. Allerdings gibt es auch Gründe, die dagegen sprechen, nämlich Zurückhaltung bei Investitionen oder die Sorge darüber, inwieweit die Daten den bisherigen Workflow im Verlag verändern würden.

Andrew Rhomberg, Gründer von Jellybooks, berichtet, dass in den Verlagen, welche die Reader Analytics von Jellybooks bereits getestet haben, besonders die Marketing- und PR-Abteilungen von den Leserdaten begeistert sind, wohingegen die Lektorate oft eher skeptisch sind (vgl. Rhomberg 2016b). Das liegt daran, dass die Frage, inwieweit Reader Analytics auch den Inhalt von Büchern beeinflussen können, in den Verlagen und bei Autoren für Unsicherheit sorgt (vgl. Anderson 2015). Laut Rhomberg besteht keine Gefahr, dass bei belletristischen Inhalten Kapitel oder Passagen verändert werden, da diese ein Gesamtkunstwerk darstellen (vgl. Lenz 2016). Bei Jellybooks ist das Buch schon fertig, wenn es testgelesen wird, und die Reader Analytics sollen hauptsächlich Erkenntnisse liefern, die dem Marketing helfen. Rhomberg gibt allerdings zu, dass es nicht ausgeschlossen ist, dass sich Autoren zukünftig dazu entscheiden, ihre Bücher auch umzuschreiben (vgl. Rhomberg 2016a). Gerade bei Fach- und Sachbüchern können mit Reader Analytics schwierig zu verstehende Passagen identifiziert werden, welche dann noch vom Lektor und Autor überarbeitet werden können (vgl. Lenz 2016). Auch bei den anderen Reader-Analytics-Konzepten ist bisher nicht davon die Rede, Bücher zu verändern. Fraglich ist, ob das so bleiben wird, vor allem, wenn Reader Analytics immer häufiger und von immer mehr Verlagen verwendet werden, oder ob die Reader Analytics nicht doch irgendwann dafür genutzt werden, um auch inhaltliche Modifizierungen bei Büchern vorzunehmen, bevor sie beispielsweise endgültig veröffentlicht werden. Denkbar wären hier sogar mehrere Testphasen eines Buches, bis das vermeintlich perfekte Buch gefunden wird, das sich dann auch gut verkaufen wird. Allerdings würden sich vermutlich viele Verlage und Autoren nicht darauf einlassen, da sie möglicherweise ein fertiges Manuskript als Kunstwerk ansehen, das nachträglich nicht mehr verändert werden kann. Allerdings ist es bisher so, dass Verlagslektoren den Autoren Änderungen vorschlagen, welche diese dann umsetzen. Die Änderungsvorschläge am Manuskript wären also nicht neu, sondern kämen nur von einer anderen Stelle, nämlich von den Lesern anstelle des Lektors, und gegebenenfalls zu einem späteren Zeitpunkt, nämlich nach der Zusammenarbeit mit dem Lektor.

Zudem kann kritisiert werden, dass die Zielgruppe, welche die Plattform Inkitt zum Lesen benutzt, nicht deckungsgleich mit der Zielgruppe ist, die ein Buch, das über Inkitt entdeckt wurde, im Handel kaufen würden. Es ist also fraglich, ob auf Grund der Leservorlieben auf der Plattform mit einer solch starken Sicherheit vorhergesagt werden kann, dass sich das Buch auch sehr gut verkaufen wird. Jellybooks verwendet höchstens einige Hundert Testleser, die das entsprechende E-Books im Austausch gegen ihre Daten geschenkt bekommen und sich dessen auch bewusst sind. Die Bedingungen, unter denen diese Testbücher gelesen werden, sind andere, als wenn ein Leser das Buch gekauft hat und im Privaten liest, ohne dass – zumindest wissentlich – seine Daten gespeichert werden (vgl. Meadows 2016). Dies kann das Leseverhalten der Testleser beeinflussen. Außerdem wird die Aussagekraft dadurch verringert, dass die Leser das Buch in der Regel kostenlos erhalten. Zudem ist die eher geringe Menge an Testlesern vermutlich nur begrenzt repräsentativ für die Gesamtleerschaft.

Eine weitere Befürchtung, vor allem auch der Testleser³², ist, dass mit Hilfe der Reader Analytics Muster erkannt werden, wie das ideale, massenkompatible Buch geschrieben und konzipiert sein muss, und Verlage zukünftig hauptsächlich solche Bücher veröffentlichen und dies die Vielfalt im Buchmarkt einschränken würde. Dies wird voraussichtlich erstmal nicht geschehen, da der Fokus der Verlage vermutlich hauptsächlich ist, die Leserdaten für Marketingzwecke zu verwenden – beispielsweise auch, um weniger massentaugliche Bücher besser zu verkaufen –, und viele, vor allem kleinere Verlage mutmaßlich die Büchervielfalt überwiegend gewährleisten wollen³³.

Dieser Aspekt führt allerdings zu der nächsten Frage:

3.3.2 Lassen sich mit Big Data Bestseller vorhersagen?

Es wurde bereits erwähnt³⁴, dass der Gründer von Inkitt der Meinung ist, mit seinen Algorithmen auf Basis von Leserdaten mit einer Sicherheit von beinahe 100 Prozent Bestseller vorhersagen zu können. Im Bezug auf das Buch *The Bestseller Code*, in dem die Autoren Jodie Archer und Matthew L. Jockers ebenfalls einen Algorithmus zur Vorhersage von Bestsellern auf Basis des Inhalts gefunden haben wollen, kam in der

32 Beispielsweise in folgendem Blogeintrag: <http://www.jos-truth.de/2015/11/jellybooks-wir-wissen-wie-du-liest> oder folgendem Forumsbeitrag: <http://www.piper-fantasy.de/forum/testleser-bei-jelly-books> zu sehen.

33 Zumindest ist die in der Verlagsbranche übliche Quersubventionierung ein Hinweis darauf. Der Börsenverein des Deutschen Buchhandels definiert diese folgendermaßen: Mit den Einnahmen von den Büchern, die dem Verlag Gewinn bringen, veröffentlichen Verlage weniger erfolgreiche und oft auch auf ein Nischen-Publikum ausgerichtete Bücher und tragen somit zur kulturellen Vielfalt bei (vgl. Börsenverein des Deutschen Buchhandels o. J.). Diese Selbsteinschätzung der Verlagsbranche kann allerdings auch kritisch gesehen werden, da die Vielfalt auch ein zufälliger, statt ein gewollter Effekt der Quersubventionierung sein könnte. Laut Ploeg (2004, S. 16) werden Bücher immer mit der Hoffnung darauf, dass sie Bestseller werden, quersubventioniert und es besteht auch keine Garantie, dass mit den Einnahmen von Bestsellern wirklich unpopuläre Bücher finanziert werden.

34 s. Kapitel 3.2.3 ab Seite 43.

Buchbranche eine Diskussion darüber auf, ob das wirklich möglich ist. Um diese Diskussion zu verstehen, muss zunächst kurz dargestellt werden, was das Buch beinhaltet.

Die Prämisse in *The Bestseller Code* ist es, dass nur ca. 0,5 Prozent der veröffentlichten Bücher zu Bestsellern³⁵ werden, was eine sehr geringe Anzahl ist (vgl. Archer/Jockers 2016, S. 3). Zudem ist bekannt, dass viele spätere Bestseller – wie die *Harry Potter*-Reihe, *Herr der Fliegen* oder *Verblendung* – zunächst von mehreren Verlagen abgelehnt wurden (vgl. Archer/Jockers 2016, S. 9 f.). Daraus schließen die Autoren, dass es bisher einem Glücksspiel glich, Bestseller zu finden oder vorherzusagen (vgl. Archer/Jockers 2016, S. 11). Das Computermodell, das die Autoren entwickelten, hat bei vielen Büchern³⁶, die zu Bestsellern wurden, mit einer Sicherheit von über 90 Prozent vorhergesagt, dass sie tatsächlich Bestseller sind; insgesamt wurden 80 Prozent der *New York Times*-Bestseller der letzten 30 Jahre erkannt (vgl. Archer/Jockers 2016, S. 6 f.).

An dieser Stelle muss erläutert werden, wie dieses Modell funktioniert. Aus ursprünglich über 20.000 Merkmalen von Büchern, die Bestseller waren, wurden letztendlich 2.800 relevante Merkmale identifiziert, welche Bestseller erkennbar machen sollen (vgl. Archer/Jockers 2016, S. 24 f.). Dafür mussten die Computer lernen, Bücher zu lesen und diese Merkmale zu extrahieren (vgl. Archer/Jockers 2016, S. 25). Daraufhin wurden Computerprogramme verwendet, welche in der Lage waren, Merkmalsets zu analysieren und verborgene Muster zu erkennen; diese Analysephase wird als Machine Learning bezeichnet (vgl. Archer/Jockers 2016, S. 25). Dazu musste der Computer das sogenannte Text Mining lernen: Dafür wurde er mit den Texten aus Bestsellern und Nicht-Bestsellern »gefüttert«, sodass er lernt, anhand der Merkmale die Bestseller von den Nicht-Bestsellern zu unterscheiden (vgl. Archer/Jockers 2016, S. 25). Das Ergebnis hierbei war, dass der Computer durchschnittlich 80 Prozent der Bestseller identifizieren konnte (vgl. Archer/Jockers 2016, S. 26). Einige der Merkmale erläutern die Autoren detaillierter in ihrem Buch – hierzu gehören etwa das Thema des Buches, der Plot, der Schreibstil und die Gestaltung der Charaktere (vgl. Archer/Jockers 2016, S. 33–182).

Einen Vorteil von Text Mining sehen die Autoren darin, dass Manuskripte ausgewählt und neue Autoren gefunden werden können, unabhängig davon, ob der Autor schon bekannt ist und was für eine Reputation er bereits hat; dies würde der Buchindustrie dabei helfen, nicht nur erfolgreicher zu werden, sondern auch weiterhin vielfältig zu bleiben (vgl. Archer/Jockers 2016, S. 13).

35 Das Kriterium für einen Bestseller ist hier, dass das Buch es auf die New York Times-Bestsellerliste schafft (vgl. Archer/Jockers 2016, S. 3).

36 Als Beispiele wurden hier *Inferno* von Dan Brown, *The Lincoln Lawyer* von Michael Connelly, *The Friday Night Knitting Club* von Kate Jacobs, *The Luckiest Girl Alive* von Jessica Knoll, *The Martian* von Andy Weir, *The First Phone Call from Heaven* von Mitch Alborn, *The Art of Fielding* von Chad Harbach und *Bared to You* von Sylvia Day genannt (vgl. Archer/Jockers, S. 7).

Verständlicherweise hat das Erscheinen von *The Bestseller Code* für Diskussionen in der Buchbranche darüber gesorgt, ob die These, die die Autoren aufstellen, wirklich haltbar ist und ob es grundsätzlich möglich sein kann, Bestseller vorherzusagen. Im nächsten Schritt soll diese Diskussion in der Buchbranche nachvollzogen werden.

Chris Sim, Gründer von Kadaxis³⁷ hat ebenfalls einen Algorithmus zur Bestseller-Vorhersage entwickelt und ist der Meinung, dass der Text alleine dafür nicht ausreicht (vgl. Sim 2016). Auch Mike Shatzkin, der bereits seit 50 Jahren in der Buchindustrie arbeitet und als Vordenker und prominenter Beobachter des Übergangs der Branche in das digitale Zeitalter gilt³⁸, ist der Meinung, dass es nicht nur auf Basis des Inhalts eines Buches möglich ist, eine Vorhersage darüber zu treffen, ob ein Buch zum Bestseller wird. Der Text eines Buches ist zwar das Kernprodukt, aber weitere Faktoren, wie Sales und Marketing, Branding, die Bekanntheit des Autors und das Coverdesign beeinflussen, ob die Kunden ein Buch entdecken und kaufen (vgl. Sim 2016; vgl. Shatzkin 2016). Um eine akkurate Vorhersage treffen zu können, müsste ein Computer auch solche Faktoren berücksichtigen (vgl. Sim 2016). Zudem ist laut Shatzkin auch die Höhe der Auflage und damit die Menge an Exemplaren, die in den Buchläden und bei Online-Händlern zur Verfügung steht, relevant dafür, wie viele Bücher verkauft werden (vgl. Shatzkin 2016).³⁹

Weiterhin kritisiert Sim an *The Bestseller Code*, das hier zu wenige Bücher für die vielen Merkmale, die verwendet wurden, ausgewertet wurden, um den Algorithmus wirklich aussagekräftig zu machen; die Gefahr dabei ist, dass der Algorithmus zwar Muster findet, diese sind jedoch möglicherweise nicht repräsentativ für das, was vorhergesagt werden soll (vgl. Sim 2016). Zuletzt erläutert Sim, dass Bestsellerlisten nicht die tatsächliche Verkaufszahlen widerspiegeln; von einem Buch, das beispielsweise für eine Woche auf einer Bestsellerliste landet, können trotzdem weniger Ausgaben verkauft worden sein als von einem Buch, das nie auf einer Bestsellerliste gewesen ist (vgl. Sim 2016). Sim plädiert dafür, Daten, welche die Leserszufriedenheit beinhalten, mit den Daten zum Text zu verknüpfen, um Aussagen dazu treffen zu können, welche Auswirkung Bücher auf bestimmte Leser haben und warum (vgl. Sim 2016). Dies wäre beispielsweise möglich, wenn man Algorithmen, die auf Text Mining basieren, mit Reader Analytics verbinden würde.

Neil Balthaser, Gründer von Intellogo⁴⁰, hat hingegen eine Replik zu Shatzkins Beitrag veröffentlicht, in der er die Meinung vertritt, dass Algorithmen durchaus

37 Kadaxis ist ein Unternehmen, das auf Machine-Learning-basierte Produktentwicklung für die Buchbranche spezialisiert ist.

38 vgl. <http://www.idealogue.com/about-us/about-mike-shatzkin/> [08.02.2017]

39 Denn wenn von einem großartigen Buch nur 5.000 Exemplare gedruckt werden, von denen nur wenige auch in große Buchläden geliefert werden, kann es laut Shatzkin gar nicht zum Bestseller werden, auch wenn es Bestseller-Potenzial hätte (vgl. Shatzkin 2016).

40 Intellogo ist ein Anbieter von Empfehlungs-Bots. Außerdem bietet Intellogo mit Author Tools eine Software für Autoren, welche Erzähl-Elemente, beispielsweise wie konsistent erzählerische Fakten, Charakterzüge und Charakterentwicklung sind oder wie fesselnd Konflikte, Höhepunkte und Auflösungen beschrieben werden, grafisch darstellen kann (vgl. Fairchild 2016).

dabei helfen können, die Wünsche und Vorlieben der Leser besser zu verstehen (vgl. Balthaser 2016). Die Auflagehöhe eines Buches werde darüber hinaus nicht hauptsächlich vom Buchmarkt, sondern durch die Nachfrage der Leser bestimmt (vgl. Balthaser 2016).

An der Diskussion ist zu sehen, dass es unterschiedliche Ansichten dazu gibt, ob es wirklich möglich ist, Bestseller vorherzusagen. Hierbei muss erwähnt werden, dass es unterschiedliche Ansätze gibt: Das Startup Inkitt will Bestseller auf Basis von Reader Analytics vorhersagen können, wohingegen es beim Text-Mining-Ansatz, der für den Bestseller-Vorhersage-Algorithmus in *The Bestseller Code* verwendet wurde, um den Inhalt eines Buches geht. Mit beiden Ansätzen sind sicher bessere Prognosen möglich als bei rein intuitiven Entscheidungen, da die datenbasierten Analysen zumindest Muster aufdecken können, die darauf hinweisen, dass ein Buch ein Bestseller werden könnte. Allerdings sind immer auch andere Faktoren relevant, wie beispielsweise Trends oder auch unvorhersehbare Faktoren wie bestimmte Ereignisse (wie Naturkatastrophen oder politische Geschehnisse). Daher wird es wahrscheinlich – auch wenn Bestseller-Vorhersage-Methoden irgendwann in Verlagen angewandt werden sollten – weiterhin unerwartete Bestseller geben oder erfolgsversprechende Bücher, die sich weniger verkaufen als erwartet.

Interessant wäre es, zu sehen, welche Ergebnisse produziert würden, wenn beide Ansätze, also sowohl Text Mining als auch Reader Analytics, miteinander verknüpft und zusammen verwendet würden. Vorstellbar wäre hier durch die Kombination der beiden unterschiedlichen Datenarten, die für sich bereits eine gewisse Aussagekraft haben, dass durchaus deutlich mehr Bestseller entdeckt werden könnten, die vielleicht durch die bisherige konventionelle Selektion von Lektoren nicht entdeckt worden wären. Was das bedeuten könnte, wird in Kapitel 4 näher beleuchtet.

3.3.3 Inwieweit wird sich die Buchbranche insgesamt mehr an den Lesern orientieren?

Anbieter von Streamingdiensten haben Zugang zu Nutzerdaten und daher die Möglichkeit, Daten darüber zu erhalten, wie, wann und wo die Nutzer die Angebote nutzen und welche Vorlieben sie haben. Mit Streaming können Zielgruppen daher gezielt angesprochen werden und die Vorlieben von Individuen ausgewertet werden. Für die Nutzer liegen die Vorteile bei Streamingdiensten darin, dass ihnen auf Basis der Datenanalyse individuelle Empfehlungen präsentiert werden können. Meist werden die Empfehlungen besser, je mehr die Nutzer das Angebot nutzen, da die Algorithmen, die die Datenanalysen umsetzen, lernfähig sind. Unter anderem wegen dieser Individualisierung werden Streamingdienste immer beliebter. Allgemein kann ein Trend zu individualisierten Angeboten festgestellt werden.⁴¹ Das zeigen die

41 So hat beispielsweise die Nachrichten-Plattform Blendle in den Niederlanden bereits damit begonnen, in einem Abomodell individuelle Newsfeeds anzubieten, wofür ebenfalls die Nutzerdaten analysiert und mit Hilfe von Algorithmen umgesetzt werden; auch hier gelten Netflix und Spotify als Vorbilder (vgl. Lomas 2016). Zusätzlich zu den vom Algorithmus ausgewählten

steigenden Abonnentenzahlen von Anbietern wie Netflix und Spotify (vgl. Lindner 2017; vgl. Shanley/Weber 2016).

Streaming- und Flatratemodelle nehmen also zu und sind bei den Nutzern beliebt. Daher ist es denkbar, dass solch ein Modell auch im Bereich Bücher Erfolg haben könnte. Auch hier könnten die Daten, die aussagen, was die Leser lesen, wo sie lesen und wie häufig und intensiv sie lesen, dazu verwendet werden, ihnen maßgeschneiderte Empfehlungen zu geben und sie damit zum Lesen zu animieren. Dies wiederum würde weitere Reader Analytics zu den gelesenen Büchern generieren. Die Chance für die Buchbranche liegt darin, Daten zur Zielgruppe und zur Erreichbarkeit der Zielgruppe zu erhalten und diese für Entscheidungen zu nutzen, wie es Anbieter wie Netflix bereits praktizieren (vgl. Balthaser 2016). Mit Oolipo ist ein solches Modell in Planung und soll voraussichtlich 2017 gelauncht werden (vgl. o. V. 2016g). Es bleibt abzuwarten, wie Oolipo sich im Bereich E-Book-Plattformen etablieren wird. Das Modell von Oolipo erinnert an Netflix. Es sind vor allem multimediale E-Books geplant, die als Serien konzipiert sind. Ebenso wie manche Nutzer bei Netflix beim sogenannten ›binge-watching‹ viele Folgen einer Serie hintereinander schauen, sind laut Meadows auch Bücher ›binge-readable‹ (Meadows 2016). Oolipo würde damit ein »interaktives, daten-getriebenes, visualisierendes Erzählen« (Mayer-Schönberger 2014) ermöglichen, welches Big-Data-Experte Mayer-Schönberger bereits 2014 als neues, erfolgversprechendes Big-Data-Modell für die Buchbranche vorgeschlagen hat. Das Modell von Oolipo ist also sehr am Serien-Streaming orientiert. Das als Produkt angebotene Medium wäre hier ein um visuelle und akustische Elemente erweiterter Text. Die Nutzerdaten würden dazu verwendet, das Angebot den Leserbedürfnissen anzupassen sowie den Lesern individuelle Empfehlungen zu geben.

Beachtet werden muss, dass einige Streamingdienste wie Netflix oder auch Amazon Video inzwischen sowohl Anbieter als auch Produzenten von Inhalten sind. Damit haben sie direkten Zugang zu den Endkunden ihrer eigenen Produkte. Diese Streamingdienste sind also – zumindest bei den Eigenproduktionen – eine in sich geschlossene Wertschöpfungskette. Im Buchmarkt ist die Wertschöpfungskette auf drei Akteure verteilt, nämlich auf den herstellenden Buchhandel, den Zwischenbuchhandel und den verbreitenden Buchhandel. Verlage sind hier nur Produzenten von Büchern, der Anbieter ist dann der verbreitende Buchhandel in Form von stationären Buchläden und Online-Buchhändlern. Verlage haben daher wenig Kontakt zu den Endkunden und keinen Zugriff auf Leserdaten von E-Book-Flatrate-Anbietern. Sie haben damit keine Möglichkeit, ihr Programm an die Wünsche der Plattformnutzer anzupassen, solange die Anbieter die Daten den Verlagen nicht zur Verfügung stellen. Vielmehr beeinflussen die Anbieter sogar die Wertschöpfungskette innerhalb der Verlage. Der Einsatz von Big Data, also die Analyse von Leserdaten

Artikeln sollen auch von Blendle kuratierte Artikel angeboten werden, um die Gefahr zu umgehen, dass die Leser sich nur in einer Filterbubble befinden (vgl. Lomas 2016).

oder Flatrate-Angebote im E-Book-Bereich, führt dazu, dass die dadurch erhaltenen Informationen über die Kunden Elemente der Wertschöpfungsstufe in Verlagen ersetzen, nämlich die des Lektorats. So plant Oolipo einige E-Books weiterzuverwerten und bei Verlagen zu veröffentlichen. Dies ist ebenso der Fall bei Reader-Analytics-Plattformen wie Inkitt, die anhand der Leserdaten auswählen, welche Bücher sie über Verlage veröffentlichen wollen. Ob die Gefahr besteht, dass die Lektorate bald durch Daten ersetzt werden, wird im folgenden Unterkapitel erörtert.

Ein ähnliches Modell wie Netflix und Spotify, nämlich Flatrate-Angebote oder kostenlose, werbefinanzierte Angebote, bieten zudem bereits die ab Seite 43 vorgestellten Anbieter Skoobe und Readfy. Diese sind aber, ähnlich wie Spotify, nur Anbieter von schon vorhandenen Inhalten, und daher keine Produzenten und auch keine Entscheider, welche Inhalte später klassisch veröffentlicht werden sollen

Das Potenzial von Streaming- und Flatratemodellen in der Buchbranche kann aufgrund des Erfolgs von Diensten wie Netflix und Spotify als gut bewertet werden, zumal bereits bestehende Anbieter wie Skoobe und Readfy sich bisher am Markt behaupten und ihr Angebot stetig vergrößern können.⁴² Die Buchbranche würde damit dem Trend zu Flatratemodellen, der bereits seit einiger Zeit in der Film-, Fernseh- und Musikindustrie vorhanden ist, und dem Trend zu individualisierten Angeboten und Empfehlungen folgen. Die Vorbilder wie Netflix und Spotify sind erfolgreich, die Nutzer sind überwiegend zufrieden mit den Angeboten und mit personalisierten Empfehlungen könnten Nutzer von E-Book-Flatrate-Plattformen zum Lesen animiert werden. Zusätzlich erhalten die Anbieter Einsichten in die Lesegewohnheiten der Leser.

Es bleibt abzuwarten, wie sich Skoobe und Readfy weiterhin am Markt behaupten werden und wie erfolgreich Oolipo sein wird. Wenn sie sich als erfolgreich erweisen, scheint es wahrscheinlich, dass sich Flatrate-Modelle im Buchbereich ebenso fest etablieren werden wie Netflix und Spotify bereits etabliert sind. Dies würde vermutlich dazu führen, dass sich mehr Verlage bereit erklären, ihre E-Books auf solchen Plattformen zur Verfügung zu stellen, was für die Verlage eine zusätzliche Einnahmequelle bedeuten würde. Für die Nutzer hätte das den Vorteil, dass sich das Angebot vergrößert, was wiederum den Plattformen mehr Nutzer einbringen würde. Ein Teil des Erfolgs von Netflix und Spotify sind die persönlichen Empfehlungen an ihre Nutzer, die auf gesammelten und ausgewerteten Nutzerdaten basieren. Flatrate-Modelle für Bücher sollten daher ebenfalls diese Daten für Empfehlungen verwenden, um die Nutzer an sich zu binden.

3.3.4 Inwieweit können Daten die Intuition ersetzen?

Um sich der Antwort auf diese Frage zu nähern, wird zunächst dargelegt, wie sich verschiedene Branchenvertreter zu diesem Aspekt äußern.

⁴² Readfy bietet derzeit mehr als 60.000 E-Books (vgl. Readfy 2016b, S. 1), Skoobe hat mehr als 175.000 E-Books im Angebot (vgl. Skoobe 2017).

Bisher galt in der Verlagsbranche die Regel, dass ca. 10 bis 20 Prozent der veröffentlichten Bücher in einem Verlag den Profit erwirtschaften, um die restlichen 80 bis 90 Prozent mitzufinanzieren (vgl. Neary 2016). Da die Menge an Büchern, die sich nicht gut verkaufen, also sehr hoch ist, gibt es Stimmen aus der Verlagsbranche, die Daten für eine bessere Entscheidungsgrundlage halten als die Intuition eines Lektors, da damit die Prozentzahl an Profit abwerfenden Büchern gesteigert werden kann. Diese Meinung vertritt beispielsweise Dominique Raccah, Gründerin und Verlegerin des amerikanischen Verlags Sourcebooks, welcher bereits viel mit Daten arbeitet (vgl. Neary 2016). Vor allem die Reader Analytics liefern dank digitaler Bücher Daten, welche bisher gefehlt haben, nämlich Einsichten darin, wie Leser grundsätzlich oder auch spezifische Bücher lesen (vgl. Raccah, zitiert nach Neary 2016).

Auch Jellybooks-Gründer Andrew Rhomberg sieht in den nun verfügbaren Daten und Methoden wie Reader Analytics eine Chance für die Buchbranche, weil jetzt Leserdaten zur Verfügung stehen, die vorher nicht verfügbar waren, und damit vor allem die Marketingabteilungen in Verlagen unterstützt werden können, die nun die Möglichkeit haben, Bücher zielgenauer bei der Zielgruppe zu bewerben (vgl. Neary 2016). Darüber hinaus können Verlage beispielsweise auch erfahren, warum ein Buch, von dem sich der Verlag einen Erfolg versprochen hat, sich schlechter verkauft als erwartet, oder ob das erfolgreiche Buch eines Erstautors auch wirklich gelesen wurde (vgl. Rhomberg, zitiert nach Neary 2016).

Alan Rinzler, ein langjähriger Lektor, erklärt hingegen, dass Lektoren Bücher vor allem danach auswählen, ob sie persönlich von dem Buch begeistert sind (vgl. Rinzler, zitiert nach Neary 2016). Er ist der Meinung, dass gute, aber inhaltlich anspruchsvolle Bücher bei Testlesern für Reader Analytics durchfallen würden und dass Verleger oder auch Autoren sich nicht wirklich dafür interessieren, was Testleser von einem Buch halten (vgl. Rinzler, zitiert nach Neary 2016).

Zudem äußern auch diejenigen Verlagsakteure, die bereits Daten einsetzen, dass die Intuition nach wie vor eine Rolle spielen wird. So sagt Dominique Raccah von Sourcebooks, dass Verlagsmitarbeiter für Entscheidungen – beispielsweise zu Format, Preis und Erscheinungstermin eines Buches –, auch wenn sie grundsätzlich auch mit Hilfe von Daten getroffen werden, einen Kontext benötigen, für den der Instinkt wichtig ist (vgl. Raccah 2012, S. 101).

Auch bei den Autoren sind die Meinungen zu Leserdaten geteilt. Es gibt Autoren, die eine Chance in Leserdaten sehen. So sagt die Autorin Therese Walsh, dass Leserdaten wertvolles Feedback für einen Autoren bedeuten können und diese damit die Chance haben, ihre Art zu schreiben zu verbessern und damit die Leser besser zu erreichen (vgl. Walsh, zitiert nach Anderson 2015). Dies ist jedoch eine seltene Autorenmeinung. Viele Autoren geben an, dass ihnen an Daten nur die Verkaufszahlen und Rezensionen ihrer Bücher wichtig sind oder dass sie nicht schreiben könnten, wenn sie zu viel Wert auf Leserdaten legen würden (vgl. Anderson 2015).

Andere wiederum glauben, dass Autoren sogar Geld dafür bezahlen würden, um an die Leserdaten zu gelangen, die beispielsweise Amazon sammelt (vgl. Anderson 2015).

Zuletzt erörtert Alexander Siebert, Gründer von *retresco*⁴³ dass situative Entscheidungen idealerweise in einem Zusammenspiel aus Daten, Software und Mensch getroffen werden sollten, um Erfolge zu erzielen (vgl. Siebert 2017). Siebert bezeichnet dies als »Data Informed-Ansatz« (Siebert 2017), bei dem die Interessen der Zielgruppen bekannt sind und das Programm danach ausgerichtet werden kann. Dies könne heutzutage nicht mehr rein nach Bauchgefühl entschieden werden, da dies im Wettbewerb mit anderen Unternehmen, die bereits Daten verwenden, wahrscheinlich von Nachteil sein wird (vgl. Siebert 2017). Verlage müssen daher lernen, wie sie Daten einsetzen und die Informationen aus Daten im Tagesgeschäft umsetzen können (vgl. Siebert 2017).

Es hat sich gezeigt, dass es sowohl in der Buchbranche als auch bei Autoren verschiedene Meinungen dazu gibt, ob und inwieweit Daten die Intuition von Lektoren ersetzen können. Zusätzlich muss beleuchtet werden, warum die Intuition in Verlagen solch eine bedeutende Rolle spielt und was für Chancen der Einsatz von Daten auf der anderen Seite bietet. Lektoren in Verlagen entscheiden sich für oder gegen die Veröffentlichung eines Manuskripts, indem sie sich zunächst eine persönliche Meinung dazu bilden, ob das Manuskript »gut« ist, also eine gewisse sprachliche Qualität aufweist, einen interessanten, fesselnden Inhalt hat oder einen künstlerischen Anspruch erfüllt.⁴⁴ Weiterhin schätzt der Lektor anhand weiterer Faktoren wie beispielsweise derzeitiger Markttrends – meist basierend auf jahrelanger Erfahrung –, wie gut sich das Buch auf dem Markt behaupten kann. Dies ist wiederum nicht immer das ausschlaggebende Kriterium, da, wie bereits schon erwähnt, viele Bücher, die wenig Aussicht auf hohe Verkaufszahlen haben, trotzdem veröffentlicht werden, finanziert durch die Bestseller des Verlags. Dies ist relevant, weil hiermit die Vielfalt auf dem Buchmarkt sichergestellt ist, was der Rolle des Buches als Kulturgut gerecht wird. Gerade für die Sicherstellung der Vielfalt des Buchmarkts sind die Intuitionsentscheidungen also wichtig.

Der Einsatz von Daten verspricht zunächst einmal, den Prozentsatz der Bücher, die sich so gut verkaufen, dass sie den Verlagen Gewinne einbringen, zu erhöhen. Dies kommt daher, dass zum einen mit Reader Analytics von bereits fertigen Büchern, wie sie bei Jellybooks durchgeführt werden, Erkenntnisse gewonnen werden können, die im Marketing umgesetzt werden können. Zum anderen können mit Reader-Analytics-Plattformen wie Inkitt, wo zunächst E-Books durch Selfpublisher veröffentlicht werden, Manuskripte nach der Beliebtheit bei den Lesern akquiriert

43 *retresco* ist ein Software-Unternehmen für die Automatisierung contentgetriebener Geschäftsmodelle.

44 Dies gilt nur für Belletristik. Bei Fachbüchern gelten andere Regeln, hier wird mit Hilfe von Peer Reviews oder fachspezifischen Lektoren die Qualität und damit die Eignung zur Veröffentlichung festgestellt.

werden. Solche Plattformen könnten also eine zusätzliche Quelle der Manuskriptakquise für Verlage darstellen. Daten sind also vor allem für die Rolle des Buches als Wirtschaftsgut wichtig, da sie im Grunde mehr verkaufte Bücher versprechen.

Eine Lösung des Problems für die Branche wäre also die Annäherung an eine Koexistenz von Intuition und Daten als Entscheidungsgrundlage. Die Vorteile, die sich hieraus für die Verlage ergeben, liegen auf der Hand: Sie erwirtschaften mehr Gewinne mit den datenbasierten Büchern und haben somit mehr Geld für die kulturell wichtigen, aber weniger erfolversprechenden Bücher. Die Vielfalt auf dem Buchmarkt wäre also gewahrt und die Rollen des Buches als Kulturgut und als Wirtschaftsgut wären nicht gefährdet.

4. Synopse

In diesem Kapitel sollen die bisherigen Erkenntnisse zusammengefasst und diskutiert werden, wie insgesamt die Zukunftsprognose von Big Data in der Buchbranche aussieht.

Wie schon in Kapitel 2 dargelegt, bietet Big Data Unternehmen die Möglichkeit, die Kunden sowie den Markt besser zu verstehen und damit Vorhersagen treffen zu können, was wiederum dazu dienen kann, Produkte besser an die Kunden anzupassen und gezielteres Marketing zu betreiben. Es hat sich in dieser Arbeit gezeigt, dass dies auch in der Buchbranche möglich ist. Dies liegt hauptsächlich daran, dass es durch E-Books nun möglich ist, Leserdaten zu erhalten. Diese Daten zum Leserverhalten und zu Lesegewohnheiten können analysiert werden, womit wichtige Erkenntnisse über die Leser gewonnen werden, welche vorher nicht oder nur schwer, beispielsweise über Befragungen, zugänglich waren. Die Leserdaten ermöglichen es den Verlagen, ihre Kunden besser zu verstehen, und Vorhersagen darüber zu treffen, welche Bücher sich gut verkaufen werden. Da die Zielgruppe und ihre Gewohnheiten besser erfasst sind, können Bücher gezielter beworben werden, um die Verkaufszahlen zu steigern.

Die Buchbranche ist bisher allerdings noch zurückhaltend im Bezug auf die Anwendung von Leserdaten. Dies hat laut Lenz (2015, S. 8) mehrere Gründe: mangelnde Erfahrung, zu wenig Ressourcen für die Auswertung und mangelnden Zugriff. Zumindest die letzten beiden Punkte können durch die Zusammenarbeit mit den neuen und teilweise neu entstehenden Reader-Analytics-Unternehmen oder auch den E-Book-Flatrate-Anbietern, welche den Verlagen auf Wunsch ausgewertete Daten zur Verfügung stellen, gelöst werden. Diese Unternehmen erlauben einen Zugriff auf Leserdaten und es müssen weniger interne Ressourcen dafür aufgewendet werden. Die größte Herausforderung ist also die mangelnde Erfahrung. Dies ist durchaus eine Herausforderung von Big Data⁴⁵, denn die Integration von Big-Data-Konzepten in ein Unternehmen erfordert neue, innovative Geschäftsmodelle. Zudem muss im Unternehmen eine Bereitschaft zu strukturellen Veränderungen von Verhaltens- und Denkmustern vorhanden sein. Da die Buchbranche stark traditionelle, historisch gewachsene Strukturen hat⁴⁶, ist verständlich, warum solch eine Veränderung auf technologischer Ebene sich nur langsam durchsetzt, auch weil Big Data »a completely different way of doing business« (Dumbill 2012, zitiert nach Reid 2012) impliziert. Die Erfahrung kann allerdings nur gesammelt werden, wenn die Branche sich von den traditionellen Strukturen löst und Innovationen zulässt.

Obwohl die Buchbranche insgesamt noch wenig mit Big Data arbeitet, zeigen die schon existierenden oder entstehenden Unternehmen, die mit verschiedenen

45 wie bereits in Kapitel 2.2.5 ab Seite 14 erläutert.

46 So zeichnet die Buchbranche in Deutschland eine »triadische Vertriebskette« (Schrape 2011, S. 11) aus herstellendem Buchhandel, Zwischenbuchhandel und verbreitendem Buchhandel aus, welche historisch gewachsen ist (vgl. Schrape 2011, S. 11; vgl. Janello 2010, S. 57 f.).

Ansätzen die Leserdaten gezielt nutzen, dass die Reader Analytics großes Potenzial für die Buchbranche haben. Sollte sich Big Data langfristig als wichtiges Konzept in Unternehmen durchsetzen, werden vermutlich auch die Verlage vermehrt beginnen, sich damit zu beschäftigen. Insbesondere im Marketing könnten durch die Nutzung von Datenanalysen bessere Ergebnisse erzielt werden, da die Datenbasis bisher hauptsächlich aus wenigen Quellen stammt und nur wenige, meist strukturierte Daten bereitstellt, nämlich die Verkaufszahlen in den Buchhandlungen, teilweise angereichert mit Daten aus Leserbefragungen. Die neuen Datenquellen, die sich durch die Digitalisierung ergeben, können vielfältigere Daten bereitstellen. Dazu gehören neben den Reader Analytics beispielsweise auch Daten aus sozialen Netzwerken oder aus Social Reading, Daten also, welche das konventionelle CRM, das bereits häufig in Verlagen eingesetzt wird, verbessern können. Davon kann auch die Buchbranche früher oder später profitieren und sollte daher die Chance nutzen, damit erfolgreichere Bücher zu produzieren.

Die zunehmende Orientierung an den Nutzern ist ein Modell, das sich in der Musik- und Filmindustrie bereits durchsetzt und dort von Akteuren wie Spotify oder Netflix umgesetzt wird, welche die Nutzerdaten für ihre Angebote und Empfehlungen verwenden. Auch Verlage haben nun die Möglichkeit, ihr Programm mehr an den Lesern zu orientieren. Entsprechende Modelle gibt es bereits in Form von E-Book-Flatrate-Anbietern oder Selfpublishing-Plattformen. Mit effektiver Auswertung von Nutzerdaten können hier ebenfalls Einsichten in die Nutzerwünsche erlangt und damit die Nutzerzahlen gesteigert werden.

Im Hinblick auf die Diskussion in der Buchbranchenpresse ließ sich feststellen, dass es überwiegend positive Resonanz auf die Möglichkeiten, die Big Data bereits bietet, und hohe Erwartungen für die Zukunft von Big Data in der Buchbranche gibt. Sofern die deutsche Buchbranchenpresse über Big Data, beispielsweise über die Reader Analytics, berichtet, wird häufig das Potenzial, das darin für die Zukunft des Verlegens liegt, betont. Gegenüber der bisherigen Methode, Entscheidungen überwiegend auf Basis von Intuition und Verkaufszahlen zu entscheiden, sehen viele Branchenvertreter in Leserdaten entscheidende Vorteile, die der Buchbranche zugute kommen, da sie dadurch insgesamt mehr Umsatz generieren können und außerdem zusätzliche Akquise-Instrumente erhalten. Die Leserdaten haben somit einen hohen Wert für Verlage.

Inwieweit Leserdaten oder auch ein vermeintlicher Bestseller-Algorithmus am Ende die Inhalte von Büchern beeinflussen können und werden und ob dies vielleicht dazu führt, dass Bücher sich inhaltlich und stilistisch angleichen, kann zu diesem Zeitpunkt noch nicht vorausgesehen werden, da die Verlage momentan noch sehr wenig mit solchen Daten arbeiten. Ob Daten irgendwann die Strukturen und Wertschöpfungsstufen der Buchbranche stark verändern werden, hängt daher auch davon ab, inwieweit sie sich in der Branche etablieren. Da Big Data ein relativ neues Phänomen ist und die Buchbranche meist eher langsam auf Innovationen reagiert,

bleibt abzuwarten, wie die Verlage sich in den nächsten Jahren zum Einsatz von Daten positionieren werden.

Letztendlich ist hier auch die Rolle des Buches als Kulturgut wichtig. Auch wenn der Einsatz von Daten beispielsweise im Marketing unproblematisch ist – falls der Datenschutz beachtet wird –, würde der Einsatz von Daten dann ein Problem darstellen, wenn tatsächlich dadurch die große Vielfalt an veröffentlichten Büchern eingeschränkt würde. Sollte es in Zukunft tatsächlich eine Art Patentrezept für ein sicheres Bestseller-Buch geben, liegt es an den Verlagen, damit verantwortungsvoll umzugehen. Das Prinzip, dass ein Verlag den Umsatz von Bestsellern verwendet, um damit spezielle oder nischenhafte Bücher zu verlegen, sollte daher beibehalten werden. Hier würden die Daten dann sogar die Vielfalt unterstützen, weil mit ihnen mehr Umsatz generiert werden kann, der dann für das Verlegen von Büchern, die eine eher kleine Leserschaft ansprechen, zur Verfügung steht.

Die Zukunft von Big Data in der Buchbranche hängt also davon ab, inwieweit und wann die Buchbranche das Potenzial der nun zugänglichen Leserdaten nutzen wird und wie sich die Reader-Analytics- und die E-Book-Flatrate-Unternehmen am Markt behaupten werden. Außerdem bleibt abzuwarten, ob es wirklich möglich ist, Bestseller mit hoher Trefferquote vorherzusagen. Wie in Kapitel 3.3.2 dargestellt, gibt es zwei verschiedene Ansätze dafür, Bestseller vorherzusagen, nämlich durch Reader Analytics und durch Analyse des Inhalts mit Text Mining. Eine Zukunftsvision wäre es hier, dass Manuskripte mit einem Text-Mining-Programm, wie es in *The Bestseller Code* entwickelt wurde, analysiert werden und danach mit Leserdaten-Analyse die Resonanz der Leser eingeholt wird. Die Kombination dieser beiden Datenanalyse-Methoden könnte dabei hilfreich sein, erfolgreiche Bücher zu finden. Dies würde Verlagen mehr finanzielle Spielräume dafür geben, wie bisher mit den Bestsellern weniger erfolgreiche Bücher mitzufinanzieren. Die Folge wäre allerdings, dass der Lektor im Verlag nicht mehr der hauptsächliche Entscheider darüber ist, welches Buch veröffentlicht wird, sondern diese Entscheidung indirekt auch von den Lesern selbst getroffen wird. Das würde auch bedeuten, dass die Verlage ihre Gatekeeping-Funktion ein Stück weit verlieren. Allerdings würde im besten Fall der Verlag mehr erfolgreiche Bücher veröffentlichen und dadurch mehr Umsatz generieren. Er würde insgesamt mehr Leser gewinnen und damit auch seine Marke stärken. Das Kulturgut Buch würde weiter verbreitet. Zudem könnten neue Autoren leichter entdeckt und etabliert werden, was die Vielfalt des Buchmarkts bewahren würde.

Wichtig bei allen Einsatz-Szenarien von Big Data in der Buchbranche ist es, die Risiken von Big Data zu berücksichtigen. Das größte Risiko ist der Datenschutz, mit der Verwendung von Kundendaten im CRM oder von Leserdaten sollte also verantwortungsvoll umgegangen werden. Vorbild könnte hier beispielsweise das Vorgehen von Jellybooks sein, bei dem die Leser explizit zustimmen, dass ihre Leserdaten gesammelt und ausgewertet werden, und als Belohnung ein kostenloses E-Books

erhalten. Aber auch das Risiko von Scheinkorrelationen sollte bedacht werden. Insbesondere bei der Vorhersage von Bestsellern kann es passieren, dass Muster aufgedeckt werden, die eigentlich gar keine Relevanz dafür haben, ob ein Buch zum Bestseller werden kann oder nicht. Daher sollte immer auch, wie bereits in Kapitel 3.3.4 diskutiert, die Einschätzung von erfahrenen Lektoren wertgeschätzt werden. Zusätzlich könnte in Verlagen auch der Einsatz von Data Scientists sinnvoll sein, welche mit dem Lektorat zusammenarbeiten.

Zum Schluss soll noch erwähnt werden, dass sich das Potenzial von Leserdaten wahrscheinlich einfacher und besser nutzen ließe, wenn die Buchbranche hier zusammenarbeiten würde. So ist Meier (2015b) der Meinung, dass es in Zukunft ein Branchenprojekt im Hinblick auf die Gewinnung von Leserdaten geben sollte. Dies könnte eine Branchen-Plattform für E-Books sein, deren E-Books eine Lesesoftware enthält, welche Leserdaten sammeln und analysieren kann (vgl. Meier 2015b). Hinzu käme im besten Fall auch eine neutrale E-Commerce-Umgebung und die Möglichkeit, die E-Books eines Nutzeraccounts auf verschiedene Geräte zu übertragen, was für die Nutzer große Vorteile hätte und damit die Plattform erfolgreich machen könnte (vgl. Meier 2015b). Die Wahrscheinlichkeit einer Umsetzung solch eines Projekts schätzt Meier (2015b) aber als sehr gering ein. Dies würde aber tatsächlich die Buchbranche sowohl beim E-Book-Vertrieb als auch beim Zugang zu Leserdaten voranbringen. Wenn den Nutzern dann auch transparent dargelegt wird, dass ihre Leserdaten gesammelt werden und wofür diese genutzt werden, wären die Nutzer zumindest besser über die Verwendung ihrer Daten informiert, was Vertrauen schaffen kann, und die Buchbranche bekäme Leserdaten. Wie bereits erwähnt, hat Amazon, ein starker Konkurrent des Buchmarkts, bereits weitreichenden Zugriff auf Leserdaten, wohingegen Verlage solche Daten von Unternehmen wie Jellybooks kaufen müssen. Ein Branchenprojekt würde also die Daten auf einfache Weise auch den Verlagen zugänglich und damit nutzbar machen.

5. Schluss

Es konnte in der Arbeit gezeigt werden, dass der Einsatz von Big Data in der Buchbranche großes Potenzial hat und dieses überwiegend als gut bewertet werden kann, sofern sich Big-Data-Konzepte in Verlagen zukünftig etablieren können und die Risiken beachtet werden. Insbesondere die Reader Analytics bieten der Buchbranche viele Vorteile. Aber auch das CRM, das schon von vielen Verlagen genutzt wird, kann mit Big Data verbessert werden.

Aufgrund des begrenzten Umfangs und der buchwissenschaftlich fokussierten Perspektive dieser Arbeit konnten nicht alle wichtigen Aspekte abgedeckt werden. Wichtig wäre es, die hier gewonnenen Erkenntnisse durch weitere, differenzierende Untersuchungen zu ergänzen: Von großer Bedeutung wäre dabei die Perspektive der Informatik, um den Umfang der technologischen Herausforderungen für den Einsatz von Big Data in Verlagen genauer bewerten zu können. Auch eine differenziertere Untersuchung der Unterschiede zwischen großen, mittelständischen und kleinen Verlagen sowie zwischen Publikums- und Fachverlagen wäre in diesem Zusammenhang von Bedeutung, um einschätzen zu können, wie sinnvoll und realistisch ein Einsatz von Big Data für sie wäre, weil sich die wirtschaftlichen Möglichkeiten wie auch das jeweilige Selbstverständnis und die Ausrichtung von Verlagen unterscheiden. Da hier aber ein allgemeiner Überblick über das grundsätzliche Potenzial von Big Data für Verlage angesteuert wurde und eine Ausdifferenzierung aller Verlagsarten im Bezug auf Big Data zu umfangreich gewesen wäre, muss die Beantwortung dieser Fragen zukünftigen Untersuchungen vorbehalten bleiben.

Die Analyse des Status quo des Einsatzes von Daten in der Buchbranche und der Buchbranchenpresse hat gezeigt, dass die Buchbranche noch wenig mit Daten arbeitet und die Branchenpresse insgesamt das Thema noch wenig aufgegriffen hat, allerdings überwiegend positiv berichtet. Da das Phänomen Big Data noch recht neu ist und auch die E-Books, welche den Zugang zu Leserdaten ermöglichen, sich erst seit wenigen Jahren im Buchmarkt zu einer festen Größe etablieren, muss daher an dieser Stelle überwiegend in die Zukunft verwiesen werden. Es ist zu hoffen, dass sich die Zukunft der Buchbranche in dem Sinne positiv entwickelt, dass Verlage sich entscheiden, die neuen Datenquellen zu nutzen, um bessere Einsichten in ihre Leser zu bekommen und damit gute und erfolgreiche Bücher veröffentlichen und neue Autoren entdecken. Auch im Hinblick auf die große Konkurrenz zu einflussreichen Marktakteuren wie Amazon, bei denen Datenanalysen bereits zum Geschäftsmodell gehören, täte die Buchbranche gut daran, sich diesem Thema zu stellen. Eine besondere Herausforderung ist es dabei, die Vielfalt des Buchmarkts und damit die Eigenschaft des Buches als Kulturgut zu bewahren. Von besonderem Interesse wird die zukünftige Entwicklung von Plattformen und Unternehmen wie Oolipo oder Inkitt sein: Wie werden die Leser die Plattformen annehmen? Wird die Buchbranche mit ihnen zusammenarbeiten? Wie in der Arbeit dargelegt, erfordert dies ein massives

Umdenken und Umstrukturieren, was für die historisch lang bestehende und traditionelle Branche eine Herausforderung sein wird. Allerdings bieten sich aber letztendlich gute Chancen für die Branche, erfolgreicher und effektiver zu arbeiten.

Literaturverzeichnis

- Albazaz, Ali 2016a:** The Future of Publishing in a Data-Driven World.
<http://www.digitalbookworld.com/2016/future-publishing-data-driven-world/>
[08.09.2016 / 27.01.2017]
- Albazaz, Ali 2016b:** «Manche Verlage sind uns zu langsam». In: Buchreport Express 47 (2016), H. 42, S. 16.
- Alter, Alexandra / Russell, Karl 2016:** Moneyball for Book Publishers: A Detailed Look at How We Read.
<http://www.nytimes.com/2016/03/15/business/media/moneyball-for-book-publishers-for-a-detailed-look-at-how-we-read.html> [14.03.2016 / 30.12.2016]
- Anderson, Porter 2015:** Reader Analytics: Not All Authors Want To Know.
<http://thoughtcatalog.com/porter-anderson/2015/08/reader-analytics-not-all-authors-want-to-know/> [25.08.2015 / 30.12.2016]
- Archer, Jodie / Jockers, Matthew L. 2016:** The Bestseller Code. Anatomy of the Blockbuster Novel. New York 2016.
- Bachmann, Ronald / Kemper, Guido / Gerzer, Thomas 2014:** Big Data – Fluch oder Segen? Unternehmen im Spiegel gesellschaftlichen Wandels. Heidelberg, München, Landsberg, Frechen, Hamburg 2014.
- Balthaser, Neil 2016:** Yes, Machine Learning Can Help Predict a Bestseller.
<http://www.digitalbookworld.com/2016/machine-learning-help-predict-bestseller/> [20.07.2016 / 17.01.2017]
- Beck, Hanno 2011:** Medienökonomie. Print, Fernsehen und Multimedia. 3., überarbeitete und ergänzte Auflage. Heidelberg, Dordrecht, London, New York 2011.
- Beier, Heiko 2016a:** Smart ist smarter als Big.
http://www.pubiz.de/home/redaktionlektorat/redaktionlektorat_artikel/datum/2016/05/11/smart-ist-smarter-als-big.htm [11.05.2016 / 15.12.2016]
- Beier, Heiko 2016b:** Mit Smart Data gezielter planen, publizieren, werben und kommunizieren.
http://www.pubiz.de/home/redaktionlektorat/redaktionlektorat_artikel/datum/2016/06/06/mit-smart-data-gezielter-planen-publizieren-werben-und-kommunizieren.htm?no_cache=1&cHash=e6fa43901f8f5ba52b44b833ed67f6e1
[06.06.2016 / 15.12.2016]
- Beyer, Andrea / Carl, Petra 2008:** Einführung in die Medienökonomie. 2., überarbeitete und erweiterte Auflage. Konstanz 2008.

- Bendel, Oliver 2016:** Big Data. Gabler Wirtschaftslexikon.
<http://wirtschaftslexikon.gabler.de/Archiv/-2046774198/big-data-v3.html>
[28.09.2016]
- Bloching, Björn / Luck, Lars / Ramge, Thomas 2015:** Smart Data.
Datenstrategien, die Kunden wirklich wollen und Unternehmen wirklich nützen.
München 2015.
- Boellstorff, Tom 2014:** Die Konstruktion von Big Data in der Theorie. In: Reichert, Ramón (Hrsg.): Big Data. Analysen zum digitalen Wandel von Wissen, Macht und Ökonomie. Bielefeld 2014,
S. 105–131.
- Börsenverein des Deutschen Buchhandels o. J.:** Was versteht man unter Quersubventionierung?
<http://www.boersenverein.de/de/293151> [07.03.2017]
- Bramann, Klaus-W. 2014:** Bedeutung des Verlagswesens. In: Breyer-Mayländer, Thomas (Hrsg.): Wirtschaftsunternehmen Verlag. 5. aktualisierte und erweiterte Auflage. Frankfurt am Main 2014, S. 15–60.
- Bramann, Klaus-W. / Cremer, Peter 2014:** Wirtschaftsunternehmen Sortiment. Der Buchhandel in Deutschland zu Beginn des 21. Jahrhunderts. Frankfurt am Main 2014.
- Breyer-Mayländer, Thomas 2014:** Marketing und Absatzpolitik in Medienunternehmen. In: Breyer-Mayländer, Thomas (Hrsg.): Wirtschaftsunternehmen Verlag. 5. aktualisierte und erweiterte Auflage. Frankfurt am Main 2014, S. 127–184.
- Buhl, Hans Ulrich / Röglinger, Maximilian / Moser, Florian / Heidemann, Julia 2013:** Big Data. ein (ir)relevanter Modebegriff für Wissenschaft und Praxis? In: Wirtschaftsinformatik 55 (2013), H. 2, S. 63–68.
DOI: 10.1007/s11576-013-0350-x.
- Bundesministerium für Justiz und Verbraucherschutz 2016:** Gesetz über die Preisbindung für Bücher (Buchpreisbindungsgesetz).
<https://www.gesetze-im-internet.de/buchprg/BJNR344810002.html>
[31.07.2016 / 30.11.2016]
- Busse, Christina 2013:** Big Data: Überrollt die riesige Datenmenge die Fachverlage?
<http://www.boersenblatt.net/640358/> [11.10.2013 / 15.12.2016]
- Callari, Ron 2016:** Inkitt: Using Algorithms to Open Doors for Indie Authors to Become Bestsellers.
<https://www.bookworks.com/2016/05/algorithmic-publishing-by-inkitt-surfaces-best-sellers/> [20.05.2016 / 16.01.2017]

- Chen, Min / Mao, Shiwen / Zhang, Yin / Leung, Victor C. M. 2014:** Big Data. Related Technologies, Challenges and Future Prospects. Cham, Heidelberg, New York, Dordrecht, London 2014.
- Cheng, Shi / Liu, Bin / Shi, Yuhui / Jin, Yaochu / Li, Bin 2016:** Evolutionary Computation and Big Data: Key Challenges and Future Directions. In: Tan, Ying / Shi, Yuhui (Hrsg.): Data Mining and Big Data. First International Conference, DMBD 2016. Bali, Indonesia, June 25–30, 2016. Schweiz 2016, S. 3–14.
- Cordon, Carlos / Garcia-Milà, Pau / Ferreiro Vilarino, Teresa / Caballero, Pablo 2016:** Strategy is Digital. How Companies Can Use Big Data in the Value Chain. Schweiz 2016.
- Davenport, Thomas H. 2014a:** big data @ work. Chancen erkennen, Risiken verstehen. München 2014.
- Davenport, Thomas H. 2014b:** Book Publishing's Big Data Future.
<https://hbr.org/2014/03/book-publishings-big-data-future> [03.03.2014 / 27.01.2017]
- Davenport, Thomas H. 2014c:** »Daten sind effektiver als die Geschäftsessen-Kultur«. Interview geführt von Daniel Lenz. In: Buchreport Magazin 45 (2014), H. 11, S. 12–13.
- Deininger, Olaf 2005:** Der Kampf um jeden Kontakt.
<https://www.heise.de/tp/features/Der-Kampf-um-jeden-Kontakt-3403069.html> [07.10.2005 / 19.12.2016]
- Dieckmann, Randolph 2015:** Die Masse macht's nicht mehr.
https://www.boersenblatt.net/artikel-was_die_wissenschaft_ueber_buecherpreise_weiss.970546.html [23.06.2015 / 07.03.2017]
- Fairchild, Christa 2016:** Intellogo's Author Tools gives author Patti Larsen new editing ammunition for launch of book, Eve.
<https://medium.com/intellogo/award-winning-writer-patti-larsen-shares-her-experiences-using-intellogos-author-tools-c93fea28338#.m7873p7bm> [07.07.2016 / 09.03.2017]
- Fasel, Daniel 2014:** Big Data – Eine Einführung. In: HMD Praxis der Wirtschaftsinformatik 51 (2014), H. 4: Big Data, S. 386–400.
- Fasel, Daniel / Meier, Andreas 2016:** Was versteht man unter Big Data und NoSQL? In: Fasel, Daniel / Meier, Andreas (Hrsg.): Big Data. Grundlagen, Systeme und Nutzungspotenziale. Wiesbaden 2016, S. 3–16.
- Fiegermann, Seth 2016:** Netflix soars 20% as new subscribers stream in.
<http://money.cnn.com/2016/10/17/technology/netflix-earnings-subscribers/> [17.10.2016 / 24.10.2016]

- Floridi, Luciano 2015:** Die 4. Revolution. Wie die Infosphäre unser Leben verändert. Berlin 2015.
- Gartner 2016:** Big Data.
<http://www.gartner.com/it-glossary/big-data/> [23.09.2016]
- Gröger, Katharina 2016:** Der (un)bekannte Leser.
http://www.boersenblatt.net/artikel-reader_analytics.1121894.html [01.04.2016 / 03.01.2017]
- Güßgen, Florian 2016:** Sommerhit oder Flop? Besuch bei den Menschen, die wissen, was wir hören wollen.
http://www.stern.de/digital/homeentertainment/spotify--das-sind-die-menschen-hinter-der-playlist-7015364.html#mc-0_1477571034006 [21.08.2016 / 27.10.2016]
- Haberich, Ralf 2015:** Von der Webanalyse zur Digitalen Intelligenz. In: Schwarz, Torsten (Hrsg.): Big Data im Marketing. Chancen und Möglichkeiten für eine effektive Kundenansprache. Freiburg, München 2015, S. 66–81.
- Hauck, Stefan 2016:** »Es reicht!«. In: Börsenblatt 183 (2016), H. 4, S. 16–17.
- Haustein, Berthold H. 2015:** Datenschutz jenseits der Papierakte. Systematische Herausforderungen des Datenschutzrechts unter den Bedingungen der Digitalisierung. In: Süßenguth, Florian (Hrsg.): Die Gesellschaft der Daten. Über die digitale Transformation der sozialen Ordnung. Bielefeld 2015, S. 253–282.
- Heimann, Holger 2015:** Metadaten verkaufen mehr Bücher.
<http://www.boersenblatt.net/artikel-vlb-infotour.1073508.html> [30.12.2015 / 20.12.2016]
- Hippner, Hajo / Wilde, Klaus D. 2013:** Data Mining im CRM. In: Helmke, Stefan / Uebel, Matthias / Dangelmaier, Wilhelm (Hrsg.): Effektives Customer Relationship Management. Instrumente – Einführungskonzepte – Organisation. 5., überarbeitete Auflage. Wiesbaden 2013, S. 179–196.
- Hippner, Hajo / Grieser, Lukas / Wilde, Klaus D. 2011:** Data Mining – Grundlagen und Einsatzpotenziale in analytischen CRM-Prozessen. In: Hippner, Hajo / Hubrich, Beate / Wilde, Klaus D. (Hrsg.): Grundlagen des CRM. Strategie, Geschäftsprozesse und IT-Unterstützung. 3., vollständig überarbeitete und erweiterte Auflage. Wiesbaden 2011, S. 783–810.
- Hoenke, Marc 2016:** "Big Data bietet enorme Chancen".
https://www.boersenblatt.net/artikel-mvb_data_summit_interview_mit_dem_keynote_speaker.1243521.html [14.10.2016 / 15.12.2016]

- Hoffelder, Nate 2015:** JellyBooks Launches Google Analytics for eBooks at DBW 2015.
<http://the-digital-reader.com/2015/01/13/jellybooks-launches-google-analytics-ebooks-dbw-2015/> [13.01.2015 / 03.01.2017]
- Homburg, Christian / Bruhn, Manfred 2013:** Kundenbindungsmanagement – Eine Einführung in die theoretischen und praktischen Problemstellungen. In: Bruhn, Manfred / Homburg, Christian (Hrsg.): Handbuch Kundenbindungsmanagement. 8., überarbeitete und erweiterte Auflage. Wiesbaden 2013, S. 3–39.
- Hüsing, Alexander 2016:** readfy auf dem Weg zum »relevanten Flatrate-Anbieter«
<https://www.deutsche-startups.de/2016/02/04/unsere-user-lesen-seit-launch-jeden-monat-mehr/>, [04.02.2016 / 17.01.2017]
- Iafrate, Fernando 2015:** From Big Data to Smart Data. London, Hoboken 2015.
- Janello, Christoph 2010:** Wertschöpfung im digitalisierten Buchmarkt. Wiesbaden 2010.
- Jellybooks 2017:** So lesen wir. <https://jellybooks.de/> [07.03.2017]
- Kappes, Christoph 2016:** Was ist heute innovatives, wirksames Buch-Marketing?
http://www.pubiz.de/home/marketingwerbung/marketingwerbung_artikel/datum/2016/03/01/was-ist-heute-innovatives-wirksames-buch-marketing.htm
[01.03.2016 / 19.12.2016]
- Keuschnigg, Marc 2012:** Das Bestseller-Phänomen. Die Entstehung von Nachfragekonzentration im Buchmarkt. Wiesbaden 2012.
- King, Stefanie 2014:** Big Data. Potential und Barrieren der Nutzung im Unternehmenskontext. Wiesbaden 2014.
- Kirkbride, Jasmin 2016:** Observing the audience: how reader analytics are influencing the industry.
<https://bookmachine.org/2016/10/13/observing-audience-reader-analytics-influencing-industry/> [13.10.2016 / 03.01.2017]
- Klausnitzer, Rudolf 2013:** Das Ende des Zufalls. Wie Big Data uns und unser Leben vorhersagbar macht. Salzburg 2013.
- Kraus, Hans 2013:** Big Data – Einsatzfelder und Herausforderungen. Arbeitspapier der FOM Hochschule für Ökonomie & Management, No. 41. Essen 2013.
- Kreutzer, Ralf T. 2016:** Kundenbeziehungsmanagement im digitalen Zeitalter. Konzepte, Erfolgsfaktoren, Handlungsideen. Stuttgart 2016.
- Krimmer, Thomas 2015:** Liebe in Zeiten des Algorithmus. In: Buchreport Magazin 46 (2015), H. 6, S. 16.

- Leisegang, Daniel 2014:** Amazon. Das Buch als Beute. Stuttgart 2014.
- Lenz, Daniel 2012:** Wie Verlage mit Daten bessere Bücher machen. In: Buchreport Magazin 43 (2012), H. 10, S. 98–100.
- Lenz, Daniel 2015:** Mehr als Kribbeln im Bauch. In: Buchreport Magazin 46 (2015), H. 11, S. 8.
- Lenz, Susanne 2016:** Jellybooks: Britische Firma analysiert Leseverhalten für deutsche Verlage.
<http://www.berliner-zeitung.de/kultur/literatur/jellybooks-britische-firma-analysiert-leseverhalten-fuer-deutsche-verlage-23975804> [29.04.2016 / 16.01.2017]
- Leußner, Wolfgang / Hippner, Hajo / Wilde, Klaus D. 2011:** CRM – Grundlagen, Konzepte und Prozesse. In: Hippner, Hajo / Hubrich, Beate / Wilde, Klaus D. (Hrsg.): Grundlagen des CRM. Strategie, Geschäftsprozesse und IT-Unterstützung. 3., vollständig überarbeitete und erweiterte Auflage. Wiesbaden 2011, S. 15–55.
- Lichtenberg, James 2014:** Big Data: Should Publishers Even Care?
<http://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/62114-big-data-should-publishers-even-care.html> [02.05.2014 / 21.09.2016]
- Lindner, Roland 2017:** Netflix schafft neuen Rekord.
<http://www.faz.net/aktuell/finanzen/netflix-aktie-explodiert-dank-steigender-nutzerzahlen-14695317.html> [19.01.2017 / 26.02.2017]
- Lindstrom, Martin 2016:** »Big Data ist ein Ausdruck von Unsicherheit«. Interview geführt von Ingmar Höhmann. In: Buchreport Magazin 47 (2016), H. 6, S. 32–35.
- LinkedIn 2017:** Jellybooks. <https://www.linkedin.com/company/jellybooks> [03.01.2017]
- Lohr, Steve 2012:** How Big Data Became So Big.
<http://www.nytimes.com/2012/08/12/business/how-big-data-became-so-big-unboxed.html> [11.08.2012 / 19.09.2016]
- Lomas, Natasha 2016:** Blendle clocks up 1M signups for its pay-per-article journalism platform.
<https://techcrunch.com/2016/08/11/blendle-clocks-up-1m-signups-for-its-pay-per-article-journalism-platform/?ncid=rss> [11.08.2016 / 31.01.2017]
- Mainzer, Klaus 2014:** Die Berechnung der Welt. Von der Weltformel zu Big Data. München 2014.
- Marr, Bernard 2016:** Big Data in practice. How 45 successful companies used big data analytics to deliver extraordinary results. Chichester 2016.

- Mayer-Schönberger, Viktor 2014:** Angst vor Big Data ist kein gutes Konzept.
<https://www.buchreport.de/2014/09/15/angst-vor-big-data-ist-kein-gutes-rezept/> [15.09.2014 / 15.12.2016]
- Meadows, Chris 2016:** Do Jellybooks e-book reader statistics actually mean anything?
<http://teleread.com/jellybooks-e-book-reader-statistics-actually-mean-anything/> [17.03.2016 / 16.01.2017]
- media control GmbH 2015:** Buch. Alle Buchmarkt-Informationen aus einer Hand.
<http://www.media-control.de/buch.html> [2015 / 20.12.2016]
- Meier, Steffen 2015a:** Heiliger Gral oder Schierlingsbecher?
http://www.boersenblatt.net/bookbytes/artikel-reader_analytics_teil_1.1058694.html [04.12.2015 / 15.12.2016]
- Meier, Steffen 2015b:** Am Anfang war die Datenbasis...
http://www.boersenblatt.net/bookbytes/artikel-reader_analytics_teil_2.1058701.html [07.12.2015 / 15.12.2016]
- Meier, Steffen 2016:** Do Androids Dream of Electric Literature?
http://www.boersenblatt.net/bookbytes/artikel-data_driven_publishing.1247483.html [08.11.2016 / 15.12.2016]
- Meyer-Dohm, Peter 1974:** Der Buchhandel in der Gesamtwirtschaft. In: Machill, Horst (Hrsg.): Handbuch des Buchhandels, Band 1, Allgemeines. In: Meyer-Dohm, Peter / Strauß, Wolfgang (Hrsg.): Handbuch des Buchhandels in vier Bänden. Hamburg 1974, S. 432–480.
- MVB Marketing- und Verlagsservice des Buchhandels GmbH o. J.:** VLB+. Das VLB auf dem Weg zur Metadatenbank der Branche. Frankfurt am Main. Abrufbar unter <http://info.vlb.de/home/vlb-plus.html>
- MVB Marketing- und Verlagsservice des Buchhandels GmbH 2016:** Wo Bücher sind, ist das VLB. Frankfurt am Main. Abrufbar unter: http://flashbooks.boersenblatt.net/VLB_2016/html5.html
- Nawotka, Edward 2015:** Ebook Platforms Know What, When and How You Read.
<http://publishingperspectives.com/2015/01/ebook-platforms-know-read/#.V7nCw6LAA-J> [20.01.2015 / 16.01.2017]
- Neary, Lynn 2016:** Publishers' Dilemma: Judge A Book By Its Data Or Trust The Editor's Gut?
<http://www.npr.org/sections/alltechconsidered/2016/08/02/488382297/publishers-dilemma-judge-a-book-by-its-data-or-trust-the-editors-gut> [02.08.2016 / 21.09.2016]

Netflix 2016: Datenschutzerklärung.

<https://help.netflix.com/legal/privacy?locale=de&docType=privacy>
[06.01.2016 / 24.10.2016]

o. V. 2012: Datenkrake E-Reader. In: Börsenblatt 179 (2012), H. 49, S. 10.

o. V. 2013: Digitaler Spion. In: Börsenblatt 180 (2013), H. 16, S. 6.

o. V. 2015: Digitalisierung? Ja, aber...

http://www.buchreport.de/nachrichten/verlage/verlage_nachricht/datum/2015/10/21/digitalisierung-ja-aber.htm?no_cache=1 [21.10.2015 / 15.12.2016]

o. V. 2016a: 4 Mio fürs Streamingportal.

<https://www.buchreport.de/2016/04/28/4-mio-fuers-streamingportal/>
[28.04.2016 / 17.01.2017]

o. V. 2016b: Data-Driven Publisher Inkitt Signs First Predicted Bestseller with Tor Books.

<http://www.digitalbookworld.com/2016/data-driven-publisher-inkitt-signs-first-predicted-bestseller-with-tor-books/> [28.04.2016 / 17.01.2017]

o. V. 2016c: Erster MVB Data Summit. In: Börsenblatt 183 (2016), H. 28, S. 14.

o. V. 2016d: Readfy: Streaming (nicht nur) zum Nulltarif. In: Börsenblatt 183 (2016), H. 11, S. 35.

o. V. 2016e: Skoobe: der E-Book-Flatrate-Dienst. In: Börsenblatt 183 (2016), H. 11, S. 35.

o. V. 2016f: Skoobe im Test: Meine Erfahrungen mit der Ebook-Flat.

<https://www.smartphonepiloten.de/digital-lifestyle/skoobe-test-erfahrungen-mit-der-ebook-flatrate>, [30.06.2016 / 17.01.2017]

o. V. 2016g: So funktioniert Oolipo.

<https://www.buchreport.de/2016/09/30/so-funktioniert-oolipo/> [30.09.2016 / 17.01.2017]

o. V. 2016h: Wissen macht was. Data Summit der MVB 2016.

https://www.boersenblatt.net/artikel-data_summit_der_mvb_2016__1253296.html [22.11.2016 / 15.12.2016]

o. V. 2017: Digitalisierung, Metadaten, Big Data.

https://www.boersenblatt.net/artikel-future_publish_startet_morgen_in_berlin.1281368.html [25.01.2017 / 26.01.2017]

Ortega, Ilias 2015: Streaming Analytics: Management in Echtzeit. In: Schwarz, Torsten (Hrsg.): Big Data im Marketing. Chancen und Möglichkeiten für eine effektive Kundenansprache. Freiburg, München 2015, S. 55–65.

- Piller, Gunther 2014:** Innovative Unternehmensanwendungen durch Echtzeitanalysen von Massendaten. In: Keuper, Frank / Schmidt, Dietmar / Schomann, Marc (Hrsg.): Smart Big Data Management. Berlin 2014, S. 83–98.
- Pleimling, Dominique 2012:** Social Reading – Lesen im digitalen Zeitalter.
<http://www.bpb.de/apuz/145378/social-reading-lesen-im-digitalen-zeitalter?p=all> [02.10.2012 / 25.02.2017]
- Ploeg, Frederick van der 2004:** Beyond the Dogma of the Fixed Book Price Agreement. In: Journal of Cultural Economics 28 (2004), H. 1, S. 1–20.
DOI: 10.1023/B:JCEC.0000009817.61107.f3
- Posth, Sebastian 2013:** Mit Daten statt Intuition entscheiden.
<https://www.buchreport.de/2013/02/21/mit-daten-statt-intuition-entscheiden/> [21.02.2013/ 15.12.2016]
- Raccah, Dominique 2012:** Bücher schneller und effizienter erstellen.
In: Buchreport Magazin 43 (2012), H. 10, S. 101.
- Rautenberg, Ursula / Wetzel, Dirk 2001:** Buch. Tübingen 2001.
- Readfy 2016a:** readfy »Reader Analytics« bietet neue Dimensionen der Datentiefe: Seitengenaue Analyse von Leserschaft und Leseverhalten. Solingen, 10.10.2016.
verfügbar unter: https://www.readfy.com/media/filer_public/59/c4/59c4fe6d-dfa1-49d5-9b34-629a485ce630/pm_readfy_reader_analytics_final.pdf.
- Readfy 2016b:** So liest Deutschland! Readfy in Zahlen – Lesegewohnheiten in Zahlen. Solingen, 30.10.2016.
verfügbar unter: https://www.readfy.com/media/filer_public/40/d1/40d1777c-2432-4d13-8a92-495745142f54/pm_readfy_infografik.pdf
- Reid, Calvin 2012:** Can »Big Data« Fix Book Marketing?
<http://www.publishersweekly.com/pw/by-topic/digital/retailing/article/52119-can-b-g-data-fix-book-marketing.html> [28.05.2012 / 21.09.2016]
- Rentzmann, René / Hippner, Hajo / Hesse, Frank / Wilde, Klaus D. 2011:** IT-Unterstützung durch CRM-Systeme. In: Hippner, Hajo / Hubrich, Beate / Wilde, Klaus D. (Hrsg.): Grundlagen des CRM. Strategie, Geschäftsprozesse und IT-Unterstützung. 3., vollständig überarbeitete und erweiterte Auflage. Wiesbaden 2011, S. 129–155.
- Rhomberg, Andrew 2015:** Jellybooks: Tracking Reader Engagement for Better Marketing.
<http://publishingperspectives.com/2015/08/jellybooks-tracking-reader-engagement-for-better-marketing/#.V7cP46LAA-I> [26.08.2015 / 05.01.2017]

- Rhomberg, Andrew 2016a:** »Belletristische Inhalte nicht anrühren«.
<http://www.buchreport.de/2016/09/20/belletristische-inhalte-nicht-anruehren/>
[17.02.2016 / 03.02.2017]
- Rhomberg, Andrew 2016b:** Data Vs. Instinct – The Publisher’s Dilemma.
<http://www.digitalbookworld.com/2016/data-vs-instinct-the-publishers-dilemma/> [17.03.2016 / 16.01.2017]
- Rhomberg, Andrew 2016c:** Data-Smart Publishing: New Ways of Working for Publishers.
<http://www.digitalbookworld.com/2016/data-smart-publishing-new-ways-of-working-for-publishers/> [01.03.2016 / 21.09.2016].
- Rhomberg, Andrew 2016d:** It's the Cover, Stupid! How book covers sometimes let both the reader and the author down.
<https://medium.com/@arhomberg/it-s-the-cover-stupid-49bbe08470de#.3uv69ogj2> [29.03.2016 / 16.01.2017]
- Rhomberg, Andrew 2016e:** What Books Have the X-Factor? Measuring a Book's Net Promoter Score [11.02.2016 / 16.01.2017]
- Richters, Kim 2015:** Geld von Christian Vollmann, Werbung mit Paulo Coelho – Inkitt will es wissen.
<http://www.gruenderszene.de/allgemein/inkitt-buch-verlag> [04.08.2015 / 17.01.2017]
- Rijmenam, Mark van 2016:** How Big Data Enabled Spotify To Change The Music Industry.
<https://datafloq.com/read/big-data-enabled-spotify-change-music-industry/391>
[30.08.2016 / 27.10.2016]
- Rimscha, Björn von/ Siegert, Gabriele 2015:** Medienökonomie. Eine problemorientierte Einführung. Wiesbaden 2015.
- Roesler-Graichen, Michael 2015:** Viele Verlierer. In: Börsenblatt 182 (2015), H. 42, S. 10–13.
- Sachs, Silas 2012:** Skoobe: Spotify für Bücher.
<http://www.gruenderszene.de/allgemein/skoobe-e-books> [16.11.2012 / 17.01.2017]
- Scherer, Lena 2015:** Reader Analytics: So tickt das Publikum. In: Buchreport Magazin 46 (2015), H. 11, S. 16–19.
- Scheuing, Sachiko 2015:** Offensive im Datenschutz. In: Schwarz, Torsten (Hrsg.): Big Data im Marketing. Chancen und Möglichkeiten für eine effektive Kundenansprache. Freiburg, München 2015, S. 115–125.

- Schlüter, Jan 2016:** Netflix, Spotify und Co: Die Vermessung des digitalen Ich.
<http://www.quotenmeter.de/n/85290/netflix-spotify-und-co-die-vermessung-des-digitalen-ich> [30.04.2016 / 27.10.2016]
- Schrabe, Jan-Felix 2011:** Gutenberg-Galaxis reloaded? Der Wandel des deutschen Buchhandels durch Internet, E-Books und Mobile Devices. Boizenburg 2011.
- Schmidt, Sabine 2013:** Ohne Ende Daten – aber was bedeuten sie? In: Börsenblatt 180 (2013), H. 52, S. 32–33.
- Schön, Dietmar 2016:** Planung und Reporting. Grundlagen, Business Intelligence, Mobile BI und Big-Data-Analytics. 2., überarbeitete Auflage. Wiesbaden 2016.
- Seemann, Michael 2014:** Gretchenfrage Big Data. In: Kappes, Christoph / Krone, Jan / Novy, Leonard (Hrsg.): Medienwandel kompakt 2011–2013. Netzveröffentlichungen zu Medienökonomie, Medienpolitik & Journalismus. Wiesbaden 2014, S. 117–120.
- Selle, Stephan 2016:** »Big Data« werden Grenzen gesetzt.
http://www.boersenblatt.net/bookbytes/artikel-neue_eu-datenschutzrichtlinien.1232559.html [29.08.2016 / 15.12.2016]
- Shaffi, Sarah 2016:** Startup of the week: Inkitt.
<http://www.thebookseller.com/futurebook/startup-week-inkitt-324342>
[15.03.2016/ 16.01.2017]
- Shanley, Mia / Weber, Jonathan 2016:** Spotify says growth has quickened since Apple Music's launch.
<http://www.reuters.com/article/us-spotify-future-idUSKCN0Y023B>
[09.05.2016 / 26.02.2017]
- Shatzkin, Mike 2016:** Full text examination by computer is very unlikely to predict bestsellers.
<http://www.idealogue.com/blog/full-text-examination-does-not-predict-bestsellers/>
[27.06.2016 / 17.01.2017]
- Siebert, Alexander 2017:** »Für einen Data Informed-Ansatz in Verlagen ist ein Kulturwandel erforderlich«.
http://www.pubiz.de/home/management/management_artikel/datum/////fuer-einen-data-informed-ansatz-in-verlagen-ist-ein-kulturwandel-erforderlich.htm
[23.01.2017 / 06.02.2017]
- Sim, Chris 2016:** Machine Learning and Bestseller Prediction: More Than Words Can Say.
<http://www.digitalbookworld.com/2016/machine-learning-and-bestseller-prediction-more-than-words-can-say/> [28.09.2016 / 07.02.2017]
- Skoobe 2017:** Über Skoobe. <https://www.skoobe.de/about-us> [26.02.2017]

- Spiller, Ulrich 2015:** Den Kunde im Fokus – CRM für Buchverlage. In: Buchreport Magazin 46 (2015), H. 6, S. 12–18.
- Statista 2016 a:** Anzahl der aktiven Spotify-Nutzer weltweit von Juli 2012 bis Juni 2016 (in Millionen).
<https://de.statista.com/statistik/daten/studie/368928/umfrage/monatlich-aktive-nutzer-von-spotify-weltweit/> [27.10.2016]
- Statista 2016 b:** Anzahl der Paid Subscribers von Spotify von März 2011 bis September 2016 (in Millionen).
<https://de.statista.com/statistik/daten/studie/297138/umfrage/anzahl-der-zahlenden-abonnenten-von-spotify/> [27.10.2016]
- Taylor-Sakyi, Kevin 2016:** Big Data: Understanding Big Data. Birmingham 2016.
ArXiv:1601.04602
- Thomas, Rob / McSharry, Patrick E. 2015:** Big Data revolution. What Farmers, Doctors and Insurance Agents teach us about discovering Big Data Patterns. Chichester 2015.
- Titel, Volker 2007:** Chance und Gefahr? Perspektiven der Digitalisierung für die Buchbranche. In: Keiderling, Thomas (Hrsg.): Buch - Markt - Theorie. Erlangen 2007, S. 197–224.
- Verhoef, Peter C. / Kooge, Erwin / Walk, Natasha 2016:** Creating Value with Big Data Analytics. Making smarter marketing decisions. London, New York 2016.
- Vossen, Gottfried 2015:** Big Data: Daten sammeln, aggregieren, analysieren, nutzen. In: Schwarz, Torsten (Hrsg.): Big Data im Marketing. Chancen und Möglichkeiten für eine effektive Kundenansprache. Freiburg, München 2015. S. 35–54.
- Weichert, Thilo 2013:** Big Data – eine Herausforderung für den Datenschutz. In: Geiselberger, Heinrich / Moorstedt, Tobias (Hrsg.): Big Data. das neue Versprechen der Allwissenheit. Berlin 2013, S. 131–148.
- Weise, Tamara 2016:** Kleine Schritte mit großer Wirkung. In: Börsenblatt 183 (2016), H. 4, S. 8–11.
- Weuster, Anna 2007:** Die Neuregelung der Buchpreisbindung in Deutschland. Stuttgart, München, Hannover, Berlin, Weimar, Dresden 2007.
- Zeckert, Patricia F. 2007:** Ein alter Unbekannter: Das Phänomen des ungelesenen Buches. In: Keiderling, Thomas (Hrsg.): Buch - Markt - Theorie. Erlangen 2007, S. 153–179.
- Zeit Online 2015:** Spotify holt sich mehr Nutzerdaten.
<http://www.zeit.de/digital/2015-08/spotify-datenschutz-personalisierung-zugang-persoenele-daten> [21.08.2015 / 16.11.2016]