

# MASTERARBEIT

Möglichkeiten einer Big Data Analyse durch Digital Analytics für klein- und mittelständische Unternehmen im Bereich E-Commerce

ausgeführt am



Studiengang  
Informationstechnologien und Wirtschaftsinformatik

Von: Thomas Fink  
Personenkennzeichen: 1510320029

Graz, am 2. Juli 2017

.....  
Unterschrift

## **EHRENWÖRTLICHE ERKLÄRUNG**

Ich erkläre ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen nicht benützt und die benutzten Quellen wörtlich zitiert sowie inhaltlich entnommene Stellen als solche kenntlich gemacht habe.

.....  
Unterschrift

## **DANKSAGUNG**

Danken möchte ich meinem Betreuer Dr. Thomas Puchleitner für die gute und unkomplizierte Zusammenarbeit im Erstellungsprozess. Danken möchte ich meinen Eltern, die mich auf allen Ebenen unterstützten und mir somit auch das Studieren ermöglicht haben. Danken möchte ich meiner Freundin Bettina, ohne die ich diese Zeilen wohl nicht schreiben könnte. Durch ihre wertvollen Ratschläge in unzähligen Gesprächen konnte nicht nur diese Masterarbeit abgeschlossen werden, sondern auch viele schwierige Situationen auf dem Weg dorthin gemeistert werden.

## KURZFASSUNG

In einer Zeit, in der die Digitalisierung von Geschäftsprozessen immer weiter voranschreitet und der digitale Wettbewerb dadurch kontinuierlich intensiver wird, müssen Unternehmen mehr denn je um ihren Platz darin kämpfen. Die aggressive Preispolitik und eine „die Großen fressen die Kleinen“-Mentalität der Internetgiganten wie Amazon und Alibaba lassen das E-Commerce immer mehr zu einer „Two-Man-Show“ verkommen. Um dem entgegenzuwirken, müssen sich die sogenannten „Kleinen“ auf jene Stärken konzentrieren, die sie „groß“ werden ließen: Qualität und KundInnennähe. Während die Qualität eine Selbstverständlichkeit darstellen sollte, wirft die KundInnennähe im E-Commerce eine gewichtige Frage auf: Wie soll man den BesucherInnen eines Webshops ein Käuferlebnis bieten, das in den Filialen von Angesicht zu Angesicht doch so selbstverständlich ist? Und genau an dieser Stelle hakt diese Masterarbeit ein. Um BesucherInnen ein Erlebnis im Online Shop zu bieten, muss man ihre Bedürfnisse kennen. Diese Bedürfnisse zu sammeln, zu speichern und auszuwerten wird unter dem Begriff Digital Analytics zusammengefasst. Digital Analytics bedeutet auf den Einzelfall heruntergebrochen, dass das Verhalten einer Person, die einen Online Shop betritt, beobachtet wird, damit die darauffolgenden Besuche an diese Beobachtungen angepasst werden können. Aus Beobachtungen werden Metriken und aus diesen Metriken können in weiterer Folge Kennzahlen zur kontinuierlichen Verbesserung abgeleitet werden. Tools zur Bewältigung dieser Aufgaben gibt es bereits, doch sind diese auch auf die Bedürfnisse von klein- und mittelständischen Unternehmen ausgerichtet und wenn ja, welche sind das? Ein Handlungsleitfaden für Klein- und Mittelunternehmen im Bereich E-Commerce, der mit Hilfe von ExpertInnen erstellt wird, soll diese Frage beantworten. Neben einer Gegenüberstellung von potenziellen Digital Analytics Lösungen beinhaltet der Handlungsleitfaden Informationen über die Identifizierung von relevanten Kennzahlen und Funktionen sowie einen Überblick über datenschutzrechtliche Aspekte.

## **ABSTRACT**

At a time when digitalization of business processes is advancing and digital competition is growing steadily, participants need more than ever to fight for their place in order not to be wiped out. Aggressive price policy and a “driven by the big ones” mentality from internet giants like Amazon and Alibaba threaten to turn E-Commerce into a “two-man show”. To counteract this, the so-called “little ones” must concentrate on the strengths which made them “big”: quality and customer proximity. While the quality should be a matter of course, the customer proximity rises thereby following question: “How can I offer online shop visitors a shopping experience that is as naturally as in usual stores?”. And this is precisely where this master thesis ties on. In order to offer visitors a tailored experience in the online shop, one need to know their desires. To explore these desires, it is necessary to collect, store and evaluate their needs during the visit. These actions are summarized under the term Digital Analytics. Digital Analytics means the evaluation of the behavior of a person who has just entered an online shop to draw conclusions regarding the subsequent visits to these observations. Metrics are derived from observations, and further measures can be derived from these metrics for continuous improvement. There are already tools for coping with these tasks, but are these tools also geared to the needs of small and medium-sized enterprises, and if so, what are they? A briefing guide for small and medium-sized enterprises in the area of e-commerce, created with the help of experts, is intended to answer this question. In addition to a comparison of potential digital analytics solutions, the guide contains information on the identification of relevant KPIs and functions, as well as an overview of data protection aspects.

# INHALTSVERZEICHNIS

|   |            |
|---|------------|
| <b>EHRENWÖRTLICHE ERKLÄRUNG</b> .....                           | <b>I</b>   |
| <b>DANKSAGUNG</b> .....   | <b>II</b>  |
| <b>KURZFASSUNG</b> .....  | <b>III</b> |
| <b>ABSTRACT</b> .....   | <b>IV</b>  |
| <b>INHALTSVERZEICHNIS</b> .....                                 | <b>V</b>   |
| <b>1 EINLEITUNG</b> .....                                       | <b>8</b>   |
| 1.1 Zielsetzung .....   | 9          |
| 1.2 Aufbau der Arbeit.....                                      | 9          |
| 1.3 Forschungsfrage.....  | 10         |
| <b>2 THEORETISCHER TEIL</b> .....                               | <b>11</b>  |
| 2.1 Grundlagen E-Commerce.....                                  | 11         |
| 2.1.1 Entwicklung zur Informationsgesellschaft .....            | 11         |
| 2.1.2 E-Commerce als Teil von E-Business .....                  | 13         |
| 2.1.2.1. Definition E-Business .....                            | 13         |
| 2.1.2.2. Definition E-Commerce .....                            | 13         |
| 2.1.3 AkteurInnen und Interaktionsmuster .....                  | 13         |
| 2.1.4 Marktzahlen und Prognosen.....                            | 14         |
| 2.2 Grundlagen Digital Analytics .....                          | 15         |
| 2.2.1 Einführung und Begriffsdefinition Digital Analytics ..... | 15         |
| 2.2.2 Digital Analytics Methoden .....                          | 16         |
| 2.2.3 Funktionsweise von Analytics-Systemen .....               | 19         |
| 2.2.3.1. Serverseitige Datensammlung mittels Logfile .....      | 20         |
| 2.2.3.2. Clientseitige Datensammlung .....                      | 23         |
| 2.2.3.3. Alternative Datensammlungsmethoden .....               | 24         |
| 2.2.4 Metriken in Digital Analytics.....                        | 25         |
| 2.2.5 Datenschutz und Digital Analytics .....                   | 27         |
| 2.3 Grundlagen Big Data .....                                   | 28         |
| 2.3.1 Definition und Eingrenzung .....                          | 29         |
| 2.3.2 Potenziale und Einsatzgebiete .....                       | 30         |
| 2.3.3 Marktsituation von Big Data.....                          | 31         |

|          |   |           |
|----------|---|-----------|
| 2.3.4    | Datenschutz und Risiken .....                                       | 35        |
| 2.4      | Big Data Analytik .....   | 37        |
| 2.4.1    | Einordnung .....  | 37        |
| 2.4.2    | Transaktionale Systeme .....  | 38        |
| 2.4.3    | Analytische Systeme: Data Warehouse und Business Intelligence ..... | 38        |
| 2.4.4    | Dokumentenmanagement .....  | 39        |
| 2.4.5    | Einflüsse auf die Software Entwicklung .....                        | 40        |
| 2.4.6    | Einflüsse auf die Anwendungsarchitektur .....                       | 41        |
| 2.4.7    | Big Data Bausteine .....  | 42        |
| 2.4.8    | NoSQL (Not only SQL) .....  | 47        |
| 2.4.9    | Apache Hadoop .....   | 48        |
| 2.4.9.1  | Szenarien und Anwendungsbeispiele für Hadoop .....                  | 51        |
| 2.4.10   | Big Data as a Service .....   | 51        |
| 2.5      | Klein- und Mittelunternehmen .....                                  | 52        |
| 2.5.1    | Chancen für Big Data in Klein- und Mittelunternehmen .....          | 53        |
| 2.5.2    | Effizientere Unternehmensführung .....                              | 54        |
| 2.5.3    | Massenindividualisierung durch Big Data .....                       | 54        |
| 2.5.4    | Big Ideas + Big Data = Big Business .....                           | 54        |
| <b>3</b> | <b>UNTERSUCHUNGSDESIGN .....</b>                                    | <b>55</b> |
| 3.1      | ExpertInneninterviews .....   | 55        |
| 3.2      | ExpertInnenstatus .....   | 57        |
| 3.3      | Vorbereitung eines leitfadengestützten ExpertInneninterviews .....  | 58        |
| 3.4      | Interviewleitfaden .....  | 60        |
| 3.5      | Durchführung der Interviews .....                                   | 62        |
| 3.6      | Auswertung der Interviews .....                                     | 62        |
| 3.7      | ExpertInnen .....   | 63        |
| <b>4</b> | <b>HANDLUNGSLEITFADEN .....</b>                                     | <b>65</b> |
| 4.1      | Themenblock 1: Basisanforderungen .....                             | 66        |
| 4.1.1    | Zielfindungsprozess im Unternehmen .....                            | 66        |
| 4.1.2    | Objectives and Key Results Modell .....                             | 68        |
| 4.2      | Themenblock 2: Identifizierung von KPIs .....                       | 69        |
| 4.3      | Themenblock 3: Funktionales Design .....                            | 71        |
| 4.3.1    | Einführung Google Analytics .....                                   | 72        |

|  |           |
|--|-----------|
| 4.3.1.1. Installation von Google Analytics.....                  | 73        |
| 4.3.1.2. Datenschutz in Google Analytics.....                    | 76        |
| 4.3.1.3. Pro und Kontra Google Analytics .....                   | 79        |
| 4.3.2 Einführung in Piwik.....                                   | 79        |
| 4.3.2.1. Installation von Piwik .....                            | 79        |
| 4.3.2.2. Datenschutz in Piwik .....                              | 83        |
| 4.3.2.3. Pro und Kontra Piwik.....                               | 84        |
| 4.4 Themenblock 4: Datenschutzrechtliche Aspekte .....           | 85        |
| 4.5 Themenblock 5: Zukunftsaussichten von Digital Analytics..... | 86        |
| <b>5 RESÜMEE .....</b>   | <b>88</b> |
| <b>ABKÜRZUNGSVERZEICHNIS.....</b>                                | <b>90</b> |
| <b>ABBILDUNGSVERZEICHNIS .....</b>                               | <b>91</b> |
| <b>TABELLENVERZEICHNIS .....</b>                                 | <b>93</b> |
| <b>LISTINGVERZEICHNIS .....</b>                                  | <b>94</b> |
| <b>LITERATURVERZEICHNIS.....</b>                                 | <b>95</b> |

# 1 EINLEITUNG

Mit dem Beginn der „Kommerzialisierung“ des Internets ab Mitte der 90er-Jahre des 20. Jahrhunderts wurde gleichzeitig auch der Grundstein für die Entwicklung des E-Commerce gelegt. In dieser Zeit wurden nicht nur die ersten Suchmaschinen gegründet, sondern auch bedeutende E-Commerce-Unternehmen, die ihren Online-Geschäftsbetrieb aufnahmen. Als Beispiele seien hier das Online-Auktionshaus Ebay und das Warenhaus Amazon genannt. Mittlerweile besitzen diese beiden Giganten zusammen etwas mehr als 420 Millionen aktive KäuferInnen. Der Erfolg von Amazon und Ebay lässt sich aber nicht nur durch das „First Come, First Serve“-Prinzip erklären. Als einer der ersten Online-Geschäftsbetriebe setzte Amazon auf die Analyse von kundenInnenspezifischen Daten. Das bedeutet, dass hier Marketingkampagnen aufgrund von Verhaltensanalysen der BenutzerInnen durchgeführt wurden und diese darüber hinaus auch auf den Einzelnen/die Einzelne abgestimmt waren (Team, 2016). Emails mit persönlichen Produktvorschlägen für den nächsten Einkauf, die anhand von früheren Seitenbesuchen generiert wurden, seien hier als allseits bekanntes Beispiel genannt. Heutzutage will ein Online-Geschäftsbetrieb aber nicht nur wissen, was seine KundenInnen in der Vergangenheit getan haben, sondern vielmehr welche Absichten sie in Zukunft hegen. Die zwei wichtigsten Fragen, die beantwortet werden sollen, sind: „Wofür will der Kunde/die Kundin sein/ihr Geld ausgeben?“ und „Wieviel will der Kunde/die Kundin dafür zahlen?“. Um diese Fragen adäquat zu beantworten, kommt Big Data ins Spiel. Ein kurzer Blick ins Wirtschaftslexikon definiert diesen Begriff wie folgt:

*Mit „Big Data“ werden große Mengen an Daten bezeichnet, die u.a. aus Bereichen wie Internet und Mobilfunk, Finanzindustrie, Energiewirtschaft, Gesundheitswesen und Verkehr und aus Quellen wie intelligenten Agenten, sozialen Medien, Kredit- und Kundenkarten, Smart-Metering-Systemen, Assistenzgeräten, Überwachungskameras sowie Flug- und Fahrzeugen stammen und die mit speziellen Lösungen gespeichert, verarbeitet und ausgewertet werden (Wirtschaftslexikon Gabler, 2016).*

Unternehmen aus verschiedensten Wirtschaftsbereichen versprechen sich hier neue, interessante Einblicke in das Verhalten der KundInnen. Neben den Datenmengen, die durch den Begriff Big Data abgedeckt werden, stellt sich aber ebenso die Frage, welche Methoden gibt es dafür? Während die sogenannten „Big-Player“ unter den Online-Retailern die Sammlung und Auswertung solcher Datenmengen schon für ihre Zwecke nutzen und auch über nötiges Knowhow und finanzielle Mittel verfügen, stellt sich für Klein- und Mittelunternehmen im E-Commerce spätestens an diesem Punkt die Frage: „Welche Möglichkeiten gibt es für meinen Online Shop, um eine Big Data Analyse durchzuführen, um so meine KundInnen noch gezielter anzusprechen?“. Genau an dieser Stelle soll die vorliegende Masterarbeit mit dem Begriff „Digital Analytics“ einhaken. Digital Analytics wird von Hassler (2017) wie folgt definiert:

*Digital Analytics hat zum Ziel, den Empfänger einer über einen digitalen Kanal gesendeten Botschaft besser kennenzulernen, den Menschen dahinter zu sehen, ihn zu verstehen und zukünftige Botschaften präziser auf ihn abzustimmen (Hassler, 2017).*

Bis ins Jahr 2005, als Google das kostenlose Analytics-System Google Analytics auf den Markt brachte, war „Web-Analytics“ noch ein Nischenmarkt. Die Verbreitung und Anwendung von verschiedensten Tools war bis zu diesem Zeitpunkt einigen wenigen ExpertenInnen vorbehalten. Erst mit dem Aufkommen von Tools wie Google Analytics und den dadurch vereinfachten Implementierungsmechanismen wurde Web Analytics auch für Marketingabteilungen und das Management zunehmend nutzbar (Hassler, 2017). Doch wie sieht die Situation für klein- und mittelständische Unternehmen im E-Commerce Bereich aus? Sind die Tools, die es zurzeit am Markt gibt, auch für ihre Zwecke geeignet? Wenn ja, was muss bei der Integration beachtet werden? In der vorliegenden Masterarbeit soll zuerst die Frage beantwortet werden, ob es für Klein- und Mittelunternehmen im E-Commerce eine eigene KMU-Definition gibt oder ob hier auch jene gilt, die von der Wirtschaftskammer Österreich empfohlen wird. Im Anschluss daran wird mit Hilfe von ExpertInneninterviews ein Handlungsleitfaden erstellt, welcher der betreffenden Zielgruppe den Einstieg in dieses Feld erleichtern soll.

*Inzwischen kaufen rund neun von 10 Internetnutzern online ein. Das entspricht 70 Prozent der deutschen Bevölkerung ab 14 Jahre. 39 Prozent der Internetnutzer kaufen regelmäßig online ein, allein im vergangenen Jahr mehr als zehn Mal (Bitkom, 2016).*

Diese Zahlen und das vermeintlich große Angebot von Digital Analytics Tools sollten Big Data Analysen auch für diese Zielgruppe relevant machen, um sich auf dem harten Pflaster des E-Commerce behaupten zu können.

## 1.1 Zielsetzung

Das Ziel dieser Arbeit ist es, durch ExpertInneninterviews einen Handlungsleitfaden für Online Shop BetreiberInnen, die dem klein- und mittelständischen Unternehmertum angehören, zu erstellen. Dabei sollen für diese Zielgruppe passende Digital Analytics Lösungen aufgezeigt werden und in weiterer Folge dargestellt werden, wie ein erfolgreicher Einstieg in die Integration von Big Data mithilfe dieser Tools im Unternehmen ermöglicht wird. Ein besonderer Fokus wird dabei auf die Usability der Analytics Lösungen gelegt. Dieser Handlungsleitfaden beinhaltet dabei eine Gegenüberstellung der am Markt befindlichen Digital Analytics Lösungen, die für KMUs im E-Commerce die meisten Vorteile bieten, welche Punkte bei deren Integration zu beachten sind sowie die Beleuchtung von datenschutzrechtlichen Aspekten.

## 1.2 Aufbau der Arbeit

Auf Basis der Literaturrecherche wird im ersten Teil dieser Masterarbeit der theoretische Hintergrund der vorliegenden Thematik beleuchtet, um dem Leser/der Leserin den Einstieg in die vielfältigen Aspekte von Digital Analytics zu erleichtern. Im Anschluss an den Theorieteil folgt der

empirische Teil. In diesem Abschnitt der Arbeit wird zunächst das Untersuchungsdesign und damit das ExpertInneninterview als ausgewählte Methode vorgestellt. Im Zuge dessen wird auch begründet, warum für den vorliegenden Sachverhalt diese Methode ausgewählt wurde. In den darauffolgenden Kapiteln wird die Durchführung der leitfadengestützten ExpertInneninterviews vorgestellt. In weiterer Folge wird die Analyse der Ergebnisse der Interviews erläutert. Auf Basis der Datenanalyse wird im vorletzten Kapitel dieser Masterarbeit der Handlungsleitfaden für klein- und mittelständische Unternehmen präsentiert. Als Abschluss folgt ein Resümee, das die wichtigsten Punkte sowie Ergebnisse dieser Arbeit nochmals zusammenfassen soll.

### **1.3 Forschungsfrage**

Welche Analytics Lösungen sind für Online Shop BetreiberInnen, die dem klein- und mittelständischen Unternehmertum zuzuordnen sind, verfügbar, um mit Hilfe von Digital Analytics Tools Big Data in ihren Unternehmen zu integrieren.

## 2 THEORETISCHER TEIL

Der Theorieteil dieser Masterarbeit soll den Grundstein für das Verständnis der verwendeten Begrifflichkeiten im empirischen Teil legen. Der theoretische Abschnitt unterteilt sich grob in folgende Themen: Grundlagen E-Commerce, Klein- und Mittelunternehmen und Grundlagen Big Data.

### 2.1 Grundlagen E-Commerce

In diesem Kapitel wird auf die Grundlagen des E-Commerce eingegangen. Zu Beginn wird ein Blick auf die Entwicklung zur Informationsgesellschaft geworfen. Im nächsten Abschnitt geht es um die Unterschiede zwischen E-Business und E-Commerce, um darauf folgend die Eigenschaften von E-Commerce zu erläutern. Dabei wird der Fokus auf verschiedene Geschäftsbeziehungen wie B2B, B2C etc. gelegt. Den Abschluss dieses Kapitels bildet eine Einführung in das Thema Web Analytics.

#### 2.1.1 Entwicklung zur Informationsgesellschaft

Aufgrund seiner Tragweite wird die Entwicklung zur Informationsgesellschaft oft mit der industriellen Revolution verglichen. Die Handlungsfähigkeit einzelner Personen und Organisationen und die Entwicklung hin zu einer offenen Gesellschaft mit kultureller Vielfalt soll durch die Anwendung von Informations- und Kommunikationstechnologien unterstützt werden. Der wirtschaftliche Wandel sorgt dafür, dass der Faktor Produktion gegenüber dem Faktor Information an Bedeutung verliert. Aus diesem Grund verlagern Unternehmen ihre Geschäftsprozesse ins Web und realisieren mit Hilfe elektronischer Kommunikationsmittel die KundInnenbeziehungen (Meier & Stormer, 2012).

Für die fortschreitende Entwicklung des E-Business werden folgende fünf Faktoren ausgemacht:

1. Die Entwicklung der Rechnerleistung: Mitte des letzten Jahrhunderts waren Computer gerade einmal in der Lage, simple Additionen durchzuführen, wobei sie dafür mehrere Stunden benötigten, den Platz einer Lagemhalle erforderten und Kosten für die Anschaffung im siebenstelligen Millionenbereich verursachten. Steigende Rechnerleistung bei gleichzeitig sinkenden Kosten für die Hardware und die Miniaturisierung der Komponenten sind dafür verantwortlich, dass ohne räumliche und zeitliche Beschränkung heutzutage die Informationsübertragung durchgeführt werden kann (Kollmann, 2013).
2. Die Kraft der Digitalisierung: Durch die Digitalisierung ist es möglich, ohne Qualitätsverlust große Mengen von Text, Bildern und anderen Informationen in hoher Geschwindigkeit zu bearbeiten, zu kopieren, zu übertragen und anzuzeigen (Bode, 1997). Erhebliche

Kostenvorteile lassen sich auch durch die elektronische Erfassung, Verarbeitung und Weitergabe von 0/1-Daten durch die Digitalisierung festhalten. Bei realen Informationsprodukten bzw. -trägern lassen sich mit Zunahme der Ausbringungsmenge von Informationsinhalten nur bedingt Kosten einsparen. Im Gegensatz dazu ist bei digitalen Informationsträgern nur die Erstellung des Informationsträgers mit Kosten verbunden. Die Verbreitung der Daten verursacht in weiterer Folge nur mehr marginale Kosten (Wirtz, 2001). Siehe dazu Abbildung 1 von Wirtz (2001):

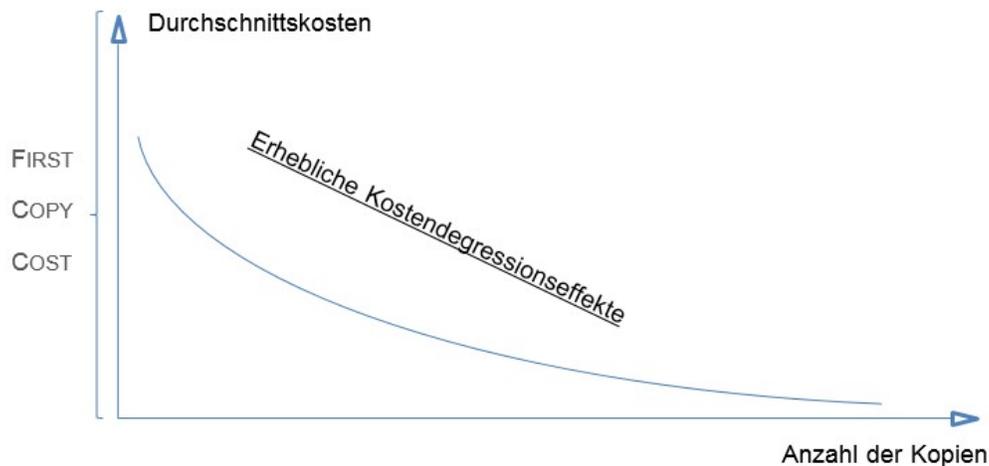


Abbildung 1: Kostendegressionseffekte bei der Verteilung digitaler Information (Wirtz, 2001)

3. Die Zunahme der Vernetzung: Durch immer günstiger werdende PCs ist es quasi für jeden/jede möglich, am „Datenhighway“ teilzunehmen. Die Kommunikationsformen verändern sich, die Globalisierung der Welt nimmt weiter zu und aufgrund dessen lösen sich die Marktgrenzen immer mehr auf. Ohne diese räumliche Beschränkung lassen sich Daten mit rasender Geschwindigkeit von einem Punkt zum anderen innerhalb dieser Netze übertragen. Ein wichtiger Punkt ist hierbei der Zugang zu diesen Netzen. Zugangswege zu diesen Netzwerken sind Telefonnetz, Kabelfernsehnetz, direkte Glasfaseranbindung, Elektrizitätsnetz, Satellitenzugang oder terrestrische Funktechnologien. Insbesondere im letzten Bereich ist mit LTE (Long Term Evolution) eine Technologie entwickelt worden, die Übertragungsraten von bis zu 100 Mbit/s ermöglicht (Kollmann, 2013).
4. Das Wachstum der Datenmenge: Auf diesen Punkt wird detailliert in Kapitel 2.2. eingegangen.
5. Die Mobilisierung der Datenübertragung: Neben der zunehmenden Vernetzung von stationären Computern ist auch bei Mobiltelefonen ein stetiger Anstieg der Datenübertragung zu beobachten. Dieser Trend wird durch die steigende Verbreitung von Smartphones und durch die Entwicklung neuer Übertragungstechnologien gefördert (Kollmann, 2013).

## 2.1.2 E-Commerce als Teil von E-Business

Nachfolgend werden nun die Unterschiede zwischen E-Business und E-Commerce erläutert.

### 2.1.2.1. Definition E-Business

Der Begriff E-Business (Electronic Business) beschreibt für unternehmensinterne und unternehmensübergreifende Geschäftsprozesse die Nutzung von Internet- und Kommunikationstechnologien. Unternehmensintern zeigt sich E-Business von der Beschaffung bis hin zum Absatz in Form der elektronischen Abwicklung. Zum E-Business zählen zusätzlich alle elektronisch unterstützten Datenflüsse zwischen einem Unternehmen und seiner Umwelt (Abts & Müller, 2010).

### 2.1.2.2. Definition E-Commerce

E-Commerce bedeutet Handel von Waren und Dienstleistungen unter Verwendung des Internets und anderen digitalen Medien (Chaffey, Mayer, Johnston, & Ellis-Chadwick, 2000). Wie in Abbildung 2 ersichtlich, werden beim Verkaufsprozess hier mehrere oder mitunter alle Phasen abgewickelt:

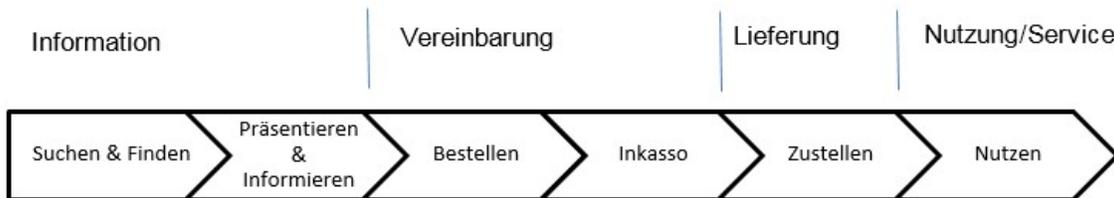


Abbildung 2: Transaktionsphasen beim Verkauf über das Internet (Müller, 2002)

Für Kollmann (2013) umfasst das Geschäftskonzept „Commerce“ die Anbahnung, Aushandlung bzw. Abwicklung von Geschäftstransaktionen über Netzwerke. Traditionelle Transaktionsphasen werden somit elektronisch unterstützt. Die schnelle und bequeme Abwicklung von Kauf- bzw. Verkaufsprozessen ist das Ziel dieses Geschäftskonzeptes (Kollmann, 2013).

### 2.1.3 AkteurlInnen und Interaktionsmuster

Laut Abts & Müller (2001) können Informationsaustausch und Transaktionen zwischen folgenden beteiligten PartnerInnen erfolgen:

- Mindestens zwei Unternehmen (Business-to-Business oder abgekürzt B2B)
  - Beispiel: pipelife.at
- Unternehmen und Privatpersonen (Business-to-Consumer oder kurz B2C)
  - Beispiel: amazon.at
- Unternehmen und öffentlichen Verwaltungen (Business-to-Administration oder kurz B2A)

- Betrifft vor allem formalisierte Ausschreibungsverfahren
- Privatpersonen und öffentliche Verwaltungen (Consumer-to-Administration oder kurz (C2A)
  - Portale von Gemeinden, die z.B. das Ändern des Hauptwohnsitzes anbieten

Informationsaustausch und Transaktionen gibt es laut Kollmann (2013) darüber hinaus noch zwischen folgenden Parteien:

- Privatpersonen und Privatpersonen (Consumer-to-Consumer oder abgekürzt C2C)
  - Beispiel: Ebay.at

### 2.1.4 Marktzahlen und Prognosen

Um ein Gefühl für die regionale und globale Entwicklung des E-Commerce zu erhalten, werden in diesem Kapitel die Jahresumsätze in Österreich von 2006 bis 2013 sowie eine globale Prognose bis 2020 präsentiert.

Wie in Abbildung 3 zu sehen, stieg seit 2006 der Brutto-Jahresumsatz in Österreich im Internet-Einzelhandel in Prozent des gesamten Einzelhandelsvolumens bis 2013 um 2,2 Prozent:

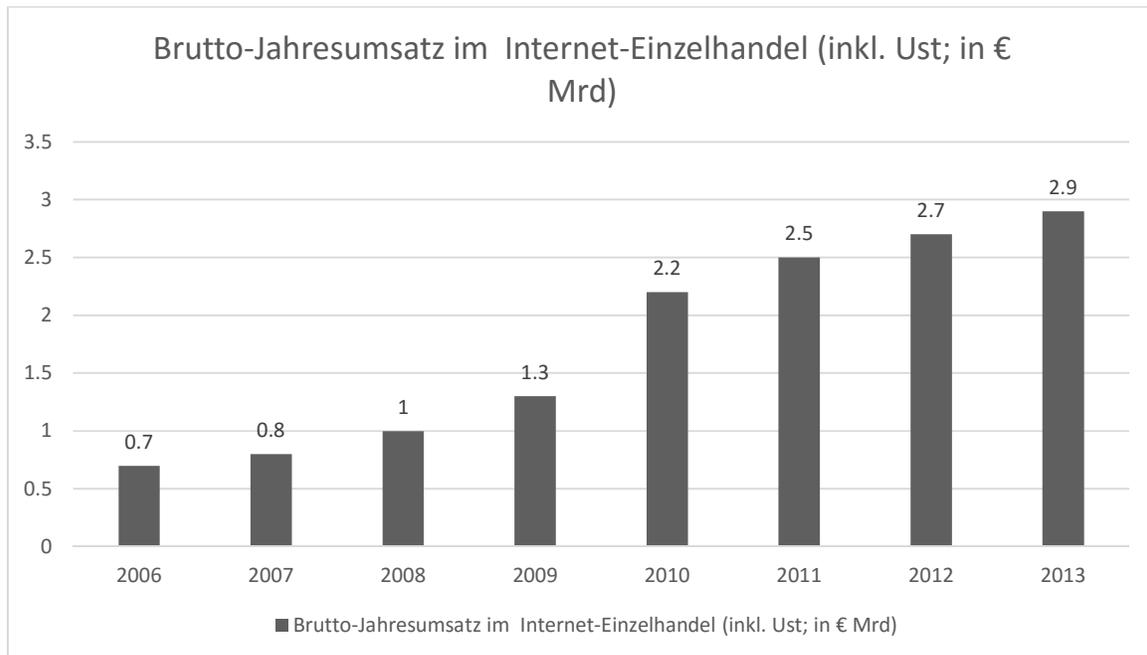


Abbildung 3: Brutto-Jahresumsatz im Internet-Einzelhandel in € Mrd (Gittenberger & Vogl, 2014)

Der Anteil am gesamten Einzelhandelsumsatz (exkl. Tankstellen) steigerte sich im selben Zeitraum von 1,3 Prozent auf 4,5 Prozent (Gittenberger & Vogl, 2014).

In Abbildung 4 ist zum Vergleich der weltweite Umsatz im E-Commerce von 2015 inklusive einer Prognose bis zum Jahr 2020 in Milliarden Dollar dargestellt:

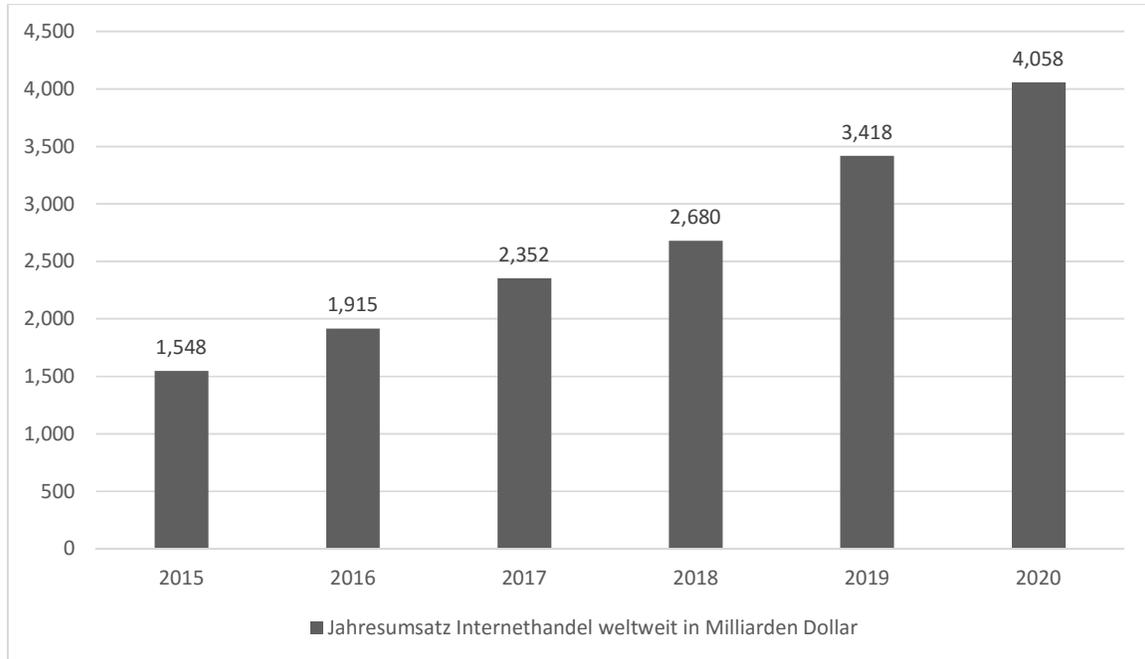


Abbildung 4: Jahresumsatz Internethandel weltweit in Milliarden Dollar (eMarketer Inc., 2016)

Der Anteil am gesamten Umsatz soll dabei von 7,4 Prozent auf 14,6 Prozent steigen. Dabei wird der asiatische Raum der größte E-Commerce Marktplatz der Welt bleiben (eMarketer Inc., 2016).

## 2.2 Grundlagen Digital Analytics

Im folgenden Kapitel werden die Grundlagen von Digital Analytics erläutert. Neben einer allgemeinen Einführung zu Beginn des Kapitels werden dem Leser/der Leserin ebenso Ursprung und Unterschiede zwischen „Digital Analytics“ und „Web Analytics“ erklärt. Weiter geht es mit den grundlegenden Methoden und der Funktionsweise von Digital Analytics, um dann mit einer Einführung in die Metriken und datenschutzrechtlichen Aspekte abzuschließen.

### 2.2.1 Einführung und Begriffsdefinition Digital Analytics

Das Ziel von Digital Analytics ist es, die EmpfängerInnen einer über das Internet gesendeten Botschaft besser kennen zu lernen. Es wird dabei versucht, den Menschen dahinter zu sehen und ihn zu verstehen, um somit zukünftige Botschaften präziser auf ihn abzustimmen. Bereits in den Anfängen des Internets stellten sich die BetreiberInnen der wenigen Websites folgende Frage: „Schaut sich das überhaupt jemand an?“. Mit der Kommerzialisierung des Internets und der damit einhergehenden Öffnung für die breite Bevölkerung änderte sich die Fragestellung von „Ob“ auf „Wie viele“. Es war nun von Relevanz, zu wissen, wie viele Personen an einem Tag eine Website besuchten. Lange Zeit verharrte die Fragestellung auf dieser Ebene bei Web-Server Verantwortlichen. Mittlerweile ist Digital Analytics allerdings viel mehr als die reine Frage nach der Anzahl an Seitenaufrufen. Nach dem Platzen der Internet-Bubble war anstatt der Cash Burn Rate – die Geschwindigkeit, in der ein Unternehmen Geld aus dem Fenster hinauswarf – der ROI

(Return on Investment) ein entscheidendes Kriterium. InvestorInnen wollten ab diesem Zeitpunkt wissen, ob das eingesetzte Kapital auch eine dementsprechende Wirkung erzeugte. Heute ist das Ziel, herauszufinden, warum BesucherInnen eine Website erfolglos – ohne einen Kauf abgeschlossen zu haben – wieder verlassen, welche Art von Produkten ein hohes Cross-Selling Potenzial aufweisen oder bei welchen Online-Kampagnen die höchste Erfolgsquote und Rentabilität erzielt wurde (Hassler, 2010). Die Digital Analytics Association (2016) definiert Web Analytics wie folgt:

*Web Analytics ist die Messung, Sammlung, Analyse und Auswertung von Internet-Daten zwecks Verständnis und Optimierung der Web-Nutzung (Digital Analytics Association, 2016).*

Historisch gesehen stammt der Begriff „Web Analytics“ aus einer Zeit, in der sich die Interaktion der BenutzerInnen auf die Website konzentrierte. Dies lag vor allem daran, dass mobile Geräte und auch Social Media zu dieser Zeit keine gewichtige Rolle spielten. Spricht man also von Web Analytics geht man im engeren Sinne nur von der Analyse und Optimierung von Webseiten aus. Demnach handelt es sich also bei Web Analytics um einen Teilbereich von Digital Analytics. Allerdings wurden auch früher unter dem Begriff Web Analytics die Betrachtung von anderen Kanälen miteinbezogen. Die Einführung von Digital Analytics kann deshalb durchaus als Neubenennung dieser Disziplin bezeichnet werden (Hassler, 2017). Um eine einheitliche Benennung zu gewährleisten, wird in der vorliegenden Masterarbeit immer der Begriff Digital Analytics verwendet.

## **2.2.2 Digital Analytics Methoden**

Unter Digital Analytics fallen sämtliche Erhebungen, die dabei helfen, die BesucherInnen der Website besser zu verstehen und das darin Angebotene zu verbessern. Web Analytics im weiteren Sinne umfasst folgende Methoden:

- **Page Tagging:** Dabei wird auf jeder Seite, auf der sich ein Webangebot befindet, ein Code (oder Tag) implementiert. Dieser Code wird dann dafür verwendet, die Seitenaufrufe oder auch das Verhalten der BenutzerInnen zu messen. Die Tools, die dazu verwendet werden, führen damit mehr oder weniger eine Vollerhebung des gesamten Traffics auf der Website durch. Das bekannteste und wohl auch das am weitesten verbreitete Tool ist hierbei der Google Tag Manager (Hassler, 2010). Nachfolgend in Abbildung 5 ist die Oberfläche von Google Analytics zu sehen:

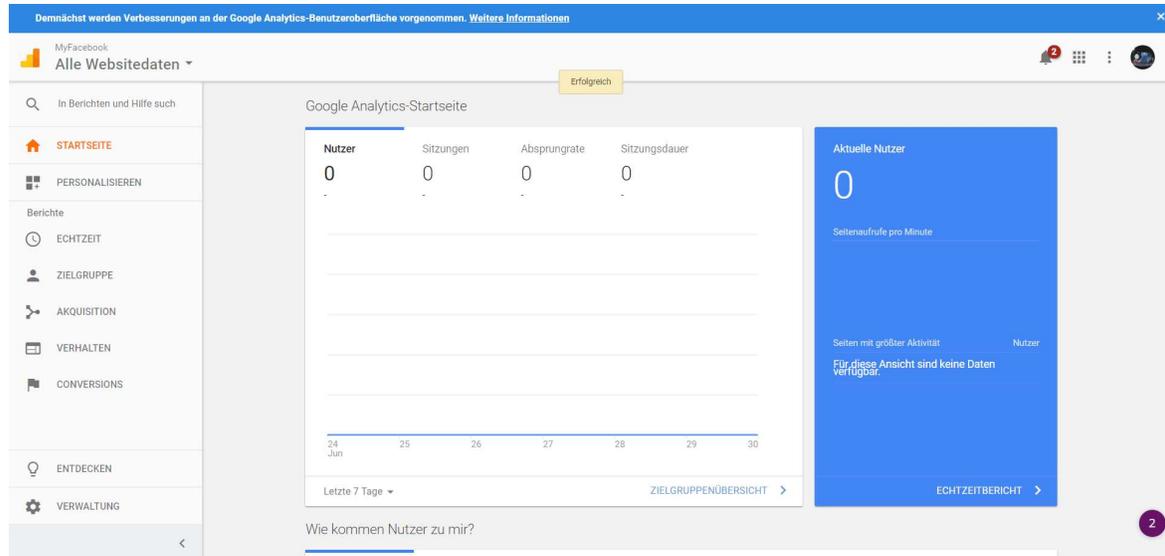


Abbildung 5: Oberfläche von Google Analytics

- **Logfile-Analyse:** Bei den Logfile-Analysen handelt es sich um den Vorgänger des heutzutage in Verwendung befindlichen Page Taggings. Diese Methode stammt noch aus einer Zeit, in der das Internet von der Technik selbst und nicht vom darauf aufbauenden Business geprägt wurde. Jeder Webserver protokolliert alle Aktivitäten, wie z.B. das Ausliefern eines Bildes oder Stylesheets sowie jeden Seitenaufruf (sei es nun korrekt oder fehlerhaft), die über den Tag entstehen. Aus dieser Masse an Daten werden dann die benötigten Informationen extrahiert (Hassler, 2010). Der Inhalt dieses Logfiles ist in Listing 1 zu sehen:

```
47.32.34.63 - - [02/Jul/2017:06:26:43 +0200] "GET
/shop/_ui/responsive/theme-red/images/***_logo.svg HTTP/1.1" 304 0 "-"
"Mozilla/5.0 (iPhone; CPU iPhone OS 10_2_1 like Mac OS X)
AppleWebKit/602.4.6 (KHTML, like Gecko) Mobile/14D27
[FBAN/FBIO;FBAV/99.0.0.57.70;FBBV/63577032;FBDV/iPhone9,2;FBMD/iPhone;FBS
N/iOS;FBSV/10.2.1;FBSS/3;FBCR/U.S.Cellular;FBID/phone;FBLC/en_US;FBOP/5;FB
RV/63577032]" upstream_addr "127.0.0.1:6081" upstream_response_time
"0.001" request_time "0.001" "10" ***.com
```

Listing 1: Inhalt eines Server-Logfiles

- **A/B- und multivariates Testing:** Durch dieses System ist es möglich, zwei oder mehrere Varianten einer Website miteinander zu vergleichen – siehe Abbildung 6. Ebenso ist es möglich, für einzelne Elemente (z.B. Banner) auf einer Website die Performance von zwei Varianten zu messen. Um dies zu eruieren, werden unterschiedlichen BesucherInnen verschiedene Variationen der Elemente gezeigt. Die statistisch beste Variante lässt sich ermitteln, indem das Verhalten des Besuchers/der Besucherin bei Varianten A und B untersucht wird, beispielsweise durch Klickevents (Hassler, 2010).



Abbildung 6: Testing zweier Varianten: Welche spricht mehr BesucherInnen an? (Eigene Darstellung)

- Online-Umfragen/Surveys: Eine noch heute praktizierte Methode zur Messung verschiedenster Faktoren ist die Online-Umfrage. Dabei wird durch Zufallsprinzip einem Prozentsatz der Website-BesucherInnen ein Popup eingeblendet. Sofern der User/die Userin möchte, kann er/sie an einer Umfrage teilnehmen, um so zur Verbesserung der Website beizutragen. Oft wird diese Teilnahme mit einem Rabattgutschein oder Ähnlichem vergütet. Der Vorteil gegenüber den zuvor genannten Methoden ist jener, dass hier auf spezifische Themen eingegangen werden kann und man sich dadurch auch konkrete Antworten erwarten darf. Das Problem an dieser Methode ist, dass immer nur ein Bruchteil der BesucherInnen daran teilnimmt. Hier kann dann z.B. das Problem entstehen, dass immer dieselbe Personengruppe an solchen Umfragen teilnimmt und somit ein nicht repräsentatives Ergebnis entsteht. Auch besteht die Gefahr, dass durch den inflationären Einsatz dieser Möglichkeit die BesucherInnen abgeschreckt werden und in Zukunft einen Bogen um die Website machen (Hassler, 2010). In Abbildung 7 ist die Darstellung einer Online-Umfrage von Amazon zu sehen.



Abbildung 7: Beispiel einer Online-Umfrage bei Amazon.com (Quelle: <https://gesterling.files.wordpress.com/2007/05/amazon-survey.jpg>, 1.7.2017)

### 2.2.3 Funktionsweise von Analytics-Systemen

Die Funktionsweise von Analytics-Systemen lässt sich, wie in Abbildung 8 zu sehen, in drei Hauptaufgaben unterteilen:

- Datensammlung
- Speicherung und Verarbeitung der Daten
- Auswertung der Daten

Bei der Datensammlung geht es in erster Linie darum, die „Spuren“, die BesucherInnen auf der Website hinterlassen, festzuhalten. Im zweiten Schritt wird festgelegt, wie die Daten gespeichert werden und wie sie zur weiteren Analyse aufbereitet werden sollen. Der letzte Schritt befasst sich mit der Analyse der Daten. Die Auswertung der Daten wird normalerweise von Web-AnalystInnen vorgenommen. Dies ist gleichzeitig auch der sichtbarste Teil der Verarbeitungskette, da es letztendlich auch um die (grafische) Präsentation der Ergebnisse geht (Hassler, 2017).

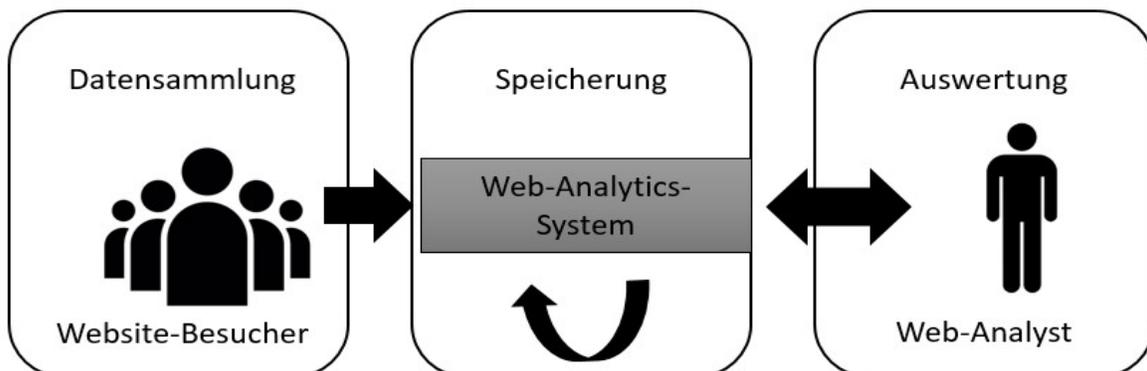


Abbildung 8: Funktionsbereiche von Digital-Analytics (Hassler, 2017)

Im ersten Teil dieser drei Bereiche wird die Datensammlung beleuchtet. Diese gehört mit dem Sammeln und Registrieren von Website-BesucherInnen zu den essentiellen Aufgaben eines Analytics-Systems. Als wichtiges Kriterium neben der Quantität der Daten zählt auch deren Qualität. Wie in vielen anderen Bereichen lässt sich auch hier das „GIGO-Prinzip“ („garbage in – garbage out“) anwenden. Selbsterklärend bedeutet dies, dass bei mäßiger Qualität der Input-Daten ebenso der Output von derselben „Qualität“ sein wird (Hassler, 2010).

Hassler (2017) differenziert bei der Datensammlung zwischen drei verschiedenen Ansätzen:

- Datensammlung serverseitig mittels Logfile
- Datensammlung clientseitig mittels Page Tagging
- Alternative Methoden wie Packet Sniffing

Diese Ansätze unterscheiden sich nicht nur in ihren technischen Eigenschaften, sondern auch in ihren Ergebnissen. Hier ist bereits im Vorhinein darauf zu achten, welche Daten für die Auswertung in weiterer Folge benötigt werden, um dementsprechend auch die dafür geeignetste Methode auszuwählen (Hassler, 2010).

### 2.2.3.1. Serverseitige Datensammlung mittels Logfile

Die Datensammlung mittels Logfile ist jene Methode, die seit Anbeginn des Internets eingesetzt wurde. Damals wie heute werden dabei die Informationen aus den Logfiles des Webserver extrahiert, da dieser alle Aufrufe von UserInnen an eine Website beantwortet (Hassler, 2017). Das Logfile besteht dabei aus folgenden Informationen:

- Datum und Uhrzeit des Aufrufs. Durch diese Angaben lassen sich unter anderem Störungen am Webserver zeitlich einschränken (Hassler, 2017). Nachfolgend in Listing 2 ein Beispiel dazu:

```
47.32.34.63 - - [02/Ju1/2017:06:26:43 +0200] "GET
/shop/_ui/responsive/theme-red/images/***_logo.svg HTTP/1.1" 304 0 "-"
"Mozilla/5.0 (iPhone; CPU iPhone OS 10_2_1 like Mac OS X)
AppleWebKit/602.4.6 (KHTML, like Gecko) Mobile/14D27
[FBAN/FBIOs;FBABV/99.0.0.57.70;FBBV/63577032;FBDV/iPhone9,2;FBMD/iPhone;FBS
N/iOS;FBSV/10.2.1;FBSS/3;FBCR/U.S.Cellular;FBID/phone;FBLC/en_US;FBOP/5;FB
RV/63577032]" upstream_addr "127.0.0.1:6081" upstream_response_time
"0.001" request_time "0.001" "10" ***.com
```

Listing 2: Datum, Uhrzeit und Zeitzone in einem Logfile

- Die URL im Logfile gibt den Pfad zur aufgerufenen Datei am Webserver an (Hassler, 2017). Als Beispiel zu sehen in Listing 3:

```
118.34.111.77 - - [02/Jul/2017:06:26:43 +0200] "GET
/shop/_ui/responsive/theme-red/fonts/icomoon.ttf?vyzbmj HTTP/1.1" 304 0
"https://***.com/shop/de/?emsrc=coop&refID=header-link/de" "Mozilla/5.0
(iPhone; CPU iPhone OS 10_3_2 like Mac OS X) AppleWebKit/603.2.4 (KHTML,
like Gecko) Version/10.0 Mobile/14F89 Safari/602.1" upstream_addr
"127.0.0.1:6081" upstream_response_time "0.000" request_time "0.001" "10"
***.com
```

Listing 3: Aufgerufene URL einer Website

Durch die IP-Adresse wird der Rechner oder das Netzwerk des Aufrufs eindeutig identifiziert. Der Webserver benötigt diese Information, um zu wissen, wohin die angeforderten Informationen zurückgesendet werden sollen. Man kann die IP-Adresse mit der AbsenderInnenadresse auf einem Brief vergleichen, auf die der Empfänger/die Empfängerin antworten kann. Dies ist vor allem für statistische Aufrufe vorteilhaft, da jeder Aufruf einer IP zugeordnet werden kann. Mehr ist jedoch meist nicht möglich, da ein Großteil der IP-Adressen nicht fix zugeordnet ist. Der Grund dafür ist, dass UserInnen, die sich über DSL einwählen von ihrem ProviderInnen in Intervallen eine andere IP-Adresse zugewiesen bekommen. Somit ist es möglich, dass derselbe User/dieselbe Userin mit zwei unterschiedlichen IP-Adressen die Website aufruft. Umgekehrt ist es natürlich auch möglich, dass eine IP-Adresse zwei UserInnen zugeordnet werden kann. Dies ist dann der Fall, wenn ein User/eine Userin vom selben Provider die IP-Adresse eines anderen Users/einer anderen Userin übernimmt. Darüber hinaus ist es ebenso möglich, dass UserInnen, die von einem Firmennetzwerk aus surfen, alle eine identische IP-Adresse besitzen. In Firmennetzwerken ist es nicht üblich, dass jeder Client eine eigene öffentliche IP-Adresse besitzt, sondern nur der Verbindungsknoten zwischen der Firma und dem Internet. Die IP-Adresse dieses Knotens bekommt jeder Client zum Surfen im Internet zugewiesen (Hassler, 2017). Eine IP-Adresse ist fett markiert in Listing 4 zu sehen:

```
66.249.92.31 - - [02/Jul/2017:06:26:54 +0200] "GET
/shop/de/personalize/p/?productCode=20141 HTTP/1.1" 200 466
"https://***.com/shop/de/damen-funktions-
fleecejacke/20141/?emsrc=retarget&refID=criteo/lowerfunnel/de" "AdsBot-
Google (+http://www.google.com/adsbot.html)" upstream_addr "127.0.0.1:801"
upstream_response_time "0.022" request_time "0.023" "30" ***.com
```

Listing 4: IP-Adresse in einem Logfile

- Der UserInnen-Agent gibt bei jedem Aufruf einer Website dessen Browser und Betriebssystem an. Die Angabe dieser Informationen hat einen einfachen Grund: Oftmals wird aufgrund der Darstellungsoptimierung die Seite für die jeweiligen Browser unterschiedlich ausgegeben. Zum Beispiel kann das benötigte Format eines Safari

Browsers auf einem Macintosh sich von dem eines Chrome Browsers, der mit Windows betrieben wird, unterscheiden – siehe auch Listing 5 (Hassler, 2017):

```
94.250.167.49 - - [02/Jul/2017:06:27:33 +0200] "GET
/shop/_ui/responsive/theme-red/images/***_logo.svg HTTP/1.1" 304 0 "-"
"Mozilla/5.0 (iPad; CPU OS 10_3_1 like Mac OS X) AppleWebKit/603.1.30
(KHTML, like Gecko) Version/10.0 Mobile/14E304 Safari/602.1" upstream_addr
"127.0.0.1:6081" upstream_response_time "0.001" request_time "0.001" "10"
***.com
```

Listing 5: UserInnen-Agent in einem Logfile

- Der Status eines Aufrufs war die eigentliche Motivation der Erstellung eines Logfiles am Server. Dieser Status kann unter anderem angeben, ob der Aufruf einer Ressource auf einer Website korrekt verlief oder, ob es dabei einen Fehler gab (Hassler, 2017).

Da aber zwischen korrekt und fehlerfrei noch weitere Unterscheidungen existieren, gibt es laut Hassler (2017) folgende Hauptgruppen:

- 1xx: Die Bearbeitung der Anfrage dauert noch an.
  - 2xx: Die Anfrage konnte erfolgreich bearbeitet werden.
  - 3xx: Von Seiten des Clients sind weitere Schritte nötig, um die Anfrage der Bearbeitung erfolgreich abzuschließen.
  - 4xx: Der Grund für das Scheitern der Anfrage liegt eher im Bereich des Clients.
  - 5xx: Der Grund für das Scheitern der Anfrage liegt eher im Bereich des Servers.
- 
- Bei den Cookies handelt es sich um kleine Textdateien, die Informationen von einer Website beinhalten und auf dem Rechner des Besuchers/der Besucherin gespeichert werden. In Cookies werden Textinformationen als Schlüssel-und-Werte-Paare gespeichert. Für Website BetreiberInnen ist das die einzige Möglichkeit, Informationen auf dem Rechner des Besuchers/der Besucherin zu speichern, ohne diesen explizit um Erlaubnis zu fragen. Aus Datenschutzgründen ist das Auslesen des Cookies nur jener Website erlaubt, die auch das Cookie gesetzt hat. Das gilt aber nur für „First-Party-Cookies“, also Cookies, die nur von jenem Webserver gelesen und geschrieben werden, auf dem man sich gerade befindet. Problematisch wird es erst bei sogenannten „Third-Party-Cookies“. Hierbei handelt es sich um Cookies, die wie der Name schon sagt, nicht vom Webserver eingebunden wurden von dem auch die Hauptseite ausgeliefert wird. Solche „Third-Party-Cookies“ können zum Beispiel von Ad-Servern zur Verfügung gestellt werden. Diese Ad-Server werden zu Werbezwecken auf verschiedenen Webseiten eingebunden und genau an diesem Punkt wird es auch kritisch; nun ist es dem Ad-Server möglich, aus dem Cookie alle Seiten auszulesen, die der User/die Userin besuchte. Das

ist unter anderem auch der Grund, warum Cookies von manchen ExpertInnen kritisch gesehen werden (Hassler, 2017).

### **2.2.3.2. Clientseitige Datensammlung**

Die im vorangegangenen Kapitel besprochene Methode der serverseitigen Datensammlung erweist sich zwar technisch gesehen als recht simpel und für viele Zwecke der Informationsgewinnung als ausreichend, dennoch betrachtet diese Methode BesucherInnen nur aus Sicht des Servers und somit einen nur sehr kleinen Ausschnitt des Besuches auf einer Website. Information über das Verhalten zwischen zwei Seitenaufrufen werden nicht gesammelt. Um auch Informationen aus diesem zeitlichen Abschnitt zu extrahieren, wurde Ende der 1990er mit der clientseitigen Datensammlung begonnen. Clientseitig bedeutet in diesem Fall, dass nicht mehr der Webserver die Tätigkeiten protokolliert, sondern der Client mit Hilfe des Browsers. Indem der Browser beim Besuch die Website lädt, ist es mit JavaScript möglich, BenutzerInnenaktionen und Browsereinstellungen zu verfolgen (Hassler, 2017).

Dabei lassen sich laut Hassler (2017) unter anderem folgende Merkmale feststellen:

- Mausklicks, die aktuelle Position des Mauszeigers und Mausbewegungen.
- Im Browserfenster getätigte Tastatureingaben.
- Die Fenstergröße des Browsers.
- Plugins, die im Browser installiert worden sind.

Zusätzlich zu diesen Informationen können natürlich auch jene gesammelt werden, die sich in den Server-Logfiles befinden. Um die gesammelten Daten auch für Analysezwecke zu nutzen, benötigt es das bereits vorgestellte Page Tagging. Beim Page Tag handelt es sich um ein verstecktes Element, welches in jede Seite einer Website integriert wird. Der Tag ist gewöhnlich in JavaScript geschrieben und ist 1x1-Pixel groß. Letzteres bedeutet, dass das Bild vom User/von der Userin praktisch nicht gesehen wird. Ruft der User/die Userin nun eine Website auf, wird zuerst die normale Seite in den Browser geladen. Im Zuge dessen wird im Hintergrund auch das Page Tag geladen. Die gewünschten Informationen werden dann vom Page Tag gesammelt. Sobald die gewünschten Informationen beisammen sind, wird das unsichtbare Bild von einem Drittserver durch JavaScript geladen. Durch dieses nachträgliche Laden ist es möglich, die Informationen an den Drittserver zu übermitteln. Auch hier muss aber wieder ein Trick angewandt werden, denn das Übermitteln an einen Drittserver wird vom Browser nicht erlaubt. Um das zu umgehen, werden die gesammelten Informationen in den URL-Aufruf des versteckten Bildes angehängt bzw. mitgesendet. Aufgrund der angehängten Informationen ergeben sich sehr lange URLs (Hassler, 2017). In Abbildung 9 ist dieser Vorgang nochmals grafisch aufbereitet:

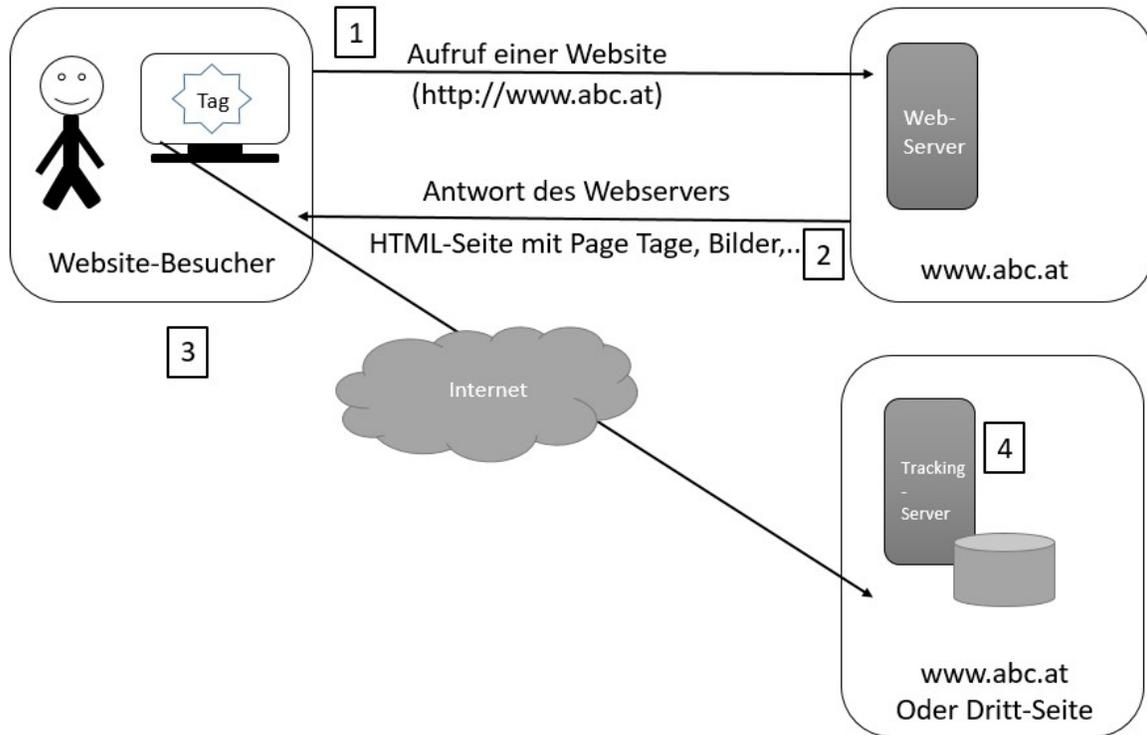


Abbildung 9: Clientseitige Datensammlung mittels Page Tagging (eigene Darstellung in Anlehnung an Hassler (2017))

### 2.2.3.3. Alternative Datensammlungsmethoden

Neben der serverseitigen Logfile- und clientseitigen Page-Tag-Datensammlung gibt es auch alternative Methoden, die versuchen, die Nachteile der beiden konventionellen Methoden wett zu machen. Da die beiden bereits vorgestellten Methoden eine weitaus höhere Einsatzquote besitzen, sind alternative Methoden am Markt eher eine Randerscheinung (Hassler, 2017).

Zu diesen alternativen Datensammlungsmethoden gehören laut Hassler (2017):

- Reverse Proxies
- Packet Sniffer
- Hybride Methoden (client- und serverseitig)

Reverse Proxies sind Rechner, die dem Webserver vorgeschaltet werden. Ihre Aufgabe dabei ist es, den eingehenden Datenverkehr entgegenzunehmen und erst dann die Anfragen weiterzuleiten. Somit kontrollieren Reverse Proxies sämtlichen Datenverkehr auf einer Website. Der Reverse Proxy besitzt nicht nur die Fähigkeit, den Datenverkehr zu lesen, sondern auch zu modifizieren. So ist es möglich, jeder Ressource, die vom Webserver ausgeliefert wird, einen Page Tag zu integrieren, noch bevor sie den Browser der BesucherInnen erreicht. In der Praxis sind Reverse Proxies eine Seltenheit, da der Hardware- und Konfigurationsaufwand für diesen Zweck sehr aufwendig ist. Packet Sniffer verfolgen das Ziel, sämtlichen Datenverkehr zwischen

Webserver und Client zu „beschnuppern“, indem sie sich vor dem Webserver einklinken. Dies hat den Vorteil, dass, sobald die Website auf mehrere physische Server verteilt ist, das Agglomerieren der Logfiles entfällt. Abgesehen davon bietet diese Methode keinen Vorteil gegenüber der serverseitigen Logfile-Datensammlung. Weiters kann auch auf hybride Methoden zurückgegriffen werden. Diese zeichnen sich dadurch aus, dass sie die Vorteile beider Methoden kombinieren. Aufgrund der Schwierigkeiten bei der Auswertung nach Zusammenführung der beiden Datensätze stellen auch hybride Methoden in Praxis bis jetzt eher eine Seltenheit dar (Hassler, 2017).

## 2.2.4 Metriken in Digital Analytics

Um aus den gesammelten Daten mit Digital Analytics auch richtige Schlüsse zu ziehen, ist es essentiell, die Metriken zu verstehen. Nachfolgend wird erklärt, welche Bedeutung diese Metriken besitzen, wie sie zustande kommen und für welche Zwecke man sie nutzen kann:

- Beim „Hit“ handelt es sich um die älteste Metrik im Feld der Digital Analytics. Dieser beschreibt jeden einzelnen Aufruf einer Datei auf einem Server. In Zeiten, in denen eine Website nur aus einer Textdatei und einer HTML-Seite bestand, war der Hit eine logische Messgröße. Wenn man sich aber heutzutage eine Website ansieht, wird man schnell feststellen, dass es pro Seitenaufruf mehrere Hits geben muss. Aus diesem Grund wird auch klar, dass der Hit als aussagekräftige Metrik heute keine Berechtigung mehr besitzt (Hassler, 2017). Als Anschauungsbeispiel dient folgende Website in Abbildung 10:



Abbildung 10: Beim Aufruf von orf.at werden mehrere Hits erzeugt (orf.at, 2017)

- Bei einer „Page View“ handelt es sich um eine im Browser vollständig geladene und durch den User/die Userin nutzbare Seite. (Hassler, 2007). Der Seitenaufruf ist im Gegensatz zu anderen geschäftsrelevanten Kennzahlen, wie Bestellungen oder Kontaktanfragen, wegen seiner Spontanität vorteilhaft. Diese Metrik reagiert schneller und ist somit als Frühindikator verwendbar. Aufgrund von dynamischen Anwendungen wie zum Beispiel AJAX (Asynchronus JavaScript and XML) lassen sich weitere Elemente einer Seite nachladen, ohne dass die Seite selbst einen Reload erfährt. Ein Beispiel hierfür ist die

Timeline auf Facebook, die durch beständiges herunterscrollen ohne weiteres Zutun von BesucherInnen immer nachgeladen wird. Mittlerweile setzen auch viele Nachrichtportale diese Technik ein: Ist man am Ende eines Artikels angelangt, wird man wieder automatisch auf die Startseite geleitet (Hassler, 2017). In Abbildung 11 ist diese Methode auf süddeutsche.de zu sehen:



Abbildung 11: Automatische Weiterleitung am Ende eines Artikels (süddeutsche.de, 2017)

Bei massivem Einsatz von dynamischen Anwendungen gilt es, den Seitaufruf als valide Metrik zu überdenken (Hassler, 2017).

- Ein „Visit“ wird als Besuch einer Website gezählt, unabhängig davon, wie viele Seiten besucht wurden. Technisch wird als Visit ein Besucher/eine Besucherin gesehen, der/die am selben Computer im selben Browserfenster mehrere Klicks ausführt. Dabei sollen die getätigten Klicks nicht mehr als 20 bis 30 Minuten auseinanderliegen. Löscht man die Cookies oder wechselt man den Computer, zählt das als zwei Visits. Um den Besuch einer Website richtig zu deuten, sollten immer Werte zwischen zwei Zeiträumen verglichen werden. Zusätzlich sollten diese beiden Zeiträume miteinander korrespondieren. Das bedeutet, dass zum Beispiel die Daten eines Dienstags immer mit den Daten eines anderen Dienstags verglichen werden sollten. Selbiges gilt für Monate: Der Dezember sollte immer mit dem Dezember eines anderen Jahres verglichen werden (Hassler, 2007).
- Beim „Visitor“ handelt es sich um einen User/eine Userin, welcher/welche am selben Computer mit demselben Browser weder die IP-Adresse wechselt, noch die Cookies löscht. Werden mehrere Geräte verwendet, wird dies folglich auch als mehrere BesucherInnen gewertet. BesucherInnen werden auch als „Uniques“ oder „Unique Visitors“ bezeichnet. Suchmaschinen, Sonden und andere automatisierte Besuche gelten nicht als Visitors (Hassler, 2007).
- Eine „Konversionsrate“ bezeichnet das Verhältnis der Anzahl der BesucherInnen, die einen bestimmten Prozess vollständig durchlaufen haben, zur Anzahl jener, die ihn

begonnen haben. Hier kann es sich z.B. um das Verhältnis von reinen BesucherInnen einer Website mit jenen, die einen Kauf abgeschlossen haben, handeln (Hassler, 2007).

### **2.2.5 Datenschutz und Digital Analytics**

Wenn Datenspeicherung – und im Falle von Digital Analytics auch die Speicherung von benutzerInnenbezogenen Daten – ein Thema ist, können rechtliche Überlegungen zum Datenschutz nicht ausgeblendet werden. Bei der Speicherung von personenbezogenen Daten gibt es zu dieser Thematik in verschiedenen Ländern unterschiedlich stark einschränkende Gesetze. Während in Europa diese Thematik sehr restriktiv gehandhabt wird, gibt es in den USA wiederum kaum Einschränkungen. Wichtig hierbei ist, dass unter diese Gesetzgebungen nur Daten fallen, die eindeutig einer natürlichen Person zugeordnet werden können. Bei herkömmlichen Analytics-Systemen wird dieser Bereich in der Regel nicht tangiert, da keine personenbezogenen NutzerInnendaten gespeichert werden. Im Fokus steht viel mehr die Auswertung von aggregierten Daten. Darunter fällt zum Beispiel die Gesamtzahl an BesucherInnen pro Tag. In den kritischen Bereich gelangt man erst, sobald BesucherInnen etwa ein Formular ausfüllen oder sich für einen bestimmten Teil einer Website einloggen müssen. Ist dies der Fall, müssen BesucherInnen darauf aufmerksam gemacht werden, indem sie der Speicherung ihrer persönlichen Daten explizit zustimmen müssen. Dies muss durch eine Check-Box geschehen, in der BesucherInnen zur Einwilligung ein Häkchen setzen müssen. Dieses Verfahren wird Opt-In-Verfahren genannt. Das Opt-Out-Verfahren – in dem das Häkchen weggeklickt werden muss – ist im Gegensatz dazu für diese Zwecke nicht zulässig. Ein weiterer Punkt, der zu beachten ist, ist die IP-Adresse. Eine IP-Adresse einer natürlichen Person zuzuordnen, ist ohne unverhältnismäßigem Aufwand grundsätzlich nicht möglich. Eine Ausnahme bildet hier der Fall, wenn UserInnen über eine DSL-Leitung eine statische IP-Adresse beziehen und darüber hinaus eine Website betreiben. Sind diese zwei Optionen gegeben, ist die URL der Website unter dem Namen des Betreibers/der Betreiberin bei einer Domain-Registrierungsstelle angemeldet. Nun ist es mit verhältnismäßig geringen Aufwand möglich, eine natürliche Person der IP-Adresse zuzuordnen (Hassler, 2017).

Trotz oder aufgrund dieser Umstände gilt in Österreich die IP-Adresse als personenbezogener Datensatz (Wirtschaftskammer Österreich, 2017).

Bei den Cookies gibt es seit 2009 eine EU-Richtlinie die besagt, dass beim Setzen von Cookies, welche zur Datenauswertung nicht absolut notwendig sind, von NutzerInnen das Einverständnis eingeholt werden muss. Dies soll, wie bei Speicherung von personenbezogenen Daten, mittels Opt-In geschehen. Da diese Cookie-Richtlinie nur als Anweisung an die EU-Staaten zum Erlassen entsprechender Gesetze zu verstehen ist, sind viele Länder dieser Anweisung entweder gar nicht, teilweise bzw. in unterschiedlicher Ausprägung nachgekommen. In Österreich fallen Cookies unter personenbezogene Daten. Wertet eine Website nun Cookies zum Zweck der Optimierung aus, so ist von BesucherInnen eine Einwilligung im Stile eines Opt-Ins einzuholen (Wirtschaftskammer Österreich, 2016). In Abbildung 12 ist als Beispiel eine Einholung der Zustimmung zur Cookie-Verwendung zu sehen:

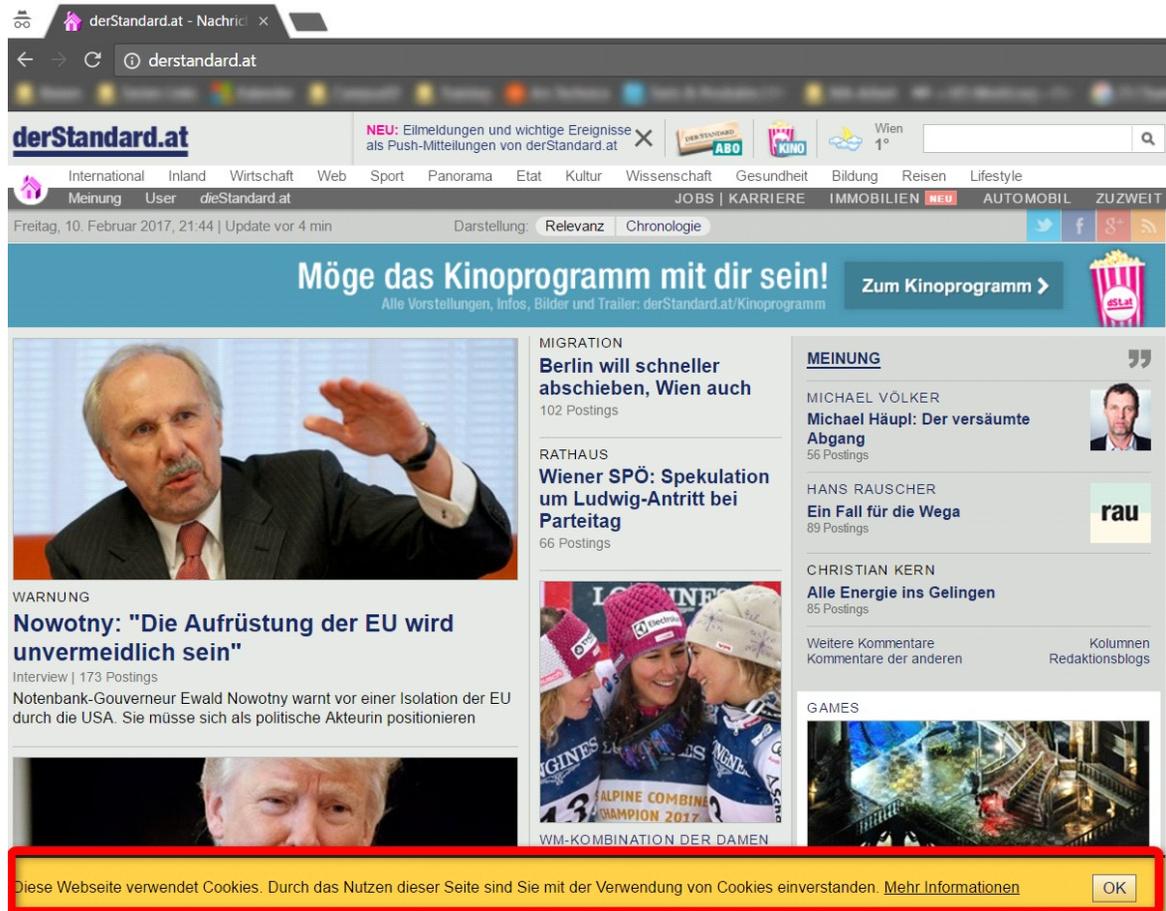


Abbildung 12: Beispiel eines Hinweises zur Cookie-Verwendung von derstandard.at

In gewissen Fällen gibt es davon aber eine Ausnahme:

*Lediglich in jenen Fällen bedarf es keiner vorherigen Zustimmung des Betroffenen, in denen der Anbieter eines Informationsdienstes (etwa ein Webshop-Betreiber) einen vom Betroffenen ausdrücklich gewünschten Dienst nur unter der Bedingung zur Verfügung stellen kann, dass Daten verwendet werden müssen und dies auch unbedingt erforderlich ist. Bei IP-Datenspeicherungen im Rahmen von Cookies zum Zwecke des virtuellen Einkaufs mit begrenzter Speicherdauer („Warenkorb“) könnte dies so gesehen werden. Hier ist lediglich der Informationspflicht nachzukommen (Wirtschaftskammer Österreich, 2017).*

## 2.3 Grundlagen Big Data

Der Erfinder des Bit und somit intellektueller Vater des digitalen Zeitalters Claude E. Shannon schätzte im Jahr 1949, dass die größte Datensammlung, die es auf der Erde gibt, ca. 12.500 MB besitzt, namentlich die U.S. Library of Congress. Im Jahr 2012 kommt Hilbert (2012) zum Ergebnis, dass im Jahr 2007 über 300 Exabyte an Daten existieren. Ein Exabyte ist eine Milliarde Gigabyte. Interessanterweise muss man hier aber hinzufügen, dass 2007 nur mehr ca. 7 Prozent der Daten in analoger Form (Druckerzeugnisse, Papierbilder etc.) gespeichert waren. Noch im Jahr 2000 waren nur ca. 25 Prozent aller weltweit gespeicherten Informationen digital verfügbar.

Der Rest war entweder in gedruckter Form, auf Film, Vinyl, Audiokassetten etc. verfügbar. Die Gesamtmenge gespeicherter Daten belief sich im Jahr 2013 auf 1.200 Exabyte. Davon waren weniger als 2 Prozent nicht digital. Würde man diese Menge an Daten in Büchern binden lassen, würde sie die Fläche der USA mit 52 Schichten abdecken. Auf CD-ROM gebrannt, könnte man damit fünf Stapel errichten, die zum Mond reichen würden (Hilbert, 2012).

Die immer weiter fortschreitende Digitalisierung wird diesen Trend und somit das Anwachsen der Datenmenge weiter vorantreiben. Unsere Gesellschaft in dieser Zeit ist vernetzter denn je. Besonders in der westlichen Welt zählt ein Computer so wie ein Smartphone zur Standardausstattung eines jeden Menschen. Mit dem Smartphone treten wir in Kontakt mit anderen Menschen, vertreiben uns die Zeit auf verschiedenen Social-Media Plattformen und kaufen mittlerweile sogar ein. Der Personal Computer oder das Notebook sind für Tätigkeiten da, die mit einem Smartphone oder Tablet umständlich zu erledigen sind oder die Rechenleistung noch nicht ausreicht. Das interessante hierbei ist aber eher weniger der Umstand, dass Informationen für den Menschen ubiquitär vorhanden und abrufbar sind. Viel eher sollte verblüffen, dass zu den Informationen, die man bekommt, im Gegenzug gespeichert wird, dass man nach dieser Information gesucht hat. Das bedeutet, dass das Verhalten eines Menschen, der in regelmäßigen Abständen mit einem Smartphone oder einem Computer hantiert und darauf Aktionen durchführt, gespeichert wird. Für einen einzelnen Menschen mag das keine Rolle spielen. Wenn man aber bedenkt, dass es bereits 2,4 Milliarden Smartphones auf der Erde gibt, ergibt das eine völlig neue Dimension an Informationen (Qureshi, 2016).

### 2.3.1 Definition und Eingrenzung

Das Gartner IT-Research Institute definiert Big Data wie folgt:

*High-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation (Gartner Inc, 2016).*

Wobei hinter den Begriffen Volume, Velocity und Variety folgende Bedeutung steckt:

- Volume (Datenmenge): Die Datenmenge ist einer der ersten Begriffe, der mit Big Data in Verbindung gebracht wird. Manche ExpertInnen sind hier der Meinung, dass Big Data erst ab der Größenordnung von einem Petabyte beginnt, wobei ein Petabyte 1000 Terabyte sind (Zadrozny & Kodali, 2013).
- Velocity (Geschwindigkeit bzw. Verfügbarkeit): Um die Daten optimal für ein Unternehmen zu nützen, müssen sie zeitnah verarbeitet werden, wenn sie in das Unternehmen einfließen (Zadrozny & Kodali, 2013).
- Variety (Vielfalt): Big Data umfasst neben strukturierten Daten auch noch unstrukturierte Daten wie Text, Clickstreams, Audio, Video etc. Heute sind mehr als 80 Prozent der Daten unstrukturiert (Zadrozny & Kodali, 2013).

Laut Mayer-Schönberger & Cukier (2013) gibt es keine exakte Definition für Big Data. Ursprünglich verstand man darunter eine Informationsmenge, die die Kapazitäten des Arbeitsspeichers in einem Computer übersteigt und man somit gezwungen war, neue Technologien zu entwickeln, um diese zu verarbeiten. Um dieses Problem zu lösen, wurden technologische Innovationen wie Google MapReduce oder das Open-Source-Pendant Hadoop von Yahoo entwickelt. Mit diesen Technologien war es möglich, mit sehr viel größeren Datenmengen zu arbeiten als je zuvor. Ein großer Vorteil dabei war, dass die Daten bereits bearbeitet werden konnten, auch wenn sie nicht in klassischen Datenbanken strukturiert und in einem einheitlichen Format bereit lagen. Im Zuge dessen wurden dabei die klassischen in diesem Bereich tätigen Firmen von Internetriesen wie Google und Yahoo überholt. Der Grund hierfür liegt darin, dass diese Firmen mit den durch ihre Produkte gesammelten Datenmengen ein großes Interesse daran haben, diese zu verarbeiten und insbesondere zu ihrem Nutzen auszuwerten (Mayer-Schönberger & Cukier, 2013).

Die Ambitionen der Unternehmen in diesem Bereich zeigen sich an deren Datenhunger: Google verarbeitete bereits im Jahr 2012 pro Tag 24 Terrabyte an Daten (Davenport, Barth, & Bean, 2012). Schon im Jahr 2012 hatte YouTube über 800 Millionen aktive NutzerInnen im Monat, die pro Sekunde über ein Stunde Video hochgeladen haben (Page, 2012).

In einem anderen Ansatz versuchen Mayer-Schönberger & Cukier (2013) Big Data wie folgt zu erklären:

*Big Data ist das, was man im großem, aber nicht in kleinem Maßstab tun kann, um neue Erkenntnisse zu gewinnen oder neue Werte zu schaffen, sodass sich Märkte, Organisationen, die Beziehungen zwischen Bürger und Staat und vieles mehr verändern (Mayer-Schönberger & Cukier, 2013).*

Die soeben vorgestellten Definitionen zeigen, dass sich die Möglichkeiten für die Verwendung von Big Data nicht in vorgefertigte Schubladen stecken lassen und jeder/jede nach seiner/ihrer eigenen Priorität und für das Unternehmen relevante Aspekte definieren muss.

### **2.3.2 Potenziale und Einsatzgebiete**

Neben den in Kapitel 1 erwähnten Vorteile für BetreiberInnen von E-Commerce Plattformen gibt es auch abseits davon verschiedenste Einsatzgebiete:

Um die Vorteile und Potenziale zu erläutern, eignet sich als Beispiel die Digitalisierung des Universums. Im Jahr 2000 startete das Projekt zur Himmelskartografie Sloan Digital Sky Survey. Das in New Mexico, USA, beheimatete Teleskop hatte in den ersten Wochen bereits mehr Daten gesammelt, als in der gesamten bisherigen Geschichte der Astronomie. Nach zehn Jahren kartographieren ist im Jahr 2010 die Menge an Daten auf über 140 Terabyte angewachsen (Mayer-Schönberger & Cukier, 2013).

Im Jahr 2022 soll das „Large Synoptic Survey Telescope“ auf dem chilenischen Berg Cerro Pachon in Betrieb gehen. Das 2,8 Tonnen schwere Teleskop wird ab dessen Inbetriebnahme pro Jahr 6000 Terabyte Bilddaten liefern. Das sind pro Nacht ca. 16,4 Terabyte. Diese gesammelten

Aufnahmen sollen über einen Online-Kartendienst öffentlich für jeden/jede zugänglich gemacht werden (Tilman, 2015).

Ein weiteres sehr großes Potenzial liegt in der Erkennung der Ausbreitung von ansteckenden Krankheiten. Im Jahr 2009 wurde ein neues Grippevirus namens H1N1 entdeckt. Dieses Grippevirus kombinierte Elemente der Vogelgrippe und Schweingrippe und breitete sich überaus schnell aus. Ein essentieller Teil zum Eindämmen einer Ausdehnung ist die zeitliche und geographische Vorhersage der Ausbreitung der Krankheit. Mit den für die Behörden verfügbaren Daten ist dies aber nur schwer möglich, weil sie meist um ca. zwei Wochen zurückliegen. Dies resultiert aus folgendem Grund: Je nachdem wie lange es dauert, bis ein/eine PatientIn die ersten Symptome erkennt, vergeht dementsprechend auch eine gewisse Zeitspanne bis der Gang zum Arzt getätigt wird. In der Zwischenzeit kann sich die Krankheit aufgrund fehlender Behandlung durch Ansteckung auf weitere Personen ausbreiten. Zur selben Zeit erkannte ein Software-Entwickler von Google, dass sich mithilfe der Suchmaschine die Ausbreitung der Epidemie nicht nur regional, sondern auch landesweit voraussagen ließe. Google wertete die Begriffe der Suchmaschinenanfragen aus und konnte somit erkennen, wo Personen nach den bereits bekannten Symptomen der Krankheit suchten. Durch diesen Umstand war es Google möglich, den Behörden eine Karte zu liefern, die die Ausbreitung der Krankheit beinahe in Echtzeit darstellen konnte (Mayer-Schönberger & Cukier, 2013).

Einen komplett anderen Ansatz macht sich die Videospieleindustrie zu Nutze. Populäre Spiele wie FarmVille, FishVille oder Mafia Wars werden heute interaktiv und zum Teil über Social-Media Plattformen gespielt. Der Spielehersteller Zynga nutzt diesen Umstand, um mehr über seine SpielerInnen zu erfahren. Auf diese Weise analysiert Zynga, an welcher Stelle im Spielverlauf ein Spiel oft abgebrochen wird, weil es zu schwer oder zu langsam wird. Mit diesen Informationen kann Zynga nun reagieren und die Stellen im Spiel so aufbereiten, dass das Spiel weniger oft abgebrochen wird, um somit ein angenehmeres Gameplay zu generieren. Bei FishVille wurde anhand von Datenanalysen festgestellt, dass transparente Fische sechsmal so oft verkauft wurden wie andere Arten. Daraufhin wurde das Angebot an transparenten Fischen aufgestockt und Zynga machte damit einen beträchtlichen Gewinn (Wingfield, 2011).

### **2.3.3 Marktsituation von Big Data**

Laut Mayer-Schönberger & Cukier (2013) sind im Laufe der Zeit drei Arten von Unternehmen im Bereich Big Data entstanden, die sich in nachfolgenden Kriterien in ihrer Art der Schaffung eines Mehrwertes unterscheiden: Daten, Fähigkeiten und Ideen. In die erste Kategorie fallen jene Unternehmen, die Daten oder zumindest einen Zugang zu diesen haben. Es ist nicht zwingend notwendig, dass Daten oder Datenbeschaffung zum Kern ihres Geschäfts gehören. Möglich ist auch, dass sie schlicht und ergreifend nicht die Mittel besitzen, um aus diesen Daten einen Mehrwert zu generieren. Ein Beispiel dazu ist Twitter. Das Unternehmen, das mit seinem Kurznachrichtendienst zu einem Welterfolg gelangt ist und dessen Server einen beachtlichen Datenverkehr aufweisen, hat sich an zwei externe Firmen gewandt, um den Datenstrom an Dritte zu lizenzieren, um somit einen Mehrwert daraus zu generieren. Die zweite Kategorie beinhaltet jene Unternehmen, die spezielle Fähigkeiten besitzen. Meist sind es Beratungsfirmen oder

TechnologieanbieterInnen, die auf die Analyse von Daten spezialisiert sind und über das nötige Knowhow verfügen. In der dritten und letzten Kategorie befinden sich die Unternehmen, die zwar nicht unbedingt über Daten und Fachwissen verfügen, aber einmalige Ideen haben, wie man aus Daten Wert schöpfen kann. Ein Beispiel dazu ist Pete Warden, ein Mitbegründer von Jetpac. Auf der Website von Jetpac werden NutzerInnen Urlaubsvorschläge aufgrund ihrer Urlaubsfotos unterbreitet, welche die NutzerInnen zuvor hochgeladen haben. Seit dem Jahr 2013 befindet sich Jetpac im Besitz von Google (Mayer-Schönberger & Cukier, 2013).

Im folgenden Abschnitt dieses Kapitels wird auf Grundlage einer Umfrage veranschaulicht, welche Unternehmenssparten Big Data bereits erfolgreich einsetzen und welche noch zögern. Außerdem werden die Beweggründe für Unternehmen aufgezeigt, in das Thema Big Data einzusteigen und ob die Investitionsbereitschaft der Unternehmen mit der fortschreitenden Digitalisierung Schritt hält. Die Online-AnwenderInnenumfrage wurde von Dezember 2014 bis Februar 2015 durchgeführt. Insgesamt beteiligten sich daran 559 TeilnehmerInnen. 37 Prozent stammten dabei aus der DACH-Region und 27 Prozent aus Nordamerika. Mit Ausnahme von Afrika und dem Nahen Osten verteilt sich der Rest gleichmäßig auf weitere europäische Staaten und weltweite Regionen. Die an der Umfrage teilnehmenden Branchen sind IT-Unternehmen, Fertigungsindustrie, Beratungsbranche und Handel (Bange, Grosser, & Janoschek, 2015).

Abbildung 13 zeigt, welchen Stellenwert Daten in Unternehmen besitzen:

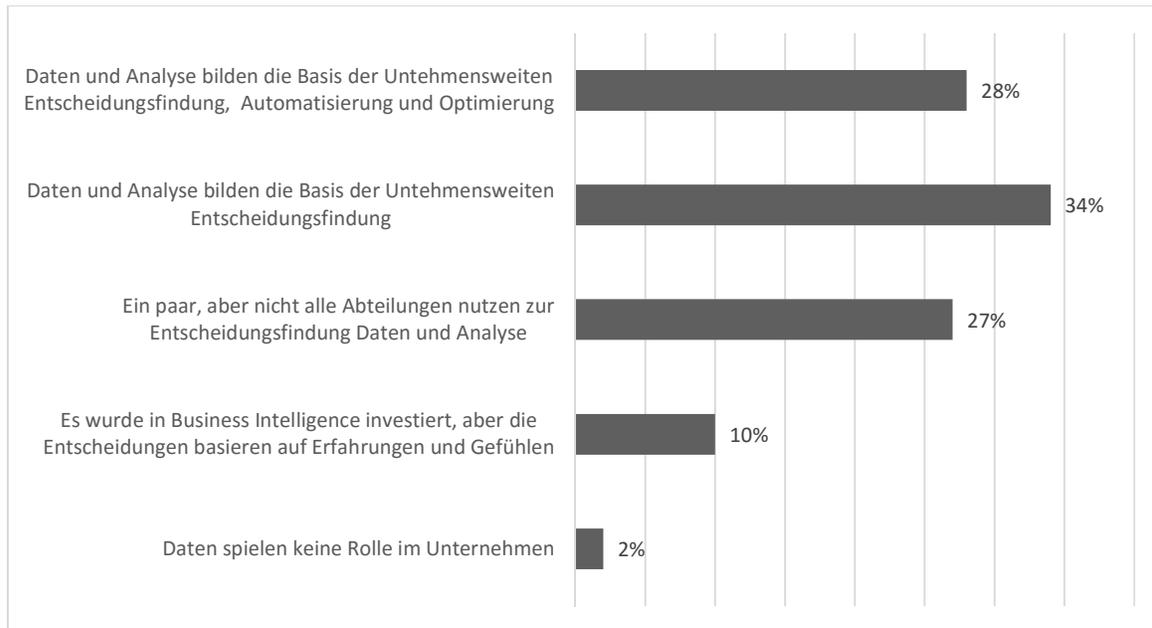


Abbildung 13: Welche Rolle spielen Daten in Ihrem Unternehmen (n=543) (Bange, Grosser, & Janoschek, 2015)

Wie in Abbildung 13 zu sehen ist, nutzen heutzutage mehr als zwei Drittel der Unternehmen Daten zur Entscheidungsfindung. Im Gegensatz dazu gaben nur 12 Prozent der 543 befragten UnternehmerInnen an, dass Daten entweder nur sehr wenig oder gar keinen Einfluss auf die Entscheidungsfindung haben (Bange, Grosser, & Janoschek, 2015).

In Abbildung 14 ist zu sehen, in welchen Sparten Big Data in Unternehmen bereits ein fester Bestandteil geworden ist:

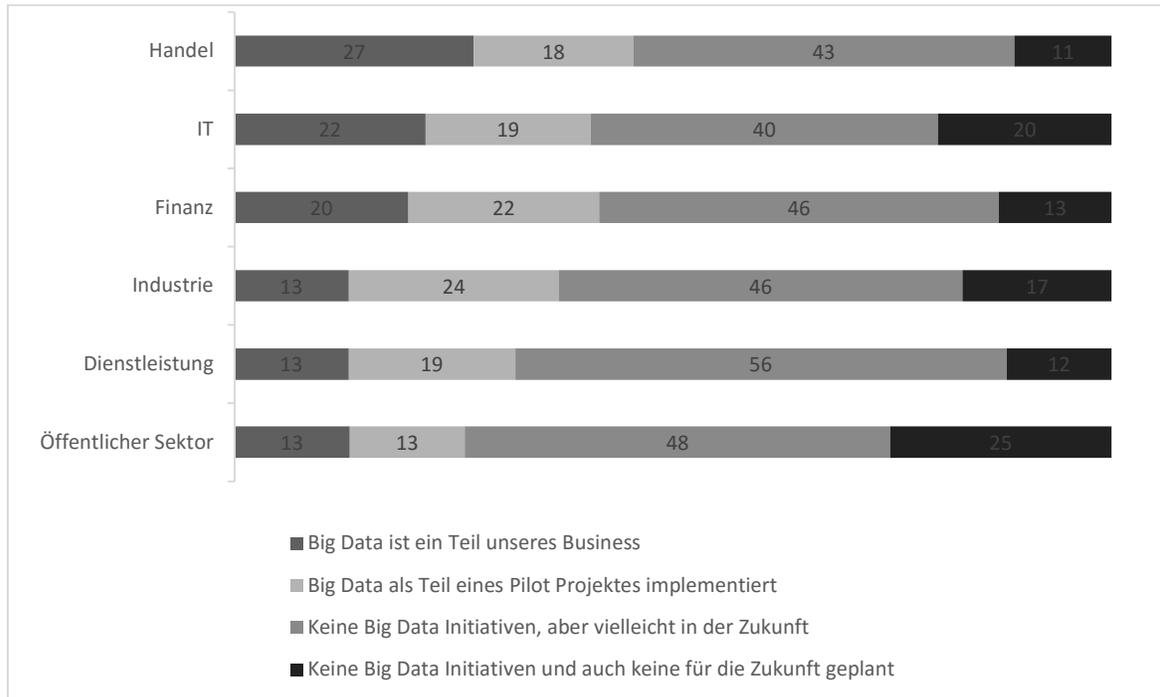


Abbildung 14: Big Data Initiativen gelistet nach Sparten (n=526) (Bange, Grosser, & Janoschek, 2015)

Interessant hierbei ist, dass die Industriesparte mit der vielbeschworenen Industrie 4.0 im Gegensatz zu den anderen Sparten etwas ins Hintertreffen geraten ist. Zumindest ist die Anzahl der Pilotprojekte mit 24 Prozent die höchste im TeilnehmerInnenfeld. Das zeigt wiederum, dass langsam aber doch Bewegung in Industrie 4.0 kommt (Bange, Grosser, & Janoschek, 2015).

Wie sich Big Data seit 2013 in Unternehmen generell entwickelt hat, präsentiert Abbildung 15:

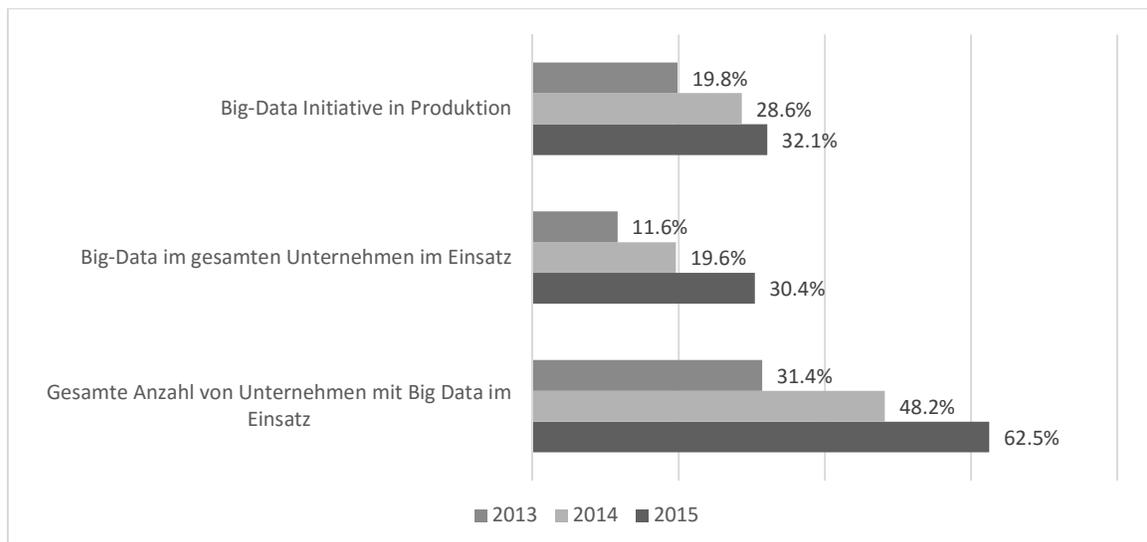


Abbildung 15: Einsatz von Big Data in den Jahren, 2013, 2014 und 2015 (n=526) (Bange, Grosser, & Janoschek, 2015)

Wie man in Abbildung 15 sehen kann, verdoppelte sich der Anteil des Big Data Einsatzes in Unternehmen von 2013 auf 2015. Der Anteil des Big Data Einsatzes der teilnehmenden Unternehmen verdreifachte sich innerhalb von drei Jahren von 11,6 Prozent auf 30,4 Prozent (Bange, Grosser, & Janoschek, 2015).

Im nächsten Punkt wird auf die Motivation von Unternehmen im Hinblick auf den Einsatz von Big Data eingegangen. Welche Probleme und Herausforderungen bewegen Unternehmen überhaupt, sich diesem Thema anzunähern und welchen Erfolg versprechen sich Unternehmen daraus. In Abbildung 16 werden diese Beweggründe nach ihrer Häufigkeit aufgelistet:

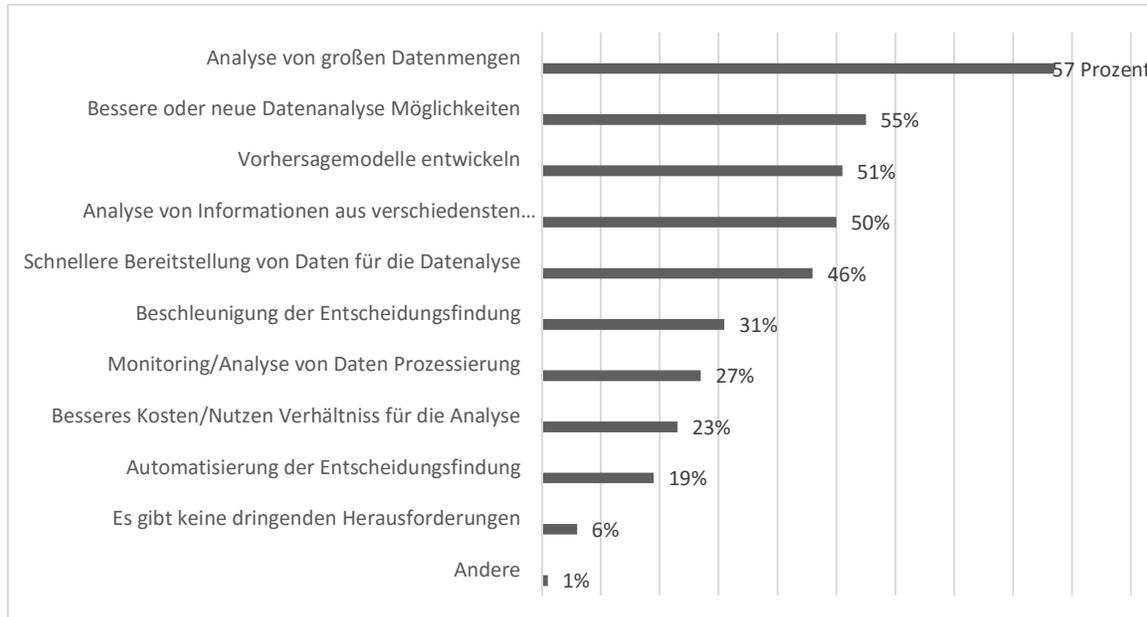


Abbildung 16: Welche Probleme/Herausforderungen wollen Sie mit Big Data Technologien bewältigen? (n=431) (Bange, Grosser, & Janoschek, 2015)

Unter den ersten fünf Plätzen befinden sich jene drei V's, die bereits am Anfang dieses Kapitels behandelt wurden. Hierbei steht die Analyse von großen Datenmengen mit 57 Prozent für den Begriff Volume. Es ist klar ersichtlich, dass es den Unternehmen mittlerweile nicht an Datenmengen mangelt, sondern an einer für sie geeigneten Technologie, um diese Datenmengen zu analysieren. An vierter Stelle befindet sich mit 50 Prozent die Analyse aus verschiedensten Datenquellen, welche für Variety steht. Unternehmen kämpfen heutzutage nicht nur mit einer Flut an Daten, sondern auch mit deren unterschiedlichen Quellen und Formaten. 50 Prozent der Unternehmer wollen mit Hilfe von Big Data diese Herausforderung meistern. Mit 46 Prozent an fünfter Stelle rangiert Velocity und damit die Beschleunigung der Entscheidungsfindung. Wie im ersten Kapitel schon erwähnt, muss die Beschaffung und Auswertung von Daten zeitnah erfolgen, damit das Unternehmen daraus einen Mehrwert generieren kann. Auf dem dritten Rang mit 51 Prozent befindet sich das Thema rund um Vorhersagemodelle. Hier wird klar ersichtlich, dass viele UnternehmerInnen Big Data zur Unterstützung bei der Entscheidungsfindung nutzen wollen – nicht nur was die strategische Ausrichtung des Unternehmens angeht, sondern auch im täglichen Geschäft mit KundInnen wie zum Beispiel in einem Online Shop (Bange, Grosser, & Janoschek, 2015).

In den vorangegangenen Beispielen wurde dargelegt, dass Unternehmen Chancen sehen, mit Big Data verschiedenste Herausforderungen zu bewältigen und darüber hinaus für sich einen Mehrwert zu generieren. Doch wie sieht es tatsächlich mit der Investitionsbereitschaft von Unternehmen aus, wenn es darum geht, Big Data in ihrem Unternehmen zu realisieren? In Abbildung 17 wird die Investitionsbereitschaft von UnternehmerInnen der DACH-Region im Jahr 2014 mit einer Prognose für das Jahr 2017 in drei Größen gegenübergestellt:

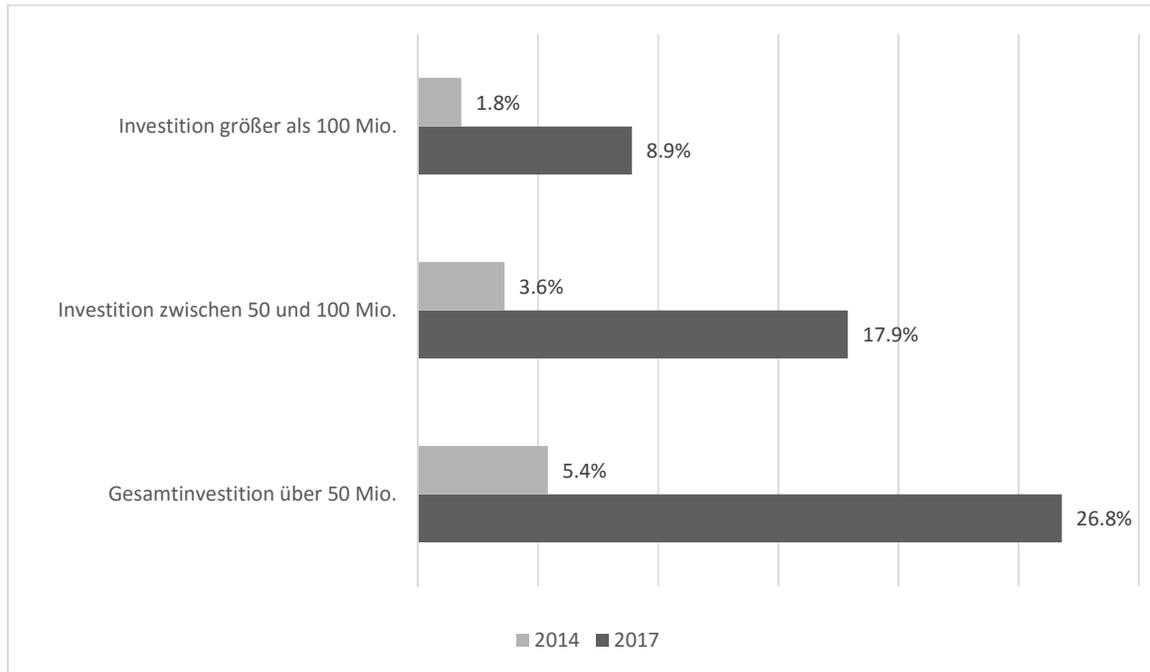


Abbildung 17: Investition in Big Data (n=526) (Bange, Grosser, & Janoschek, 2015)

Die Investitionsbereitschaft steigt von 2014 bis 2017 signifikant. Zählt man alle Investitionen zusammen, wird im Jahr 2017 um knapp 20 Prozent mehr in Big Data investiert als im Jahr 2014.

### 2.3.4 Datenschutz und Risiken

Im Jahr 2007 stellte eine britische Zeitung fest, dass in einem Umkreis von 200 Metern rund um George Orwells Haus mehr als 30 Überwachungskameras installiert sind (Graham, 2007). Neben Kuriositäten wie diesen, werden in diesem Kapitel die Risiken erläutert, die mit Big Data einhergehen. Da das Thema rund um den Datenschutz nicht das Ziel dieser Arbeit ist, wird der Sachverhalt anhand von Beispielen in komprimierter Form dargelegt.

40 Jahre DDR ergaben durch systematische Bespitzelung der BürgerInnen mindestens 39 Millionen Karteikarten und 100 Regalkilometer Akten, die detailliert das Privatleben der betroffenen Personen wiedergeben. Wie an diesen Beispielen zu erkennen ist, gab es Überwachung auch schon vor der digitalen Revolution und ihren Begleiterscheinungen. Nicht nur Geheimdienste mit drei Buchstaben versuchten im Namen der nationalen Sicherheit Privatangelegenheiten vieler BürgerInnen zu überwachen, auch Unternehmen, die täglich Daten

über NutzerInnen sammeln, um diese dann für ihr Unternehmen gewinnbringend zu nützen. Amazon weiß über die Produktvorlieben der NutzerInnen Bescheid, Google über das Surfverhalten von zig Millionen Menschen und Twitter weiß, welche Themen rund um den Erdball gerade im Trend sind. Die voranschreitende Digitalisierung, der Hunger nach Daten von Firmen wie Google, Facebook und Amazon und die immer billiger werdenden Speicher deuten eindeutig darauf hin, dass das Sammeln von Daten stetig mehr werden wird. Entscheidend wird aber immer sein, ob die gesammelten Daten personenbezogen sind und somit im schlimmsten Fall gegen diese Person verwendet werden können (Mayer-Schönberger & Cukier, 2013).

In den folgenden Zeilen folgen nun Beispiele, in denen die Sammlung von Daten eindeutig einen Graubereich erreicht hat:

In Europa und den USA gibt es StromversorgerInnen, die damit beginnen, intelligente Stromzähler, sogenannte Smart Meter, einzusetzen. Diese können den Stromverbrauch rund um die Uhr in Intervallen von sechs Sekunden aufzeichnen. Dabei sind diese Smart Meter nicht nur in der Lage, den klassischen Stromverbrauch aufzuzeichnen wie es traditionelle Stromzähler tun. Mit einem Smart Meter ist es möglich, die Art des Stromverbrauchs zu unterscheiden und somit zu erkennen, welches Gerät wie viel Strom braucht. Ein PC hat zum Beispiel eine unterschiedliche Lastkurve wie eine Küchenmaschine. Mit Hilfe von Mustererkennung ist es in weiterer Folge auch möglich, die Lastkurve einer Lampe zur Cannabiszucht zu erkennen. Da viele Menschen mittlerweile ständig von elektronischen Geräten umgeben sind und diese eben auch Strom zum Betrieb benötigen, können ganze Tagesabläufe und Verhaltensweisen in den eigenen vier Wänden – seien sie legaler oder illegaler Art – aufgezeichnet werden (Quinn, 2014).

Nach diesem Beispiel muss die Frage gestellt werden, ob die Gefährdung der Privatsphäre durch Big Data nur erhöht wird (was zweifelsohne der Fall sein wird) oder auch die Qualität dieser Gefahr steigt.

Der Internetanbieter AOL veröffentlichte im Jahr 2006 einen Datensatz, der aus 20 Millionen Suchanfragen von 657.000 AnwenderInnen bestand. UserInnennamen und IP-Adressen wurden vor der Veröffentlichung anonymisiert. Die Absicht von AOL war, dass WissenschaftlerInnen neue und interessante Erkenntnisse aus diesem Datensatz gewinnen sollen. Jedoch identifizierten innerhalb weniger Tage JournalistInnen der New York Times eine Person aus der Liste. Mit Suchanfragen wie „männliche Singles um die 60“, „Gesundheitstee“ und „Gartenbaufirmen in Lilburn, Ga“ wurde die 62 Jahre alte, verwitwete Thelma Arnold aus Lilburn im US-Bundesstaat Georgia entdeckt (Mayer-Schönberger & Cukier, 2013).

Dieser Fall zeigt, dass die Spuren, die überall im Netz hinterlassen werden, jederzeit zu der Person zurückführen können, die sie dort hinterlassen hat. Die Frage ist, wie damit jetzt und in Zukunft umgegangen werden soll. Die Frage wird lauten, inwieweit der einzelne Mensch es noch selbst in der Hand hat, sich vor DatenschnüfflerInnen zu schützen? Ist es gerechtfertigt, zu verlangen, am iPhone bei jedem Update seitenlange Bestimmungen über geänderte Aspekte im Datenschutz lesen zu müssen? Auch die Frage, ob es sinnvoll ist, Geld zu sparen, indem man mit der Krankenversicherung über eine App die eigene gesunde Lebensweise teilt, darf und muss gestellt werden. Abschließend bleibt in diesem Kapitel zu sagen, dass in Zukunft jede Person darum kämpfen muss, dass Privatangelegenheiten auch privat bleiben.

## 2.4 Big Data Analytik

In diesem Kapitel wird die technische Einordnung von Big Data dargestellt. Big Data findet heute dort statt, wo etablierte Technologien für die Datenanalyse nicht mehr ausreichen. In den folgenden Kapiteln wird ein Blick auf bereits genutzte Big Data Technologien und Verfahren geworfen. Eine der Fragen, die dabei beantwortet werden soll, ist, welche Rolle diese Technologien bis jetzt gespielt haben und welche sie in Zukunft einnehmen könnten.

### 2.4.1 Einordnung

Für viele Unternehmen, die mit steigenden Datenmengen zu kämpfen haben, stellt sich oft nicht die Frage, ob sie in den Bereich Big Data einsteigen wollen, sondern wie dieser am besten zu bewältigen ist. Mit dieser Frage einhergehend muss ebenso festgestellt werden, ob Big Data zwingend etwas mit einer kompletten Neuausrichtung der Infrastruktur und Prozesse zu tun hat oder ob es auch möglich ist, durch klugen Einsatz von bereits bestehenden Technologien dies zu ermöglichen. Aus Kosten- und Zeitgründen wird für viele Unternehmen Letzteres die Herausforderung sein, der sich viele stellen werden. Anknüpfungspunkte bieten hierbei transaktionale Systeme, Data Warehouse, Business Intelligence, Dokumenten-Management- und Enterprise-Content-Management-Systeme. Durch die Nutzung dieser bestehenden Systeme wird es für Unternehmen möglich sein, die mit Big Data verbundenen Prozesse, Technologien und Qualifikationen zu meistern, ohne dabei völlig neue Prozesse entwickeln zu müssen. Es ist somit möglich, Schritt für Schritt diese neue Technologie einzuführen. Im Laufe der Zeit wird sich dann herausstellen, ob die bereits bestehenden Technologien sich auch in Zukunft bewähren können oder ob Neuentwicklungen diesen den Rang ablaufen werden. Big Data kann somit zum einen als logische Weiterentwicklung gesehen werden, zum anderen aber auch als direkte Konkurrenz zu den etablierten Technologien (Bundesverband Informationswirtschaft, 2012). Abbildung 18 zeigt Big Data als Ergänzung und Konkurrenz zur traditionellen IT:

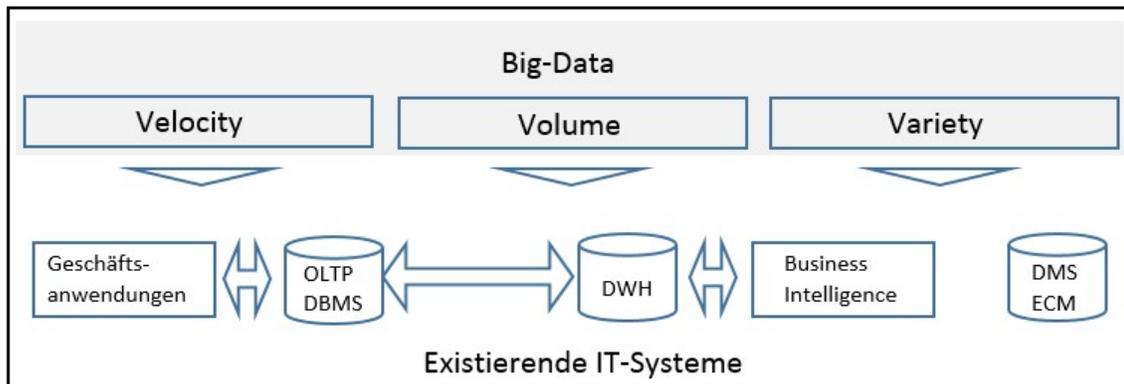


Abbildung 18: Big Data als Ergänzung und Konkurrenz zur traditionellen IT (Bundesverband Informationswirtschaft, 2012)

## 2.4.2 Transaktionale Systeme

Die durch das Internet und die Globalisierung zu Tage getretene Digitalisierung und deren Datenmengen sind vor allem durch transaktionale und analytische Systeme bedient worden. Da sich transaktionale Systeme bereits in großer Anzahl in Unternehmen befinden, ergibt sich dadurch auch ein dementsprechend großes Angebot an Software-Lösungen (Bundesverband Informationswirtschaft, 2012).

Nachfolgend ist in Tabelle 1 eine Gegenüberstellung von transaktionalen Systemen mit Big Data ersichtlich:

| Transaktionale Systeme  | Big Data  |
|---|---|
| Basiert auf strukturierten und gut dokumentierten Daten (z. B. relationalen Tabellenmodellen)   | Starke Betrachtung von unstrukturierten oder polystrukturierten Daten             |
| Integrität durch das Datenmodell sicher gestellt  | Integrität durch den Programmablauf sicher zu stellen                             |
| Optimiert auf transaktionale Datenkonsistenz, parallel konkurrierende Schreib- und Lesezugriffe | Optimiert auf Datendurchsatz und Performanz, asynchrone Schreib- und Lesezugriffe |
| Für bestimmte Geschäftsprozesse und Transaktionen implementiert                                 | Anwendungsfälle meist stark analytischer Art                                      |
| Größe von Dateneinheiten eher gering (z. B. Datensatz)  | Sehr große Dateneinheiten möglich   |
| Datenabfrage und -manipulation meist durch SQL S  | Spezifische Abfragesprachen wie MapReduce   |

Tabelle 1: Vergleich transaktionaler Systeme und Big Data (Bundesverband Informationswirtschaft, 2012)

## 2.4.3 Analytische Systeme: Data Warehouse und Business Intelligence

Durch neue Tools die unter den Schlagwörtern Data Warehouse und Business Intelligence zusammengefasst wurden, ist die steigende technische Anforderung bei der Analyse von Daten frühzeitig adressiert worden. Diese neuen Systeme, welche auf transaktionalen Systemen basieren, erlauben einen Blick auf die Daten, um dadurch direkt in die Entscheidungsfindung des Unternehmens Einfluss nehmen zu können. Aufgrund ihrer Entwicklung werden diese Systeme vornehmlich für strukturierte Daten verwendet, welche auf Tabellenstrukturen und SQL-Abfragen optimiert sind. In ihrem Einsatz sind sie auf nur wenige, leistungsfähige und vor allem teure Rechner optimiert. Um die Geschwindigkeit der Auswertung zu erhöhen, wird auf eine feine Detailierung der Daten absichtlich verzichtet. Meist wird aus den Originaldaten eine periodisch verdichtete und bereinigte Kopie erstellt (ETL-Prozess). Für gewöhnlich werden von diesen analytischen Systemen Datenmengen im niedrigstelligen Terrabyte Bereich verarbeitet. Dies ist eine Größenordnung, die sich bereits bei kleineren Big Data Anwendungen finden lässt. Bei bestehenden Business Intelligence Lösungen kann ein punktueller Einsatz von Big Data die

Auswertung von großen Bestandsdaten beschleunigen oder die Aktualisierungshäufigkeit von Bewegungsdaten drastisch erhöhen. Somit ergeben sich zwangsweise auch ähnliche Aufgabengebiete (Bundesverband Informationswirtschaft, 2012).

Wie in Tabelle 2 aber zu sehen ist, unterscheiden sie sich deutlich in ihren Schwerpunkten:

| Analytische Systeme  | Big Data  |
|--|---|
| Zentrale Datenhaltung, alle Daten müssen exakt zueinander passen | Daten existieren an mehreren Stellen, Ungenauigkeiten sind akzeptabel |
| Qualitativ hochwertige Daten                                     | Einfachheit der Nutzung   |
| Strukturierte, bereinigte und aggregierte Daten                  | Verarbeitung der Rohdaten mit vielen unterschiedlichen Formaten       |
| Wiederkehrende Berichte  | Interaktion in Echtzeit   |
| Periodische Erstellung   | Optimiert für Flexibilität  |
| Zentralistische Organisation                                     | Heterogene, dezentrale Organisation                                   |

Tabelle 2: Vergleich analytischer Systeme und Big Data (Bundesverband Informationswirtschaft, 2012)

#### 2.4.4 Dokumentenmanagement

Für die Verwaltung elektronischer Dokumente und freier digitaler Formate und Inhalte (Texte, Bilder, Druckdokumente) entstanden für Unternehmen auch datenbankgestützte Systeme. Dazu gehörten Dokumenten-Management-Systeme (DMS) und Enterprise-Content-Management-Systeme (ECM), welche den täglichen Workflow in Unternehmen unterstützen sollen. Wie in Tabelle 3 zu sehen ist, stellt der dokumentgestützte Workflow bereits einen wesentlichen Unterschied zwischen DMS und Big Data dar. Big Data dient nicht zur Unterstützung für den rechtssicheren Umgang mit Dokumenten innerhalb eines Unternehmens, sondern bietet Hilfe bei der Analyse von meist flüchtigen Daten (Bundesverband Informationswirtschaft, 2012).

| Dokumenten-Management-Systeme  | Big Data  |
|--|---|
| Rechtssichere Verwaltung von Dokumenten, z. B. Versionierung                             | Verarbeitung flüchtiger Daten   |
| Einbindung in Workflow-Prozesse  | Analyse von Datenmengen   |
| Strukturierte Metadaten zur Beschreibung von elektronischen Inhalten                     | Integration vieler verschiedener und häufig wechselnder Formate           |
| Meist für den unternehmensinternen Einsatz und einen beschränkten Nutzerkreis konzipiert | Prinzipiell unbeschränkter Nutzerkreis vorstellbar                        |
| Zugriff auf einzelne Dokumente aus einem großen Bestand                                  | Auswertung eines großen Datenbestands mit unterschiedlichen Datenformaten |

|  |  |
|--|--|
| Auswertungen und Suchen meist auf Basis von Textindizes oder Metadaten | Anpassbar an sehr unterschiedliche Formate |
|--|--|

Tabelle 3: Vergleich Dokumenten-Management-Systeme und Big Data (Bundesverband Informationswirtschaft, 2012)

### 2.4.5 Einflüsse auf die Software Entwicklung

Eine neue Herangehensweise in der Datenanalyse bringt auch Änderungen in der Software Entwicklung mit sich. Der Prozess, der sich in vielen Unternehmen etabliert hat, läuft wie folgt ab: Der Fachbereich definiert Fragen und legt dabei die Zielbereiche fest, die damit analysiert werden sollen. Im nächsten Schritt entwirft die IT-Abteilung mit Hilfe der Anforderungen aus dem Fachbereich eine Lösung mit der geforderten Funktionalität. Da die Anforderungen an diese Lösung meist sehr konkret sind, resultiert daraus eine feste, nicht veränderbare Struktur. Der Fachbereich hat dann mit dieser Lösung seine Abfragen so lange gegen den Datenbestand gefahren, bis sich neue Ziele ergeben haben und somit ein Re-Design und Re-Build der bestehenden Lösung erforderlich wird. Der Big Data Ansatz für Software Engineering lässt sich heute folgendermaßen zusammenfassen: „Mehr“ Daten liefern bessere Ergebnisse als intelligentere Algorithmen auf bestehenden Daten. Der Grund dafür ist, dass sich im Big Data Umfeld die Qualität der Daten im Vorhinein nur schwer abschätzen lässt. Deswegen wird darauf Wert gelegt, dass möglichst viele Daten untersucht werden. Die Antwort darauf ist eine flexible Analyseplattform, welche es erlaubt, auf Änderungen während des Analyseprozesses einzugehen und eine dementsprechende Anpassung vorzunehmen. Im Vorfeld wird vom Fachbereich und der IT-Abteilung gemeinsam eine interessante Datenquelle für die Analyse identifiziert (Bundesverband Informationswirtschaft, 2012). In Abbildung 19 sind die maßgeblichen Unterschiede für Analyseansätze zwischen einer traditionellen und einer explorativen Analyse abgebildet:

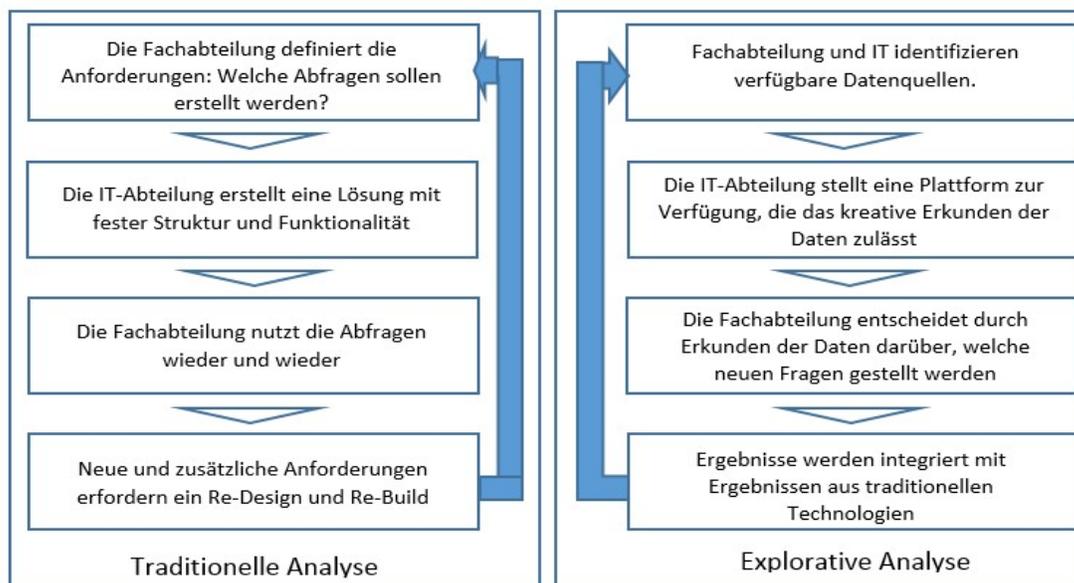


Abbildung 19: Analyseansätze für traditionelle Systeme und Big Data Systeme (Bundesverband Informationswirtschaft, 2012)

## 2.4.6 Einflüsse auf die Anwendungsarchitektur

Proprietäre Big Data Systeme haben sich in den letzten Jahren etwas geöffnet. Eines davon ist das von Google veröffentlichte „MapReduce“ Framework. Dieses Framework ist ein allgemeingültiges Programmiermodell, welches sich für die hochgradig verteilte Datenverarbeitung eignet. Eine Entwicklung, die bis heute Big Data maßgeblich beeinflusst, ist das Open Source Framework Hadoop. Hadoop ging im Jahr 2006 als eine freie Implementierung aus der Firma Yahoo hervor. Mit der Steuerung von Apache ist es möglich, dass Hadoop zusammen mit neuen Open Source Komponenten große Mengen von unstrukturierten Daten (> 1000 Terabyte) auf losen Server-Cluster speichert und analysiert. Durch das Angebot von Lösungen innerhalb einer Cloud ist darüber hinaus die technische und finanzielle Hürde für den Einsatz von Big Data Anwendungen niedrig. Konventionelle IT-Anwendungen stehen aufgrund der zum Teil quelloffenen Verfügbarkeit von Big Data Lösungen unter großem Druck, dieselben Herausforderungen ebenso kostengünstig anzubieten. Neben der Open Source Bewegung für verteilte Datenverarbeitung entwickeln sich darüber hinaus Lösungen zur Speicherung von Daten ohne feste Struktur. Der Überbegriff „NoSQL“ (siehe Kapitel 2.4.8) hat sich hier in Anspielung auf die starren, seit langem bestehenden relationalen Datenbanken gebildet. Die Abfragesprache SQL wird unabhängig dazu trotzdem weiter von Relevanz sein und von AnwenderInnen genutzt werden. Die häufigste Unterscheidung bei AnbieterInnen von Speicherlösungen findet sich in der eingesetzten Speicherstruktur. Hier gilt: Ist die Geschwindigkeit ein Hauptkriterium, bieten HerstellerInnen als Ergänzung In-Memory Technologien an. Dies hat den Vorteil, dass bestehende Anwendungen drastisch schneller werden und neue Disziplinen wie Big Data können unter Umständen mit bestehenden Technologien realisiert werden. Von diesem Geschwindigkeitszuwachs profitieren im Besonderen die analytischen Verfahren (Bundesverband Informationswirtschaft, 2012).

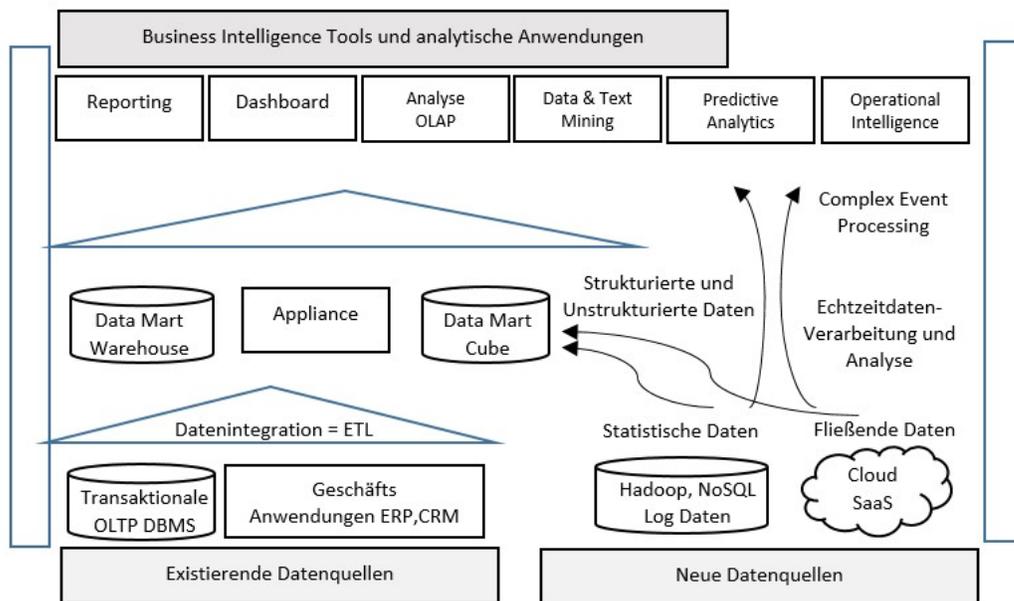


Abbildung 20: Integrierte Anwendungslandschaft mit traditionellen Systemen und Big Data Lösungen (Bundesverband Informationswirtschaft, 2012)

Wie in Abbildung 20 zu sehen ist, wird in Unternehmen in Zukunft eine Kombination aus konventionellen und neuen Technologien zum Einsatz kommen. Es besteht zum Beispiel die Möglichkeit, Hadoop-Cluster zu verwenden, um große Datenmengen für ein Data Warehouse zur Verfügung zu stellen. Zusätzlich lassen sich bestehende Datenbanksysteme durch Integration von In-Memory Technologien beschleunigen (Bundesverband Informationswirtschaft, 2012).

### 2.4.7 Big Data Bausteine

Zu Beginn dieses Kapitels werden die Technologie-Bausteine von Big Data vorgestellt. Die Bausteine werden dabei durch zwei Ebenen hierarchisiert: Schicht und Komponente. Danach wird auf jene zwei Hauptblöcke eingegangen, die in einem Unternehmen hinzugefügt werden, um Big Data zu adaptieren: NoSQL und Hadoop. Um aber zunächst einen Überblick über die Komponenten und deren Begrifflichkeiten zu erhalten, werden in Tabelle 4 die Komponenten ihrer jeweiligen Schicht zugeordnet und darüber hinaus ihre Bedeutung erläutert.

| Schicht      | Komponente                   | Erläuterung  |
|--------------|------------------------------|--|
| Datenzugriff | Batch Processing (MapReduce) | Batch Processing (Stapelverarbeitung) bezeichnet die automatische, sequentielle und vollständige Abarbeitung der in den Eingabedaten enthaltenen Daten. Dabei läuft das Programm nach dem Start völlig selbständig und die zu verarbeiteten Daten werden und Eingriff des Benutzers der Reihe nach abgearbeitet. Abgelegt werden können die Ergebnisse entweder in einer Datenbank oder in Dateien. Ein typisches Beispiel für Stapelverarbeitung bei Big Data ist Apache Hadoop MapReduce |
|              | Streaming Processing und CEP | Stream-Processing bezeichnet den Vorgang der kontinuierlichen Verarbeitung von Eingangsdaten- oder Signalen bei gleichzeitiger Bereitstellung von Ergebnisdaten- oder Signalen.  |
|              | Search & Discovery           | Umfasst das Suchen und Finden von Informationen in meist unstrukturierten Daten. Hauptziel von Search ist es, neben der genauen Antwort auf eine Frage auch ähnliche, verwandte Antworten vorzuschlagen und damit das Entdecken neuer Informationen und Zusammenhänge zu ermöglichen. Der Kern dieser Suchmaschinen sind Algorithmen, welche Text-Matching durchführen und Indizes bauen, welche dann mit Schlagworten durchsucht werden können.   |

|                         |             |   |
|-------------------------|-------------|---|
|                         | Query       | Query zielt darauf ab, strukturierte Daten aus verschiedenen Quell-Systemen und mit verschiedenen Formaten sinnvoll zu verdichten und zu befragen. Dazu wird meist die populäre Abfragesprache SQL genutzt.   |
| Analytische Ausrichtung | Audio/Video | Hier geht es um die Analyse multimedialer Inhalte, die Extraktion von Fakten und das Erkennen von Zusammenhängen. Oft werden Audio-Daten in Text konvertiert und dann mit Text-basierten Verfahren analysiert   |
|                         | Geospatial  | Bei geospatialen Analysen geht es um die Anwendung statistischer und mathematischer Verfahren auf Daten, die einen geografischen oder räumlichen Bezug haben. Diese Daten können mit exakten Global Positioning System (GPS)-Koordinaten versehen sein. Der Bezug auf Ort oder Raum kann auch implizit, indirekt oder ungefähr sein.  |
|                         | Data Mining | Diese Kategorie umfasst statistische Analyseverfahren und Modelle zum Auffinden von Mustern in großen Datenmengen.  |
|                         | Predictive  | Anders als bei traditionellen rückwärts gerichteten Analysen geht es bei Predictive Analytics darum, Entwicklungen vorher zu sehen und so Entscheidungen möglichst zu optimieren.   |
|                         | Web         | Web Analytics umfasst die Messung, die Erfassung, die Analyse und das Reporting von Daten, die es erlauben, den Internet-Auftritt eines Unternehmens zu optimieren, meist mit dem Ziel mehr Umsatz zu erzielen. Social Media Analytics analysiert die Informationen welche Nutzer online Preis geben, wie zum Beispiel Vorlieben zu bestimmten Produkten, Aktivitäten oder dem Freundeskreis, basiert auf Quellen wie Facebook oder LinkedIn. |

|                |                  |  |
|----------------|------------------|--|
|                | Machine Learning | Maschinelles Lernen umfasst eine Vielzahl von Anwendungen und Methoden, in denen Computerprogramme durch iterative Verfahren ihr Wissen stetig erweitern und somit hinzu lernen – in der Regel durch statistische oder logische Analysen gegebener Daten sowie durch die Anwendung rechenintensiver Algorithmen.   |
|                | Text/Semantic    | Diese Kategorie umfasst linguistische und semantische Verfahren, mit deren Hilfe aus Texten relevante Informationen extrahiert, Strukturen erkannt und Verknüpfungen der Daten untereinander sowie mit anderen Datenquellen hergestellt werden, um Business Intelligence auf Text zu ermöglichen.  |
|                | Reporting        | Reports sind systematische (Detail-) Berichte, die eine analytische Aufbereitung, meist in tabellarischer Form, auch als Mischform aus tabellarischen und grafischen Elementen mit Textblöcken, zur Verfügung stellt. In der herkömmlichen Nutzung werden diese häufig gedruckt, bzw. intern via Mail-Anhang verteilt. Quelle sind typischerweise strukturierte Daten, welche zusammengefasst und dann zum Beispiel als Dashboard visualisiert werden. Sie werden wie Dashboards ad-hoc oder periodisch, regelmäßig oder aufgrund spezifischer Anforderung genutzt. Die Kategorie ist nicht neu, sondern hat de facto den Business-Intelligence-Markt begründet. |
| Visualisierung | Dashboards       | Ein Dashboard (englisch für Instrumententafel) besteht aus einem Arrangement von mehreren visuellen Bausteinen mit dem Ziel diese zu konsolidieren, dabei nötigenfalls zu verdichten und damit relevante Informationen auf einem Bildschirm im Überblick darzustellen. Gängige Dashboards erlauben eine mehrschichtige Darstellung (MultiLayer/Linking). Interaktiv kann zwischen einzelnen Schichten navigiert werden und im besten Falle stehen dem Betrachter Selektionen zur Verfügung, die ihn z.B. Zeiträume einschränken oder dargestellte Inhalte variieren lassen.  |

|                  |                        |  |
|------------------|------------------------|--|
|                  | Advanced Visualization | Unter Advanced Visualization (fortgeschrittene Visualisierung) versteht man interaktive visuelle Darstellungen, die komplexe Sachverhalte und Zusammenhänge zu verstehen und kommunizieren helfen. Abgrenzend zu einfachen Dashboards beinhalten fortgeschrittene Visualisierungen auch die Möglichkeit, interaktiv Veränderungen an der Darstellung vorzunehmen, um so sukzessive verschiedene Teilfragen zu analysieren. Fortgeschrittene Visualisierungen umfassen zudem fast immer koordinierte Mehrfachansichten, mittels derer zusammenhängende Teilaspekte der visualisierten Daten zeitgleich dargestellt und zueinander in Beziehung gesetzt werden können. |
|                  | Real-time Intelligence | Der Begriff bezeichnet die kontinuierliche Echtzeit-Animation bzw. visuelle Analyse auf eingehenden Streaming-Daten. Typischerweise wird dazu ein gleitendes Zeitfenster aus dem Datenstrom extrahiert und grafisch dargestellt.   |
| Datenintegration | Daten-Konnektivität    | Konnektoren sind Technologien mit der Aufgabe, Daten aus unterschiedlichen Systemen zugänglich zu machen. Dies können sowohl klassische Datenbanken oder Anwendungen sein, als auch Middleware-Technologien. Konnektoren abstrahieren die spezifische Beschaffenheit des Quellsystems und stellen Daten und Funktionen über Standardschnittstellen (z.B. SQL, Web Services, XML, JMS Messaging Middleware) einheitlich zur Verfügung.  |

|  |   |   |
|--|---|---|
|  | <p>Data Ingestion (ETL, ELT)</p>        | <p>Ingestion (Aufnahme) hat das Ziel, Daten aus verschiedensten Quellen in eine BigData-Lösung zu importieren. Hierbei kann es sich um Echtzeit- (Real-time-) oder Stapel- (Batch-)Daten handeln. In traditionellen Data Warehouses folgt auf diesen DatenImport ein Umwandeln der Daten in drei Schritten: Extract, Transform, Load (ETL) bezeichnet den Prozess der Aufbereitung von Rohdaten für die nachgelagerte Speicherung dieser Daten in einer analytischen Datenbank bzw. einem Data Warehouse. Rohdaten werden normalisiert, validiert und mit einer Struktur versehen, damit also in ein relationales Datenbank-Schema überführt. Mit Hadoop etabliert sich ein neuer Batch-Prozess mit den Schritten Extract, Load, Transform (ELT). Rohdaten landen ohne Struktur in Hadoop und erst zum Zeitpunkt der Analyses erfolgt eine Transformation. Im Bereich von Echtzeitdaten verarbeiten vorgelagerte Messaging- oder Event-Processing-Systeme die eingehenden Datenströme, die dann zur Weiterverarbeitung an NoSQL-Datenbanken oder Hadoop geleitet werden können.</p> |
| <p style="writing-mode: vertical-rl; transform: rotate(180deg);">Daten-Governance und Sicherheit</p> | <p>Identity &amp; Access Management</p> | <p>Identity &amp; Access Management regelt den Zugang zu Daten und beantwortet zwei Fragen:</p> <ol style="list-style-type: none"> <li>1. Wer ist der Benutzer? Wichtig ist die Authentifizierung der Identität, welche typischerweise in einem Verzeichnis aller Mitarbeiter<sup>29</sup> hinterlegt ist.</li> <li>2. Welche Rechte hat der Benutzer? Bei dieser Frage geht es um die Autorisierung zu bestimmten Handlungen. Das Recht, Daten zu lesen oder zu verändern kann zum Beispiel an die Rolle des Benutzers gebunden sein. Identity &amp; Access Management ist nicht neu, wird aber auch in Big Data-Lösungen eingesetzt</li> </ol>  |
|  | <p>Multi-Mandantenfähigkeit</p>         | <p>Big Data-Technologie ist multi-mandantenfähig<sup>30</sup>, wenn sie erlaubt, auf demselben Server oder Cluster mehrere Mandanten<sup>31</sup> zu bedienen, ohne daß diese wechselseitig Einblick in ihre Daten, Jobs, analytischen Modelle, Benutzerverwaltung etc. haben. Die Mandanten werden physisch aus einem Infrastruktur-Pool bedient, während logisch<sup>32</sup> eine Isolation der Ressourcen erfolgt</p>   |

|  |                       |  |
|--|-----------------------|--|
|  | Daten-Verschlüsselung | Personenbezogene und andere besonders schützenswerte Daten werden in der Verschlüsselung so umgewandelt, daß bei einem Verlust der Daten <sup>33</sup> kein Schaden entsteht, da die Informationen ohne den Schlüssel nicht im Klartext lesbar sind. Daten können sowohl für den Transport <sup>34</sup> als auch für die Lagerung verschlüsselt werden. |
|--|-----------------------|--|

Tabelle 4: Bausteine von Big Data (Bundesverband Informationswirtschaft, 2012)

### 2.4.8 NoSQL (Not only SQL)

Durch die neuen Anforderungen an Datenbanken im Internet sind Alternativen zu den herkömmlichen relationalen Datenbanksystemen, die über SQL abgefragt werden, nötig geworden. Die Kernzielgruppe der NoSQL-Bewegung sind datendurchsatzstarke Dienste wie Blogs, Online Shops, Foren und vor allem Suchmaschinen sowie soziale Netzwerke. Besonders für die beiden Letztgenannten sind flexible Strukturen unumgänglich, um mit den riesigen Datenmengen umgehen zu können. Im Gegensatz zu herkömmlichen relationalen Datenbanksystemen, welche über SQL abgefragt werden, beschreibt NoSQL eine Datenbankstruktur ohne festes Datenschema mit alternativen Abfragemechanismen (Kurowski, 2012).

NoSQL lässt sich grob durch folgende drei Punkte definieren:

- Mit NoSQL ist es möglich, in Echtzeit einen immens großen Zulauf von unstrukturierten Daten und Daten ohne Schemen zu erfassen, zu lesen und zu aktualisieren. Zu diesen Daten zählen Clickstreams, Sensor-, Ereignis- und Maschinendaten (WAGmob, 2013).
- NoSQL Datenbanken werden im Hinblick auf Big Data benötigt, um Big Data zu erwerben und im Anschluss daran zu speichern. NoSQL Datenbanken eignen sich besonders für dynamische Datenstrukturen und sind gut skalierbar (WAGmob, 2013).
- Daten, welche in einer NoSQL Datenbank gespeichert sind, verfügen für gewöhnlich über eine sehr hohe Vielfalt (Variety). Der Grund dafür ist, dass die Systeme darauf ausgerichtet sind, alle Daten zu sammeln ohne diese in ein festes Schema zu kategorisieren, um sie anschließend syntaktisch zu analysieren (WAGmob, 2013).

Die nächsten Verwandten zu NoSQL Datenbanken sind objektorientierte Datenbanken bzw. XML Datenbanken. Innerhalb von NoSQL können durch eine Art Datenlogik die Datenbanken in folgende Bereiche eingeteilt werden:

- Dokumentenbasierte Datenbanken: Anstatt in Spalten und Tabellen werden die Daten in sogenannten Dokumenten gespeichert. Bei diesen Dokumenten handelt es sich um eine Ansammlung von Feldern und Werten. JSON und XML sind typische Dokumentstrukturen dafür. In der Datenbank gibt es keine Definition dieser Dokumente (Kurowski, 2012).

- **Key-Value Stores:** Diese sind mit assoziativen Arrays zu vergleichen. Hierbei wird mittels eines Schlüssels auf einen Datensatz mit einem einfachen Typ wie Zahl, Text, Liste gezeigt (Kurowski, 2012).
- **Spaltenorientierte Datenbanken:** Bei den spaltenorientierten Datenbanken werden die Werte physikalisch nicht in Zeilen, sondern in Spalten gespeichert. Wenn ein Datensatz gelesen wurde, ist der nächste Datensatz nicht die folgende Spalte der gleichen Zeile, sondern die gleiche Spalte der nächsten Zeile. Diese Art der Speicherung eignet sich besonders für die Analyse großer Datenmengen, bei denen viele Felder eines Datensatzes unberücksichtigt bleiben. Ein weiterer Vorteil ist die Komprimierung der Datensätze. Da sich in einer Spalte im Normalfall immer derselbe Datentyp befindet, lassen sich diese effizienter komprimieren (Kurowski, 2012).
- **Graphendatenbanken:** Der Ursprung der Graphendatenbanken liegt in der Graphentheorie. Ein Graph ist durch drei Attribute definiert: Knoten, Kanten und Eigenschaften. Ein Beispiel wäre der Streckenplan der Deutschen Bahn. Knoten A ist eine Stadt mit der Eigenschaft Hannover, Knoten B ist auch eine Stadt mit der Eigenschaft Berlin. Die Kante zwischen A und B besitzt die Eigenschaften „Länge=286km“ und „Preis=51euro“ (Kurowski, 2012).

Die Gründe für die Etablierung von NoSQL liegen auf der Hand: Smartphones synchronisieren ihre internen Datenbanken mit anderen Datenbanken, das Internet beherrscht die Kommunikation mit dem Menschen und hat dementsprechend auch neue und andere Anforderungen an Software und Daten. Texte, Bilder, Videos werden ausgetauscht, Kommentare, Bewertungen und Nachrichten werden geschrieben. Diese Informationen können nicht mehr in feste Schemen gepresst werden bzw. will der Mensch dies nicht mehr. Das System soll sich an den Menschen anpassen. Aus diesem Grund braucht es diese flexible Struktur. NoSQL wird relationale Datenbanken keineswegs ersetzen, aber immerhin um eine zusätzliche Schicht erweitern (Kurowski, 2012).

## **2.4.9 Apache Hadoop**

Das 2011 entstandene Open Source Projekt ist ein in Java geschriebenes Framework für skalierbare, verteilt arbeitende Software. Der von Google stammende Algorithmus „MapReduce“ war hierfür die Basis. In einem Satz lässt sich das Apache Projekt wie folgt zusammenfassen: Es handelt sich um ein Konglomerat von Algorithmen und Dateisystemen für das verteilte Rechnen mit großen Datenmengen auf verclusterten Computern. Demzufolge soll es in der Lage sein, mit intensiven Rechenprozessen und vielen Daten umgehen zu können. Darüber hinaus kann Hadoop mit verschiedensten Datentypen in unterschiedlicher Quantität arbeiten – egal ob strukturiert oder unstrukturiert, mit oder ohne Schema sowie mit allen Arten von analytischen Applikationen. Aufgrund der Apache Lizenzierung besteht nicht die Gefahr eines Lock-in-Effekts. Hadoop arbeitet mit Flat Files in verschiedensten Formaten ohne Schema (WAGmob, 2013).

Wie in Abbildung 21 zu sehen ist, werden die Dateien automatisch in Blöcke aufgeteilt:

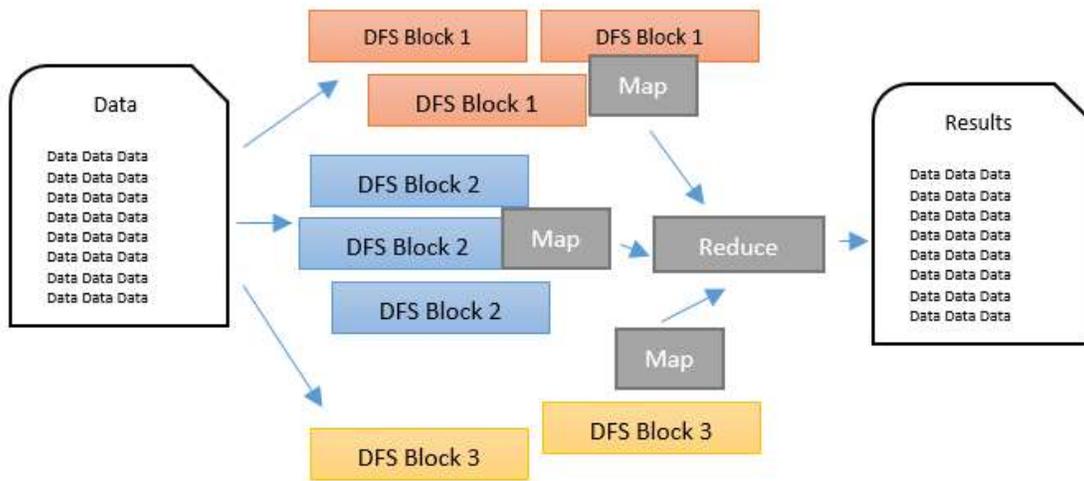


Abbildung 21: Aufteilung der Blöcke (WAGmob, 2013)

Innerhalb dieser Blöcke werden dann spezifische Jobs ausgeführt. Dabei ist jeder Job in einzelne Tasks aufgeteilt, wobei jeder Task entweder eine Map oder Reduce Aktion darstellt (WAGmob, 2013).

Hadoop MapReduce ist ein Weg, um jeden Request in kleinere Requests zu splitten. Diese werden wiederum zu kleinen Servern gesendet, um so die zur Verfügung stehende Rechenleistung bestmöglich auszunutzen (WAGmob, 2013).

Die Aufgaben von Map und Reduce lassen sich in Tabelle 5 wie folgt unterscheiden:

|   |   |        |
|---|---|--------|
| Iterate over a large number of records  | ↓ | Map    |
| Extract something of interest from each |   |        |
| Shuffle and sort intermediate results   | ↓ | Reduce |
| Aggregate intermediate results          |   |        |
| Generate final output                   |   |        |

Tabelle 5: Aufgaben von Map und Reduce (WAGmob, 2013)

Das Hadoop Framework besteht aus zwei Hauptkomponenten. Zum einen aus dem Hadoop Distributed File System (HDFS) und zum anderen aus der Execution Engine (MapReduce) (WAGmob, 2013).

HDFS ist ein File System zur Speicherung von sehr großen Datenmengen. Die Daten werden hierbei auf mehrere Datenblöcke verteilt. Es werden mehrere Kopien von den jeweiligen Datenblöcken angelegt, um so eine höhere Zuverlässigkeit und Geschwindigkeit zu erreichen. Ein Datensatz wird beispielsweise in viele kleine Datensätze gesplittet und jeder dieser Datensätze dann auf die anderen Server repliziert (WAGmob, 2013).

Grafisch stellt sich HDFS und die Replikation in Abbildung 22 wie folgt dar:

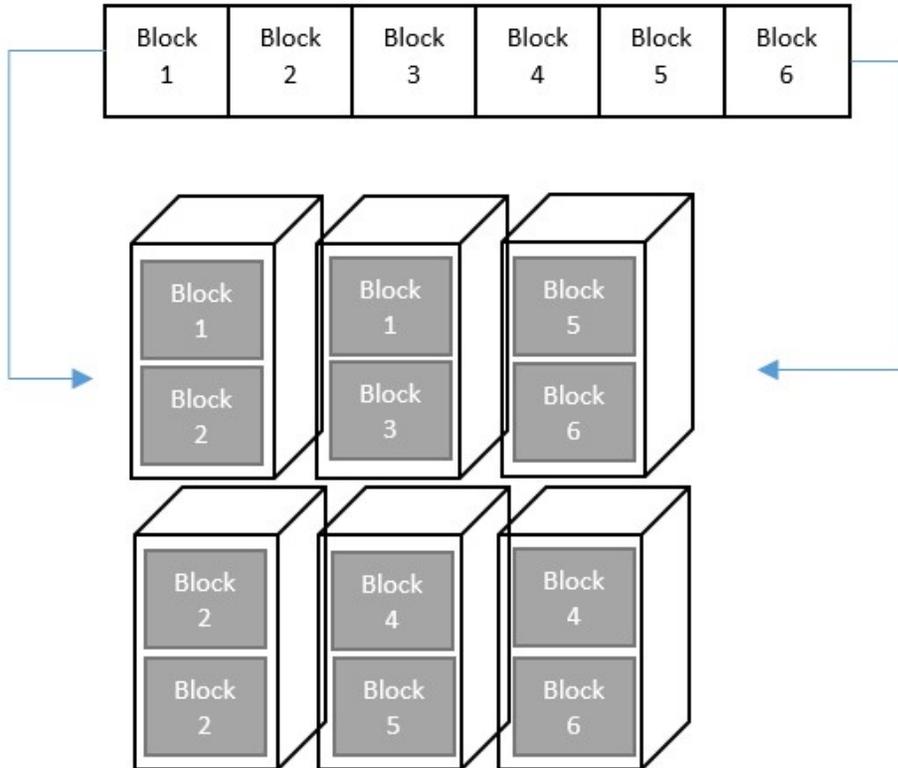


Abbildung 22 HDFS und Replikation (WAGmob, 2013)

Wie in obenstehender Grafik zu sehen, werden die Daten in Blöcken organisiert. Diese Blöcke stellen im HDFS die kleinste Einheit an Daten dar, die gelesen oder geschrieben werden können (WAGmob, 2013).

Typischerweise sind diese Blöcke 64 Megabyte groß und werden auf drei Knoten repliziert:

- NameNode (NN): Verwaltet Metadaten über Daten und Blöcke. Sind Master Knoten (WAGmob, 2013).
- SecondaryName Node (SNN): Stellt das Backup vom NN zur Verfügung (WAGmob, 2013).
- DataNodes (DN): Ist für die Speicherung und Bereitstellung der Blöcke zuständig. Sind Slave Knoten (WAGmob, 2013).

Die eingehenden Daten werden von einem Masterknoten bearbeitet, Ablagen von Dateien organisiert und anfallende Metadaten gespeichert. Dabei ist zu beachten, dass die Input Datei während der gesamten Lebenszeit im HDFS nicht modifiziert wird. HDFS unterstützt dabei Dateisysteme mit mehreren 100 Mio. Daten (speicherguide.de GbR, 2012).

Die Datenbank besteht aus folgenden Komponenten:

- Hive: Bei Hive handelt es sich um eine Erweiterung der Software um die Anfragesprache QL. QL erlaubt eine SQL-basierte Abfrage, welche eine Benützung der SQL-Syntax ermöglicht. Dadurch können auch Datenbankinhalte entsprechend bearbeitet werden (speicherguide.de GbR, 2012).
- Pig: Mit der Software-Komponente Pig, die auch Bestandteil des Frameworks ist und eine Programmiersprache sowie einen Compiler enthält, ist es möglich, High Level Abfragen auf Datenbeständen durchzuführen. Auch komplexe Analysen lassen sich parallel prozessieren (speicherguide.de GbR, 2012).
- HBase: Bei HBase handelt es sich um eine skalierbare Datenbank innerhalb eines Hadoop-Clusters zur Verwaltung sehr großer Datenbestände. HBase basiert auf einer Integration von Googles „Big Table“. Besonders geeignet ist diese Datenstruktur für Daten, die sich selten verändern, aber sehr oft ergänzt werden (speicherguide.de GbR, 2012).

#### **2.4.9.1. Szenarien und Anwendungsbeispiele für Hadoop**

Folgende Szenarien werden für Hadoop vorgeschlagen:

- Hadoop als eine ETL- und Filterplattform: Eine der Herausforderungen im Umgang mit großen Datenmengen ist die Unterscheidung und Extraktion von qualitativen Daten mit jenen, die nicht der Auswertung dienen. Das bereits erwähnte MapReduce Verfahren, welches in Hadoop integriert ist, kann hierbei eine Hilfestellung bieten (WAGmob, 2013).
- Hadoop zum Erforschen der Daten: Sobald sich die Daten im MapReduce Cluster befinden, können verschiedenste Werkzeuge eingesetzt werden, um Daten dort zu analysieren, wo es auch Sinn macht (WAGmob, 2013).
- Hadoop als Archiv: Historische Daten werden aus Kosten und Performance Gründen oft nicht im aktuellen SAN (Storage Area Network) gespeichert. Meist befinden sich diese auf sekundären Datenträgern und sind für Analysen somit nicht sofort zugänglich. Diese dann wieder zu integrieren bedeutet für Unternehmen ein großes Maß an Aufwand hinsichtlich Zeit und Kosten. Mit dem verclusterten System von Hadoop ist es möglich, diese kostengünstig für weiterführende Analysen „aktiv“ zu halten (WAGmob, 2013).

Nicht überraschend ist, dass mit Facebook ein soziales Netzwerk eine Hadoop Datenbank in Verwendung hat. Bereits im Jahr 2012 gehörte die Datenbank mit mehr als 100 Petabyte zu den größten der Welt. Bis zum Jahr 2014 wuchs diese auf 300 Petabyte an (Vagata & Wilfong, 2014). Auch Netflix setzt auf Hadoop. Hier ist ein stündlich geplanter Hadoop Job für das Importieren der Log-Files in die Datenbank zuständig (WAGmob, 2013).

#### **2.4.10 Big Data as a Service**

Der Begriff BDaaS umfasst eine Vielzahl von Möglichkeiten, um Big Data in der Cloud zu nützen. Das reicht von der Bereitstellung der nötigen Infrastruktur über die Software, Analysetools, um

Berichte zu erstellen, bis hin zu konsultativen Services. Durch den Mix aus den bereits bekannten Services „Software as a Service“, „Platform as a Service“ und „Data as a Service“ entstand der neue Begriff „Big Data as a Service“. Wie auch bei anderen Services, steht BDaaS für die Bereitstellung und Nutzung von Hard- und Software Services, ohne dass sich diese im eigenen Besitz befinden. Neben dem Vorteil der Kostenersparnis für die Infrastruktur, können sich die Unternehmen weiterhin auf ihr Kerngeschäft konzentrieren, ohne dafür Ressourcen abstellen zu müssen. Im Grunde stellen BDaaS ProviderInnen ein ready-to-go Paket bereit. Dieses umfasst neben dem Cloud-basierten Storage auch eine analytische Engine um Auswertungen durchzuführen. Das Zahlungsmodell basiert dann entweder auf der Zeit der Benützung oder der Menge der ausgewerteten Daten. Ein weiterer Vorteil von BDaaS ist, dass der/die AnbieterIn für die Sicherung von Daten zuständig ist (Marr, 2015).

Ein Beispiel dafür ist das von IBM gestartete „Analytics for Twitter“ Service. Dieser stellt Unternehmen einen Zugang zu den Daten und Analysen von Twitters 500 Millionen Tweets pro Tag und 280 Millionen aktiven UserInnen pro Monat zur Verfügung. Der Service stellt analytische Tools und Anwendungen zur Verfügung, um Ordnung und Struktur in große Mengen an Daten zu bringen und so Analysen zu ermöglichen (IBM, 2015).

Ein weiteres Beispiel ist der Landmaschinenproduzent John Deere. Dieser stattet seine Traktoren mit Sensoren aus, welche Informationen über die technischen Daten des Traktors sowie auch Informationen über Ernteerträge an MyJohnDeere.com sendet (John Deere GmbH & Co.KG, 2016).

Laut dem Forbes Magazin gehen Schätzungen davon aus, dass der Anteil an der Nutzung von X-as-a-Service von 15 Prozent im Jahr 2015 auf 35 Prozent im Jahr 2021 steigen wird. Daraus wird geschlussfolgert, dass der globale Big Data Markt bis zu diesem Zeitpunkt 88 Milliarden Dollar wert sein wird, wobei der Anteil von BDaaS bei ungefähr 30 Milliarden Dollar liegt (Marr, 2015).

## 2.5 Klein- und Mittelunternehmen

Da im Rahmen dieser Masterarbeit die Möglichkeiten einer Big Data Analyse für Klein- und Mittelunternehmen analysiert werden, wird in vorliegendem Kapitel die Definition dieser erläutert.

Die europäische Kommission definiert ein Unternehmen wie folgt:

*Als Unternehmen gilt jede Einheit, unabhängig von ihrer Rechtsform, die eine wirtschaftliche Tätigkeit ausübt. Dazu gehören insbesondere auch jene Einheiten, die eine handwerkliche Tätigkeit oder andere Tätigkeiten als Einpersonen- oder Familienbetriebe ausüben, sowie Personengesellschaften oder Vereinigungen, die regelmäßig einer wirtschaftlichen Tätigkeit nachgehen (Liikannen, 2003).*

Unternehmen, die zur Größenklasse der Kleinstunternehmen sowie der kleinen und mittleren Unternehmen (KMU) zählen, beschäftigen weniger als 250 Personen und erzielen entweder einen Jahresumsatz von 50 Mio. EUR oder die Jahresbilanzsumme beläuft sich höchstens auf 43 Mio. EUR. Ein kleines Unternehmen wird in der Kategorie innerhalb der KMU als ein

Unternehmen definiert, das die Grenze von 50 Beschäftigten nicht überschreitet und dessen Jahresumsatz bzw. Jahresbilanz nicht über 10 Mio. EUR liegt. Als mittleres Unternehmen werden jene Unternehmen bezeichnet, deren MitarbeiterInnenzahl sich zwischen 50 und 249 Personen bewegt und deren Umsatz bzw. Bilanzsumme eine Grenze von 50 Mio. EUR bzw. 43 Mio. EUR nicht unterschreitet (Liikannen, 2003).

Zur Veranschaulichung werden die soeben genannten Zahlen und jene der Kleinst- und Großunternehmen in Tabelle 6 dargestellt:

|                             | <b>MitarbeiterInnen</b> | <b>Umsatz</b> | <b>Bilanzsumme</b> |
|-----------------------------|-------------------------|---------------|--------------------|
| <b>Kleinstunternehmen</b>   | Bis 9                   | ≤ 2 Mio EUR   | ≤ 2 Mio. EUR       |
| <b>Kleinunternehmen</b>     | 10 bis 49               | ≤ 10 Mio EUR  | ≤ 10 Mio. EUR      |
| <b>Mittlere Unternehmen</b> | 50 bis 249              | ≤ 50 Mio EUR  | ≤ 43 Mio. EUR      |
| <b>Großunternehmen</b>      | ab 250                  | > 50 Mio EUR  | > 43 Mio. EUR      |

Tabelle 6: Gegenüberstellung der Grenzen von Unternehmensbezeichnungen (Wirtschaftskammer Österreich, 2016)

Der Faktor MitarbeiterInnen wird hierbei als ein Hauptkriterium bezeichnet. Als zweites Hauptkriterium zeichnet sich der Umsatz aus. Dieser soll die Bedeutung, die Leistung und die Marktstellung des Unternehmens berücksichtigen. Die Bilanzsumme wird hinzugezogen, um die Gesamtheit des Unternehmenswertes zum Ausdruck zu bringen. Dies ergibt sich daraus, dass Handelsunternehmen naturgemäß einen viel höheren Umsatz erwirtschaften als Produktionsunternehmen. Ein weiterer wichtiger Wert zur Einstufung der Unternehmen ist jener der Eigenständigkeit bzw. Unabhängigkeit. Dieses Kriterium ist insofern von Bedeutung, um Klein- und Mittelunternehmen aus dieser Kategorie auszuschließen, die einem Konzern angehören. Naturgemäß verfügen diese über Mittel und Unterstützungsleistungen, die ihre gleich großen und unabhängigen KonkurrentInnen nicht haben. Ebenso sollen damit rechtliche Gebilde oder Zusammenschlüsse von KMUs ausgeschlossen werden. Ein Richtwert ist, dass sich weniger als 25 Prozent der Kapitalanteile oder Stimmrechte in Fremdbesitz befinden. (Wirtschaftskammer Österreich, 2016).

Im Zuge der Recherche zu Klein- und Mittelunternehmen wurde versucht, zu eruieren, ob es dafür im E-Commerce eine gesonderte Definition und somit andere Regeln zur Unterteilung gibt. Dies kann hiermit verneint werden. In allen bisher untersuchten Publikationen zu diesem Thema wurde für die Definition der klein- und mittelständischen Unternehmen immer jene herangezogen, welche in Kapitel 2.5. angeführt wurde.

### 2.5.1 Chancen für Big Data in Klein- und Mittelunternehmen

Big Data wurde von der IT-Branche zu einem der vorherrschenden Trends erklärt. Doch in den meisten Studien zu diesem Thema sind es große Unternehmen, die darin erwähnt und analysiert werden. Klein- und Mittelunternehmen befinden sich demgemäß hier in der Unterzahl. In den

folgenden drei Punkten wird erläutert, warum auch KMUs auf diesen Trend setzen sollten und welche spezifischen Vorteile daraus entstehen können.

### **2.5.2 Effizientere Unternehmensführung**

Eine effiziente Nutzung von Big Data schafft einen Überblick über betriebliche Szenarien. Durch die Steigerung von Transparenz wird der Informationsgehalt von Daten erhöht, wodurch Zusammenhänge und Abhängigkeiten besser zu erkennen sind. Daraus resultieren fundierte Entscheidungsgrundlagen, um Geschäftsprozesse effizienter zu gestalten. Ein weiterer Aspekt ist die Kalkulation von Risiken. Dabei lassen sich Eintrittswahrscheinlichkeiten oder auch die Relevanz dieser exakt bewerten. Darüber hinaus können damit Zusammenhänge zwischen unterschiedlichen Risikofaktoren identifiziert werden, um so Optionen aufzuzeigen, die trotz nachweisbaren Risiken im Bereich des Profitablen sind. Die Vermeidung von Lieferengpässen setzt eine präzise Prognose von Bedarf und Nachfrage voraus. Mit Hilfe von Big Data Analysen lässt sich eine fundierte Entwicklung des Absatzes prognostizieren, um Umsatzverluste durch Überschuss oder Mangel zu vermeiden und den finanziellen Erfolg des Unternehmens sicherzustellen (PricewaterhouseCoopers AG, 2013).

### **2.5.3 Massenindividualisierung durch Big Data**

Ein Stichwort das immer wieder in Zusammenhang mit Big Data und Online Shops fällt, ist der Traum vom IT-gestützten Tante-Emma-Laden. Um die KundInnenbeziehung eines Tante-Emma-Ladens in einem Online Shop zu erzeugen, ist es nötig, die Gewohnheiten und Vorlieben der BesucherInnen zu kennen (Fraunhofer IAIS, o.J.).

Ein Beispiel dazu ist die dynamische Preisgestaltung. Diese ermöglicht es, die Preise in Echtzeit automatisch und dynamisch nicht nur an die aktuelle Markt- und Wettbewerbssituation anzupassen, sondern auch an die individuelle Kaufkraft des Kunden/der Kundin. So ist es möglich, tages- oder wochentagspezifische „Prime-Time“ Preise anzuwerben und so die Auslastung der internen Logistik zu verbessern. Auch aus dem Bereich des Social-Media können Verhaltensmuster erkannt, simuliert und somit eine Prognose des zukünftigen Nutzungsverhalten erstellt werden. Ebenso können Marketingkampagnen genutzt werden, um auf die KundInnen noch personalisierter einzugehen (PricewaterhouseCoopers AG, 2013).

### **2.5.4 Big Ideas + Big Data = Big Business**

Für Unternehmen wird es in der globalen Welt immer wichtiger, sich durch die Positionierung in Marktlücken und mit passgenauen Angeboten für KundInnen vom Wettbewerb abzuheben. Den Überblick über die Märkte zu erhalten wird dabei immer schwieriger oder gar unmöglich. Mit Hilfe von Big Data ist es Unternehmen nun möglich, durch exakt abgestimmte Positionierung in verschiedensten Regionen Fuß zu fassen und somit das Geschäftsfeld zu erweitern. Dabei können verborgene Marktpotenziale identifiziert und ausgeschöpft werden (PricewaterhouseCoopers AG, 2013).

## 3 UNTERSUCHUNGSDESIGN

Im empirischen Teil dieser Arbeit wird zunächst das Untersuchungsdesign und die gewählte Methode zur Erhebung der Daten für den Handlungsleitfaden vorgestellt und begründet. Im Hinblick auf die vorliegende Untersuchung wurde das ExpertInneninterview als sinnvollste Methode erachtet. In den nachfolgenden Kapiteln wird in Folge auch die Durchführung der ExpertInneninterviews sowie die Erstellung des Interviewleitfadens erläutert und im Anschluss der Prozess der Interviewanalyse vorgestellt.

### 3.1 ExpertInneninterviews

Im Zuge dieser Arbeit wurden sechs ExpertInnen einer Befragung unterzogen. Das Ziel dieser ExpertInneninterviews war es, Wissen von Personen einzuholen, die bereits mehrjährige Erfahrung im Bereich Digital Analytics besitzen, um damit auf Basis der Literaturrecherche im theoretischen Teil einen Handlungsleitfaden für Klein- und Mittelunternehmen im E-Commerce zu erstellen.

Das ExpertInneninterview als Untersuchungsmethode wird in verschiedensten Forschungsfeldern eingesetzt. Dies kann zum einen als Methodenmix geschehen oder als eigenständiges Verfahren. Bei ExpertInneninterviews bildet nicht die Person an sich den Gegenstand der Analyse. Das heißt, dass die Orientierung und Einstellung in Bezug auf das kollektive oder individuelle Leben nicht direkt Fokus der Befragung sind (Meuser & Nagel, ExpertInneninterviews - vielfach erprobt, wenig bedacht : ein Beitrag zur qualitativen Methodendiskussion, 1991).

Innerhalb der Sozialwissenschaften findet eine hitzige Debatte über die Objektivität von qualitativen Methoden statt. Der Grund dafür ist, dass ein Interview immer soziale Interaktion zwischen zwei Personen voraussetzt und diese soziale Interaktion ist oft mehr als reiner Wissensaustausch. In Interviews werden von den Gesprächspartnern persönliche Sichtweisen eingebracht sowie auch Machtverhältnisse geregelt. Zusätzlich muss der Funktionskontext beachtet werden, denn der Befragte/die Befragte nimmt im Zuge des Interviews eine Rolle im betrieblichen Kontext ein und empfängt den Interviewer/die Interviewerin somit unter bestimmten Voraussetzungen. Der Ursprung des Problems ruht also in der Gegenwart des Interviewers/der Interviewerin, denn es ist nicht abzuschätzen, ob der Befragte/die Befragte auf Fragen eines anderen Gegenparts anders reagiert und geantwortet hätte (Mieg & Brunner, 2001).

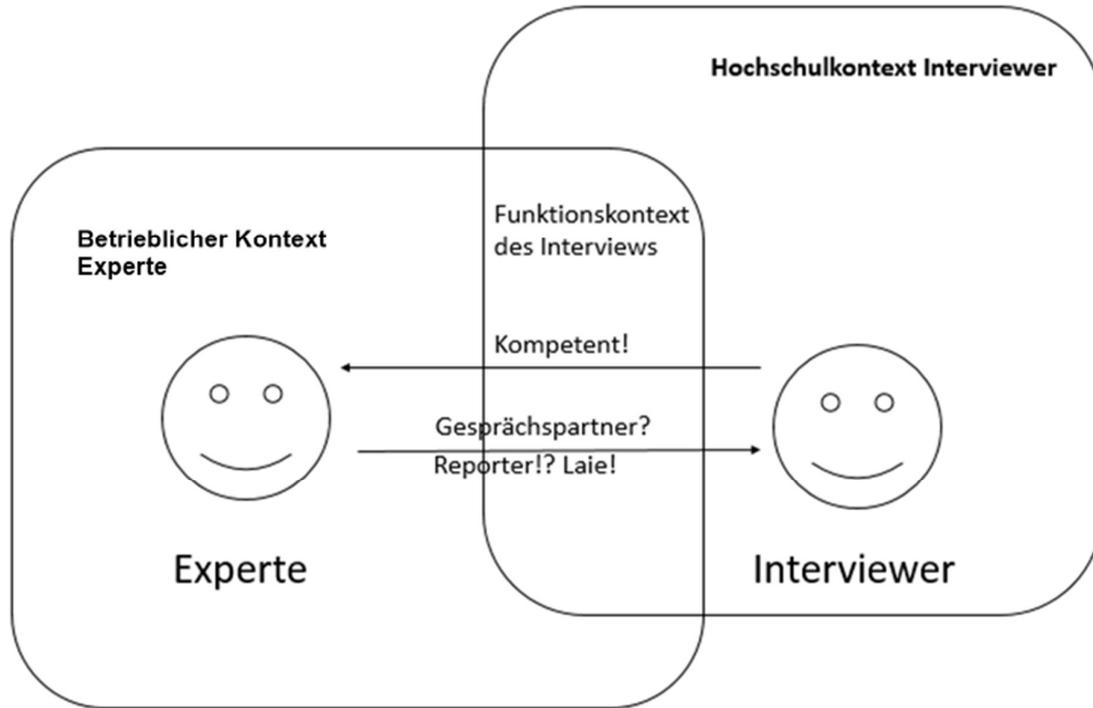


Abbildung 23: Das Interview als soziale Interaktion. Der Interviewer sollte ein kompetentes Gesprächsklima bieten (Mieg & Brunner, 2001).

In diesem Zusammenhang wird von Gläser und Laudel (2010) der Begriff ExpertIn wie folgt definiert:

*Experte beschreibt die spezifische Rolle des Interviewpartners als Quelle von Spezialwissen über die zu erforschenden sozialen Sachverhalte. Experteninterviews sind eine Methode, dieses Wissen zu erschließen (Gläser & Laudel, 2010).*

Eine Person mit relevantem Fachwissen im entsprechenden Untersuchungsgebiet kann somit als PartnerIn für ein ExpertInneninterview herangezogen werden. Das Ziel des ExpertInneninterviews ist es, dieses Fachwissen zu erschließen, um es in weiterer Folge für sich nutzen zu können. Aus diesem Grund werden Einzelbefragungen in den meisten Fällen als zweckmäßig betrachtet da es naturgemäß einfacher ist, sich während einer Befragung auf eine Person zu konzentrieren. Bei nichtstandardisierten Interviewmethoden ist es darüber hinaus empfehlenswert, im Vorfeld einen Interviewleitfaden anzufertigen. Eine vorgefertigte Fragenliste stellt sicher, dass im Zuge des Interviews der/die ExpertIn zu allen relevanten Themen Informationen abgeben kann. Im Gegensatz dazu wird von narrativen oder freien Interviews abgeraten, da relevante Themen oft nur vage oder peripher behandelt werden (Gläser & Laudel, 2010).

Die Besonderheiten des ExpertInneninterviews lassen sich auch durch die Gegenüberstellung mit dem journalistischen Interview in Tabelle 7 aufzeigen:

| <b>Journalistisches Interview</b>   | <b>ExpertInneninterview</b>   |
|---|---|
| Motiv: öffentliches Interesse   | Motiv: sachliches Interesse   |
| investigativ: Tatsachen werden ans Tageslicht gebracht (notfalls auch gegen den Willen der Befragten) | konstruktiv: Sachzusammenhänge werden erhellt (nie gegen den Willen der Befragten)                        |
| Motivation der Befragten im Interview: Meinungsäußerung   | Motivation der Befragten im Interview: Sachmotivation   |
| KO-Kriterium Befragter/Befragte: Person / Meinung nicht von öffentlichem Interesse                    | KO-Kriterium Befragter/Befragte: Befragter/Befragte ist kein Experte/keine Expertin / hat keine Erfahrung |
| KO-Kriterium Interviewer/Interviewerin: Interviewer hat keine Medienmacht                             | KO-Kriterium Interviewer/Interviewerin: fachliche Inkompetenz   |

Tabelle 7: Gegenüberstellung journalistisches Interview mit ExpertInneninterview (Mieg & Brunner, 2001)

Im Gegensatz zum journalistischen Interview ist der Sinn von ExpertInneninterviews, ein Gesprächsklima zu erzeugen, das Sachverhalte erhellt und Zusammenhänge aufzeigt. ExpertInnen sind zudem eher geneigt, ihr Wissen in einem Interview zu teilen, da ihre Motivation davon geleitet wird, die gewonnenen Erkenntnisse in ihrem Fachgebiet mit der Öffentlichkeit zu teilen (Mieg & Brunner, 2001).

### 3.2 ExpertInnenstatus

Die Einigkeit darüber, wer oder was ExpertInnen sind, ist in den Sozialwissenschaften nur sehr gering ausgeprägt. Zum einen werden Begriffe wie ExperteIn, BeraterIn oder GutachterIn oft wie Synonyme verwendet und zum anderen wird der Ausdruck des/der ExperteIn oft ohne weitere Diskussionen übernommen. In der Soziologie wird von einer Vielzahl von ExpertInnen-Funktionen gesprochen. Dabei geht es vor allem um den Status der ExpertInnen oder von Rollenzuweisungen. Daraus geht hervor, dass es sich beim ExpertInnentum um ein soziales Konstrukt aus Erwartungen und Zuschreibung von besonderen Fähigkeiten an ExpertInnen handelt (Mieg & Brunner, 2001).

Für Mieg & Brunner (2001) gelten folgende Eigenschaften als Voraussetzung für ExpertInnenenkompetenz:

- Die geringe Bedeutung von persönlichen Generalfertigkeiten (Intelligenz, Gedächtnis etc.) für die Expertenleistung;
- Die strikte Bereichsabhängigkeit der Expertenleistung (Schachkompetenz ist nicht auf andere Bereiche übertragbar) (Mieg & Brunner, 2001).

Für Meuser & Nagel (1991) sind ExpertInnen Personen:

*[...] die selbst Teil eines Forschungsfeldes sind, das den Forschungsgegenstand ausmacht. Wir meinen nicht den Experten, der von außen – im Sinne eines Gutachters – Stellung zum Handlungsfeld nimmt (Meuser & Nagel, 1991).*

Darüber hinaus erwähnen Meuser & Nagel (1991):

*Ob jemand als ExpertIn angesprochen wird, ist in erster Linie abhängig vom jeweiligen Forschungsinteresse. ExpertIn ist ein relationaler Status (Meuser & Nagel, 1991).*

Zusammengefasst sind für Meuser & Nagel (1991) zwei Punkte relevant, um als ExpertIn bezeichnet zu werden:

*Wer in irgendeiner Weise Verantwortung trägt für den Entwurf, die Implementierung oder die Kontrolle einer Problemlösung (Meuser & Nagel, 1991).*

Oder:

*Wer über einen privilegierten Zugang zu Informationen über Personengruppen oder Entscheidungsprozesse verfügt (Meuser & Nagel, 1991).*

Der ExpertInnenstatus der Personen, welche im Kontext dieser Masterarbeit interviewt wurden setzt sich aus folgenden Punkten zusammen:

- Die Praxiserfahrung, die die befragten Personen im Rahmen des Berufslebens im Forschungsfeld vorweisen können.
- Die Position in einem Unternehmen, das im Gebiet der Digital Analytics tätig ist.

Zur Unterstützung der Auswahl von geeigneten ExpertInnen wurde zum einen Recherche in Fachmedien und im Internet durchgeführt sowie das im Beruf aufgebaute Netzwerk genutzt. Um das Untersuchungsergebnis so ausgewogen wie möglich zu halten, wurde darauf geachtet, ExpertInnen auszuwählen, die in unterschiedlichen Unternehmen und Positionen in diesem Feld tätig sind.

### **3.3 Vorbereitung eines leitfadengestützten ExpertInneninterviews**

Zur Vorbereitung auf ein ExpertInneninterview ist es in erster Linie wichtig, sich mit dem Fachgebiet auseinander zu setzen und sich ein entsprechendes Wissen anzueignen. Weiters ist es essentiell, die gängigen Fachausdrücke des Forschungsfeldes zu kennen und anwenden zu können. Neben dem konkreten Interviewleitfaden, muss als Voraussetzung auch Klarheit über Fragestellungen und die eigene Erkenntnisinteresse herrschen. Der Interviewleitfaden unterstützt den Interviewer/die Interviewerin dabei, dass im Gespräch alle Fragen vollständig und hinreichend behandelt werden (Mieg & Brunner, 2001).

Als eine Art Anleitung schlagen Mieg & Brunner (2001) folgende Strategie (Abbildung 24) bei der Bewältigung der Aufgabe vor:

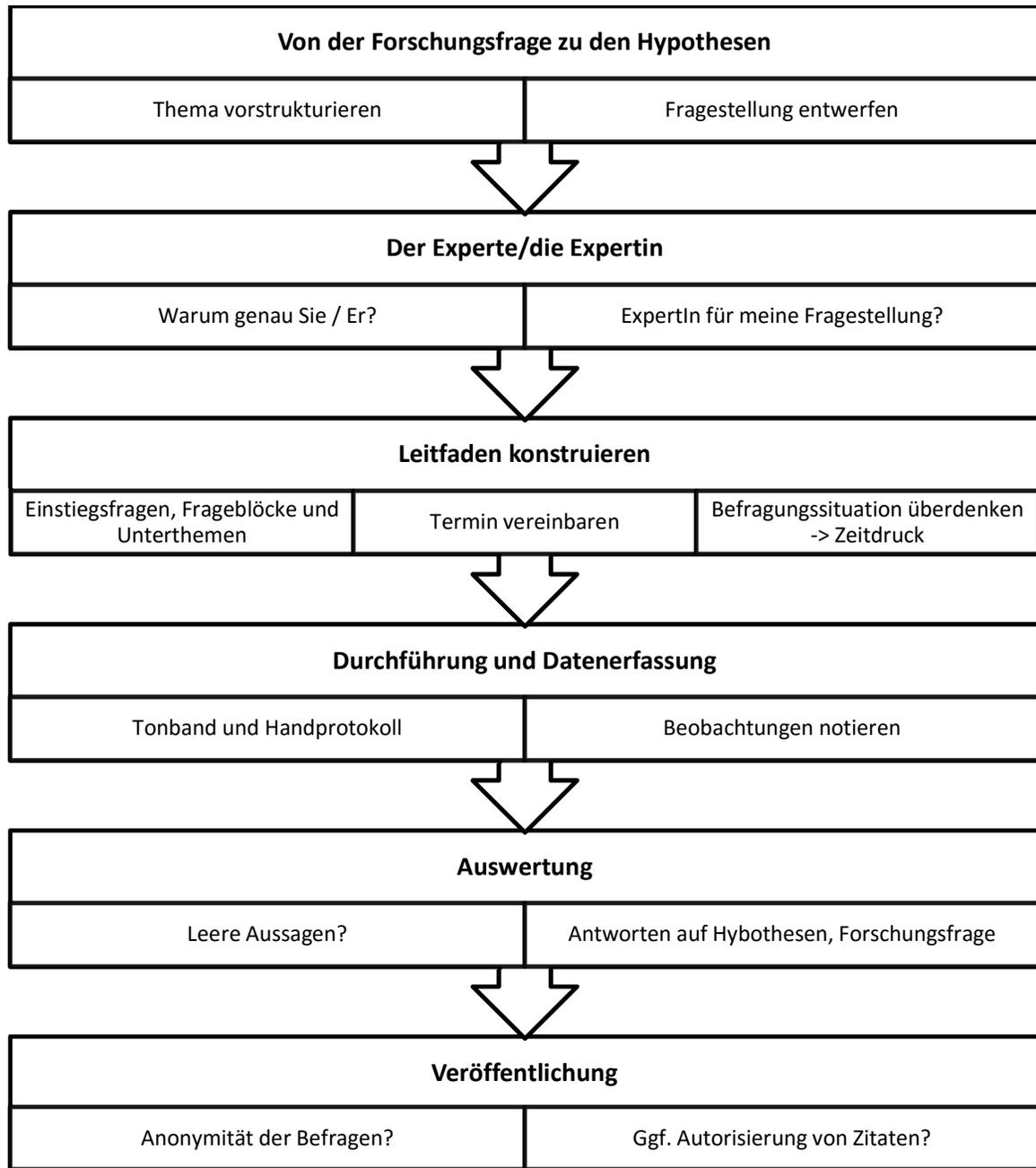


Abbildung 24: Kurzanleitung zur Durchführung eines ExpertInneninterviews (Mieg & Brunner, 2001).

### 3.4 Interviewleitfaden

|  | <i>Gesprächsimpuls</i>  | <i>Anmerkungen</i>   |
|--|---|--|
| F1   | Wie lange beschäftigen Sie sich schon mit dem Thema Digital Analytics?  |  |
| F2   | Gibt es Analytics Lösungen auf die Sie, respektive das Unternehmen in dem Sie tätig sind, spezialisiert sind?   |  |
| F3   | Wie beurteilen Sie das Verhältnis Ihrer Kunden im Hinblick auf deren Firmengröße?   | Verhältnis<br>KMU/Großunternehmen  |
| <b>Themenblock 1: Basisanforderungen</b>       |   |  |
| F4   | Welche beeinflussenden Kriterien sehen Sie für KMUs bei der Implementation von Digital Analytics Lösungen?  | Gibt es einen Punkt, ab dem die Implementation erst Sinn macht?<br><br>Kosten/Nutzen |
| F5   | Für Personen, die in diesem Bereich keinerlei Erfahrung aufweisen, ist die Komplexität auf den ersten Blick oft nicht ersichtlich.<br><br>Können Sie hier der Zielgruppe Empfehlungen abgeben, um häufige Fehler zu vermeiden?<br><br>Ist in diesem Fall eine Agentur zur Unterstützung zu empfehlen oder können Kurse hier Abhilfe schaffen? | Welche Ressourcen sollten im Unternehmen vorhanden sein?                             |
| F6   | Die Entscheidung für eine Digital Analytics Lösung hängt oft von den eigenen Zielen ab. Die Findung dieser Ziele ist aber für KMUs ohne Erfahrung in diesem Bereich schwierig.<br><br>Gibt es hier eine Methode, die Sie empfehlen können, um KMUs in diesem Prozess zu unterstützen.   |  |
| <b>Themenblock 2: Identifizierung von KPIs</b> |   |  |

|   |   |  |
|---|---|--|
| <b>Themenblock 3: Funktionales Design</b>                 |   |  |
| F7  | <p>Wie soll bei der Findung von KPIs vorgegangen werden?</p> <p>Welche KPIs sollten für ein KMU im E-Commerce im Fokus stehen?</p> <p>Gibt es hier eine Mindest- oder Maximalanzahl?</p>                  | <p>Null-Resultat bei Suchanfragen,<br/>Bestellrate für Kampagnen,<br/>Kosten pro Conversion etc.</p>               |
| F8  | <p>Gibt es in diesem Prozess Fehler, denen Sie als Experte/Expertin häufig begegnen?</p> <p>Wenn ja, wie kann man diesen entgegensteuern?</p>   |  |
| <b>Themenblock 3: Funktionales Design</b>                 |   |  |
| F9  | <p>Unterschiedliche Analytics Lösungen bieten naturgemäß auch unterschiedliche Funktionen an.</p> <p>Welche Basis-Funktionen würden Sie der Zielgruppe ans Herz legen?</p>                                | <p>evtl. Zusatzfrage:<br/>Worauf sollte man bei der Auswahl achten?<br/>Welche Fehler werden hier oft gemacht?</p> |
| F10   | <p>Kann in Bezug auf die vorherige Frage bereits eine Einschränkung von Analytics Lösungen für KMUs getroffen werden?</p>   |  |
| F11   | <p>Wie bereits angesprochen, sind personelle Ressourcen und somit auch das Know-How in dieser Zielgruppe oft begrenzt.</p> <p>Kann unter diesem Gesichtspunkt eine Analytics Lösung empfohlen werden?</p> | <p>Stichwort:<br/>Support<br/>Dokumentation<br/>Usability</p>  |
| <b>Themenblock 4: Datenschutzrechtliche Aspekte</b>       |   |  |
| F12   | <p>Gibt es datenschutzrechtliche Aspekte, die bei der Auswahl einer Analytics Lösung in jedem Fall zu beachten sind?</p>  | <p>Unterschiede europäische und amerikanische Anbieter</p>   |
| <b>Themenblock 5: Zukunftsaussichten und Schlussfrage</b> |   |  |

| F13 | In welche Richtung wird sich Digital Analytics weiterentwickeln?                                  | Stichwort:<br>Vielfalt an Anbietern oder<br>Google als Platzhirsch |
|-----|---|--|
| F14 | Aus welchen Gründen würden Sie als Experte/Expertin KMUs Digital Analytics empfehlen?             |  |
| F15 | Welche Themen abseits von den bereits besprochenen, erachten Sie für die Zielgruppe als relevant? |  |

### 3.5 Durchführung der Interviews

Im Zuge des empirischen Teils dieser Masterarbeit wurden sechs ExpertInneninterviews durchgeführt. Die ExpertInnen wurden durch Empfehlungen, berufliche Kontakte sowie über die Suche im Internet gefunden. Die Kontaktaufnahme erfolgte hierbei über Email, wobei das Interview in vier von sechs Fällen über Skype und die restlichen zwei per Telefon durchgeführt wurden. Im Anschluss daran folgte eine Vorstellung mit Name, Studienrichtung und Studienherkunft und eine Einführung in das Thema der Masterarbeit. Sofern die ExpertInnen zu diesem Zeitpunkt keine Fragen hatten („Haben Sie noch Fragen, bevor wir starten?“), wurden die TeilnehmerInnen darüber aufgeklärt, dass das Interview aufgenommen wird. Waren die TeilnehmerInnen damit einverstanden, wurde das Interview gestartet. Wie im Interviewleitfaden in Kapitel 3.4 ersichtlich, wurde mit drei Eisbrecher-Fragen gestartet. Die darauffolgenden Themenblöcke orientierten sich an der Reihenfolge im Implementierungsprozess und schlossen mit einem Ausblick über die Entwicklung des Themas ab. Um eine korrekte Aufnahme zu gewährleisten, wurden die Interviews am Laptop mit der Software *Amolto Call Recorder* und am Smartphone mit der App *Diktiergerät* aufgenommen. Die simultane Aufnahme auf beiden Geräten sollten den Datensatz für die spätere Auswertung sichern. Da von Seiten des Studiengangs IT & Wirtschaftsinformatik an der FH Campus02 keine Transkription der Interviews erforderlich war, sind die Interviews auf einer beigelegten CD-ROM gespeichert.

### 3.6 Auswertung der Interviews

Die Interviews wurden nach der qualitativen Inhaltsanalyse von Meuser/Nagel (2009) ausgewertet. Bei dieser Methode ist der zentrale Punkt der Inhalt des Interviews. Das Ziel dabei ist es, aus den Interviews das Gemeinsame und Typische herauszuarbeiten und gleichzeitig das Datenmaterial vom Umfang her zu reduzieren (Meuser & Nagel, 2009).

Bei der Analyse der Interviews müssen nach Meuser & Nagel (2009) die folgenden Arbeitsschritte aufsteigend durchlaufen werden:

1. *Paraphrasierung: Gliederung in einzelne Textabschnitte und textgetreue Wiedergabe vom Inhalt geordnet nach thematischen Einheiten mit eigenen Worten.*
2. *Thematisches Ordnen: Stichworte und Überschriften finden; thematisches Sortieren von einzelnen Textsegmenten in Terminologie des Befragten; nah am Text bleiben.*
3. *Thematischer Vergleich: Vergleich der Kategorien mit dem eigenen Wissen/anderen empirischen Studien/Erfahrungen aus dem Feld; in wissenschaftliche Sprache umformulieren; interpretieren, werten; Verallgemeinerbarkeit bleibt noch auf das Material beschränkt.*
4. *Konzeptualisierung: Vergleich der Kategorien mit dem eigenen Wissen/anderen empirischen Studien/Erfahrungen aus dem Feld; in wissenschaftliche Sprache umformulieren; interpretieren, werten; Verallgemeinerbarkeit bleibt noch auf das Material beschränkt.*
5. *Theoretische Generalisierung: Einbeziehung entsprechender Theorien; einzelne Themen werden in ihren theoretischen Zusammenhang gebracht; interpretieren, werten; eigene Termini benutzen (Meuser & Nagel, 2009).*

### 3.7 ExpertInnen

In nachfolgender Tabelle 8 sind die an dem Interview teilnehmenden ExpertInnen mit Informationen zu ihren Tätigkeitsfeldern, Erfahrungen und Positionen im Unternehmen angeführt sowie die Analytics-Lösungen, die die befragten Personen einsetzen.

| Name            | Tätigkeitsfelder  | Branchen-<br>erfahrung | Position im<br>Unternehmen     | Analytics-<br>Lösungen im<br>Einsatz |
|-----------------|---|------------------------|--------------------------------|--------------------------------------|
| Stefan Krausler | Softwareentwicklung,<br>Suchmaschinenoptimierung  | 6 Jahre                | CTO                            | Google Analytics und Piwik           |
| Thomas Tauchner | Software- und<br>Softwarearchitektur, Digital<br>Analytics  | 6 Jahre                | CEO                            | Webtrekk und Google Analytics        |
| Harald Koch     | Online Marketing, Usability-,<br>Workflow- und Prozess-<br>Optimierung,<br>Suchmaschinenoptimierung | 20 Jahre               | Geschäftsführer                | Google Analytics und Piwik           |
| Ines Kriebernig | Webentwicklung, Digital<br>Analytics,<br>Suchmaschinenoptimierung                                   | 3 Jahre                | Online Marketing<br>Consultant | Google Analytics und Piwik           |
| Miha Hartmann   | Webentwicklung,<br>Suchmaschinenoptimierung,<br>Suchmaschinenmarketing                              | 7 Jahre                | Führungskraft                  | Google Analytics                     |

|                        |  |         |                  |                                  |
|------------------------|--|---------|------------------|----------------------------------|
| Nils<br>Brettschneider | Web Analyse,<br>Suchmaschinenmarketing,<br>Web Entwicklung | 7 Jahre | Geschäftsführung | Google<br>Analytics und<br>Piwik |
|------------------------|--|---------|------------------|----------------------------------|

*Tabelle 8: Informationen zu ExpertInnen, die am Interview teilnahmen*

## 4 HANDLUNGSLEITFADEN

Um dem Handlungsleitfaden eine sinnvolle Struktur zu geben, wurde diese aus dem Interviewleitfaden übernommen, um dadurch eine bessere Übereinstimmung mit den Antworten der ExpertInnen zu erlangen und um den chronologischen Ablauf hinsichtlich der Implementation beizubehalten. Zur Erstellung des Handlungsleitfadens wurde von jedem Interview immer der gerade bearbeitete Themenblock angehört und davon die wichtigsten Erkenntnisse nach Meuser/Nagel (2009) notiert. Dieser Vorgang wurde gewählt, um die Informationen der ExpertInnen nicht zu vermischen, doppelten Inhalt zu vermeiden sowie konträre Ansichten besser herauszustreichen. So wurde auch sichergestellt, dass Meinungen, die von mehreren ExpertInnen vertreten wurden, sich auch im Interviewleitfaden dementsprechend niederschlagen.

Im folgenden Handlungsleitfaden werden die Ergebnisse der ExpertInneninterviews aufbereitet sowie zwei ausgewählte Analytics Lösungen vorgestellt. Der Leitfaden gliedert sich dabei in folgende Blöcke:

- Themenblock 1: Basisanforderungen
- Themenblock 2: Messgrößen definieren
- Themenblock 3: Funktionales Design
- Themenblock 4: Datenschutzrechtliche Aspekte
- Themenblock 5: Zukunftsaussichten

In Themenblock 3 befindet sich neben den allgemeinen Informationen zu Analytics Lösungen auch eine Einführung in die Analytics Lösungen Google Analytics und Piwik. Der Grund, warum diese beiden ausgewählt wurden, liegt zum einen daran, dass für beide Lösungen kostenlose Varianten verfügbar sind und dies für KMUs einen nicht zu unterschätzenden positiven Faktor darstellt.

Bei Google Analytics handelt es sich zum anderen um jene Lösung, die mit überwältigender Mehrheit am Markt eingesetzt wird (siehe Kapitel 4.3) und dazu auch ausnahmslos alle ExpertInnen Erfahrungen aufweisen können. Bei Piwik handelt es sich um eine Open Source Software, die nach Google Analytics die am zweit häufigsten verwendete Analytics Software in Deutschland und Österreich ist (W3 Techs, 2017).

Zudem unterscheidet sich Piwik von Google Analytics durch den Umstand, dass die Installation der Software und die Speicherung der gesammelten Daten nicht auf Fremdservern, sondern auf selbst zur Verfügung gestellten Servern stattfinden kann.

Am Ende jedes Hauptthemenblocks sind konkrete Handlungsempfehlungen zu finden, um den LeserInnen aus jedem Themenblock eine kurze und prägnante Zusammenfassung zu geben.

## 4.1 Themenblock 1: Basisanforderungen

In diesem Themenblock werden folgende Fragen beantwortet:

- Gibt es beeinflussende Kriterien und/oder Voraussetzung, die KMUs vor der Implementation zu beachten haben?
- Gibt es einen Prozess/eine Methode, um diese Voraussetzungen zu erfüllen?
- Ist es zwingend nötig, eine Agentur zur Unterstützung zu Rate zu ziehen?
- Welche Fehler treten im Zuge der Implementation auf?

Ob es Vorbedingungen bezüglich der Implementation gibt, hängt laut ExpertInnen von der jeweiligen Analytics Lösung ab. Beschränkt man sich auf Google Analytics, gibt es laut ExpertInnen keine Vorbedingungen, da diese zum einen keine Hardware-technischen Voraussetzung erfordert und zum anderen eine Implementationsanleitung anbietet. Verwendet man ein Content Management System wie zum Beispiel WordPress, so bieten diese bereits vordefinierte Lösungen für den Einsatz von Analytics Lösungen an. Ist der Shop auf Servern eines Hosting-Partners konfiguriert, kann in diesem Fall die Implementation des Codes auch von diesem erfolgen. Auf der anderen Seite gibt es aber auch Tools, die eine beträchtlich aufwendigere Implementation erfordern. In vielen Fällen handelt es sich dabei um nicht kostenlose Systeme im Gegensatz zu Google Analytics oder Piwik. Bei diesen Tools bedarf es daher auch einer vorherigen Schätzung von Nutzen und Kosten. Sollte aus Funktionsgründen eine Analytics Lösung dieser Art bevorzugt werden, so sollte zur Beratung ein externer Dienstleister hinzugezogen werden.

### 4.1.1 Zielfindungsprozess im Unternehmen

Bevor das Unternehmen sich aber überhaupt mit der Auswahl und der Implementation einer Analytics Lösungen befasst, wird aus Sicht von ExpertInnen empfohlen, zuerst die (digitalen) Ziele des Unternehmens zu überprüfen. Unternehmen sollen sich dabei die Frage stellen, ob sich das digitale Geschäft und die damit verbundenen Kennzahlen in den Zielen widerspiegeln. Unternehmen müssen sich im Klaren sein, wohin sich der eigene digitale Bereich entwickeln soll und diese Entwicklung in ihren Zielen festhalten. Für den Fall, dass diese Ziele im Unternehmen noch nicht definiert worden sind, können diese in einem Workshop erarbeitet werden. Dieser Prozess kann, muss aber nicht zwingend mit einer Digital Analytics Agentur durchgeführt werden, sondern auch mit einem Unternehmensberater. Die ExpertInnen empfehlen bei diesem Schritt in jedem Fall externe BeraterInnen zu konsultieren, um der eigenen „Betriebsblindheit“ entgegenzuwirken. Die Punkte, die im ersten Part eines Zielfindungs-Workshop erarbeitet werden, sind folgende:

- Welche Ziele verfolge ich im digitalen Bereich bzw. wo will ich damit an einem definierten Zeitpunkt stehen?
- Was ist/wäre ein Erfolg für mich in diesem Bereich?
- Welche Geschäftsprozesse werden bereits jetzt digital abgearbeitet und welche nicht?

- Wie gliedern sich diese Prozesse in die Unternehmensstruktur ein?
- Was funktioniert bei diesen Prozessen gut und was nicht?
- Gibt es auch ein „offline“ Geschäft und welche Schlüsse können daraus für das „online“ Geschäft gezogen werden?

Im zweiten Teil des Workshops wird versucht die Frage zu beantworten, welche Daten bei der Umsetzung der festgelegten Ziele helfen können. Erst ab diesem Punkt kann auch bestimmt werden, für welche Zwecke eine Digital Analytics Lösung eingesetzt werden kann. ExpertInnen machen hier häufig die Erfahrung, dass im Prozess der Integration von Analytics Lösungen dieser essentielle Schritt oft übersprungen wird und somit an einem späteren Zeitpunkt, an dem richtungsweisende Entscheidungen bereits getroffen wurden, diese nochmals zu korrigieren oder im Worst Case zu revidieren sind. Ein Modell, welches sich für die Erarbeitung dieser Ziele eignet und von ExpertInnen empfohlen wird, nennt sich OKR-Modell, welches im Unterkapitel 4.1.1. vorgestellt wird.

Ein weiterer Fehler, den Unternehmen laut ExpertInnen oft begehen, ist es, sich entweder bei der Integration einer Lösung komplett auf die dafür beauftragte Agentur zu verlassen oder umgekehrt, von der Agentur in den Prozess nicht aktiv eingebunden zu werden. Die ExpertInnen raten hier – sofern der Prozess nicht ohne externe Unterstützung durchgeführt wird – diesen Prozess in enger Zusammenarbeit gemeinsam zu absolvieren, aber dass die Letztentscheidung immer vom Auftraggeber/von der Auftraggeberin gefällt werden muss. Dass für die erstmalige Implementation eine Agentur zu Hilfe gezogen wird, ist wie am Anfang des Kapitels bereits erwähnt, keine Voraussetzung für eine erfolgreiche Integration. Diese kann auch ohne Fachwissen im Bereich der Digital Analytics selbst vollzogen werden. Oft entsteht ein Bedarf an externer Beratung erst beim Lesen und interpretieren von Auswertungen. In diesem Bereich raten die ExpertInnen aufgrund der einfachen Handhabe zu Google Analytics. An zweiter Stelle könnte laut ExpertInnen aufgrund von datenschutzrechtlichen Bedenken Piwik gewählt werden. Fehler, die bei der Implementation selbst passieren, hängen oft damit zusammen, dass Tracking Code auf einzelnen Webseiten fehlerhaft eingebunden wird. Es kann sein, dass der Code doppelt eingebunden wird oder sich dieser nicht im Header befindet. Ein weiteres Problem kann eine fehlerhafte JavaScript Syntax sein. Als Beispiel seien hier Fehler bei den öffnenden und schließenden Tags genannt. Dies sorgt dann dafür, dass Seitenaufrufe entweder gar nicht oder doppelt gezählt werden und somit nützliche Informationen aus dem Analytics System hinfällig werden.

Mit Nachdruck empfehlen die ExpertInnen, die Wirtschaftlichkeit bei diesem Thema nicht aus den Augen zu lassen und alle Investitionen und Kampagnen einer ständigen Kosten-Nutzen-Rechnung zu unterziehen. Die ExpertInnen bemängeln, dass viele Unternehmen für teure Analytics Lösungen und Kampagnen jeglicher Art keine Strategie verfolgen und die Auswirkungen von Investitionen nur schlecht oder falsch interpretieren.

### 4.1.2 Objectives and Key Results Modell

Das OKR-Modell verkörpert eine Management Methode, die einem Unternehmen dabei helfen soll, messbare Ziele zu definieren und diese auf alle Unternehmensebenen abzustimmen sowie deren Progression zu beobachten. Erfunden wurde das OKR-Modell von Andy Grove, der seines Zeichens einer der Mitgründer von Intel ist. An prominenter Stelle eingesetzt wurde es von Investor John Doerr im Jahr 1999 bei Google. Eingesetzt wird das Management Tool heute von IT-Konzernen wie Oracle und Twitter sowie auch bei Unternehmen im Lebensmittelhandel wie MyMuesli (T3N Digital Pioneers, 2014).

Die Grundidee des OKR-Modells ist es, dass jedem Ziel (Objectives) Schlüsselergebnisse (Key Results) zugeordnet werden. Diese Schlüsselergebnisse werden mit den Zielen in regelmäßigen Abständen verglichen und gegebenenfalls neue OKRs definiert. Beim Festlegen der Ziele muss beachtet werden, dass diese immer mit messbaren Schlüsselergebnissen unterlegt sein müssen. Die OKRs müssen auf allen Ebenen des Unternehmens definiert werden. Diese beginnen beim Management Team und werden dann zu den Teams bzw. bis zum Mitarbeiter/zur Mitarbeiterin nach unten „synchronisiert“. Auf allen Ebenen darunter müssen in Folge team- bzw. personenspezifische OKRs festgelegt werden, die zum „Großen Ziel“ einen Beitrag leisten. Wichtig ist hierbei, dass diese Ziele nicht aufoktroiert werden, sondern ausverhandelt. Das Team und der einzelne Mitarbeiter/die einzelne Mitarbeiterin sollen sich selbst überlegen, wie sie das Unternehmen unterstützen können, um dieses Ziel zu erreichen. Es wird empfohlen, zu Beginn eines Quartals 4 bis 6 OKRs zu erstellen und diese zu evaluieren. Sollten die Ziele nicht erreicht werden, so sind diese nicht zu sanktionieren, sondern in die Bewertung der nächsten OKRs miteinzubeziehen (T3N Digital Pioneers, 2014).

Nachfolgend ein Praxisbeispiel anhand eines Bloggers:

- Ziel: Die Reputation des Software Bloggers verbessern
  - Schlüsselergebnis:
    1. Der Blogger/die Bloggerin nimmt in den nächsten 6 Monaten an drei BloggerInnen Events teil.
    2. Der Blogger/die Bloggerin veröffentlicht 5 Blogeinträge mehr pro Monat.
    3. Der Blogger/die Bloggerin eröffnet ein Twitterkonto und generiert 100 FollowerInnen.
    4. Der Blogger/die Bloggerin erweitert sein/ihr Wissen um eine neue Technologie.

Am Ende des Quartals werden die einzelnen Schlüsselergebnisse mit einer Zahl zwischen 0 bis 1 bewertet. Das Ziel hierbei ist, die einzelnen Ergebnisse mit 0,6 oder 0,7 zu bewerten. Bewertungen darüber bedeuten, dass das Schlüsselergebnis zu einfach angesetzt wurde. Ein Wert unter dieser Schwelle heißt, dass das Ziel nicht erreicht wurde (Yarow, 2014). Zusammengefasst nochmals alle relevanten Punkte zum Einsatz von OKR:

- 4 bis 6 OKRs pro Quartal definieren.

- Ein OKR besteht aus einem Ziel und seinen Schlüsselergebnissen.
- Schlüsselergebnisse messbar machen.
- In Unternehmen werden OKRs von oben nach unten „synchronisiert“, nicht aufoktroziert.
- Am Ende des Quartals werden die Schlüsselergebnisse von 0 bis 1 bewertet.
- Ziel ist es, die Schlüsselergebnisse mit 0,6 oder 0,7 zu bewerten (Yarow, 2014).

Der Erfolg dieses Modells hängt letztendlich vom Willen und der Kommunikation im Unternehmen ab.

**Aus Themenblock 1 ergeben sich somit folgende konkrete Handlungsempfehlungen:**

- Identifizierung der (digitalen) Ziele des Unternehmens sofern diese nicht vorhanden sind. Hier kann das OKR-Modell, welches in Kapitel 4.1.1 vorgestellt wird, zur Unterstützung eingesetzt werden.
- Notwendigkeit, um eine externe Beratungsagentur hinzuziehen, überprüfen.
- Analyse der Online-Daten, die zur Erreichung dieser Ziele zur Verfügung stehen.
- Abschätzung des Budgets, welches für Implementation und Betrieb einer Analytics-Lösung zur Verfügung steht – Stichwort Kosten-Nutzen-Rechnung.

## 4.2 Themenblock 2: Identifizierung von KPIs

Im zweiten Themenblock werden folgende Fragen beantwortet:

- Gibt es einen Prozess, der die Definition von KPIs (i.S.v. Messgrößen) unterstützt?
- Gibt es KPIs, die für die Zielgruppe in jedem Fall relevant sind?
- Gibt es immer wiederkehrende Fehler, die den ExpertInnen in diesem Prozess oft begegnen und wie kann man diese vermeiden?

Bei der Identifizierung von KPIs empfehlen ExpertInnen heutzutage einen agilen Ansatz. Agil bedeutet in diesem Fall, dass KPIs nur einzeln, aber nie in Paketen zusammengefasst implementiert werden. Im Detail werden in den Workshops, welche in Themenblock 1 beschrieben wurden, auch mögliche relevante KPIs definiert. Folgende KPIs sollten für Webshop BetreiberInnen aber in jedem Fall interessant sein:

- Wie viele BesucherInnen finden überhaupt den Weg in den Online Shop?
- Wie viele BesucherInnen legen Produkte in den Warenkorb?
- Wie viele BesucherInnen schließen einen Bestellvorgang ab?

Die Implementation dieser KPIs erfolgt im Anschluss an eine mehrwöchige „Warm-Up“ Phase, in welcher die Performance der unterschiedlichen Kanäle im Unternehmen beobachtet und verglichen wird. Um die Qualität und den Output der einzelnen Kennzahl zu kennen, wird diese nach der Implementation einer Evaluierungsphase unterzogen. Können die erwarteten Informationen aus dem implementierten KPI extrahiert werden, beginnt dieser Prozess von

neuem mit dem bereits implementierten KPI, sowie einem zusätzlichen. Hierbei wird wie in der agilen Software Entwicklung in sogenannten Sprints vorgegangen. Für gewöhnlich dauern diese Sprints zwei Wochen. Die dabei involvierten Personen werden als Business Unit bezeichnet. Diese Business Unit besteht meist aus drei Personen, die die Rolle eines Technikers/einer Technikerin, eines Business Analysten/einer Business Analystin und eines Entscheiders/einer Entscheiderin einnehmen. In diesen zweiwöchigen Sprints wird solange an einer Messgröße gearbeitet, bis diese einen zufriedenstellenden Output liefert. Dieser Vorgang unterscheidet sich wesentlich vom klassischen Wasserfallmodell, in dem ein Katalog mit verschiedensten KPIs erstellt wird und dieser dann bei der Implementation umgesetzt wird. Der Vorteil dieser Methode – der jedoch zugleich auch Nachteil sein kann – zeigt sich an der Menge an Daten, die bereits nach der ersten Implementation generiert werden. Der Nachteil besteht darin, dass aus dieser großen Datenmenge häufig keine nützlichen Schlüsse gezogen werden können.

Laut ExpertInnen wurden in vielen Fällen sogenannte Tracking Maps, welche beim Wasserfallmodell sämtliche Implementierungsdetails enthalten und oft mehrere 100 Seiten stark sind, nicht ansatzweise komplett umgesetzt, da sich die Zielvorgaben im Laufe der Zeit stark veränderten. Vor allem bei KMUs erfordert das oftmals geringe Budget eine klare Zieldefinition, sodass eine wahllose Streuung von KPIs nicht wirtschaftlich wäre. Im Gegensatz dazu können beim Agilen Ansatz bereits nach kurzer Zeit erste Ergebnisse vorgelegt werden.

Bei der Definition der KPIs sollte auch auf eine Kaskadierung der Konversionsrate (siehe Kapitel 2.2.1) geachtet werden. Das heißt, das Ziel sollte am Anfang nicht sein, so viele BesucherInnen wie möglich in den Shop zu bekommen, sondern zu analysieren, wie das Verhältnis von Personen, die in den Shop gelangen und ein Produkt in den Warenkorb legen, zu jenen ist, die dann mit diesem Warenkorb auch eine Bestellung abschließen. Aus dieser Analyse können dann jene Stellen offengelegt werden, an denen eine Prozessoptimierung sinnvoll wäre. Vor allem wird darauf hingewiesen, sich am Beginn der Integration nicht in komplexen Details zu verlieren. Digital Analytics ist dazu da, um Zusammenhänge zu plausibilisieren und aufzuzeigen. Speziell KMUs sollten hier auch das Kosten-Umsatz-Verhältnis im Auge behalten, denn oft ist beim eingesetzten Kapital weniger mehr.

Bei Verbesserungsentscheidungen raten ExpertInnen dazu, nicht das Gesamtbild aus den Augen zu verlieren und Entscheidungen immer im Einklang mit anderen Kriterien, die dabei ebenso beeinflusst werden, zu fällen. Bei diesen Entscheidungen wird empfohlen, neben Analytics zusätzlich einen A/B Splittest (Kapitel 2.2.2) zur Entscheidungsunterstützung durchzuführen. Der A/B Splittest wird allerdings aufgrund seiner Einfachheit von vielen Agenturen stiefmütterlich behandelt, obwohl dieses Instrument die Entscheidungsfindung in Grenzfällen maßgeblich unterstützen kann.

**Aus Themenblock 2 ergeben sich somit folgende konkrete Handlungsempfehlungen:**

- Identifizierung von KPIs, die für die Erreichung der zuvor definierten Ziele relevant sind. Hierbei kann wieder das OKR-Modell aus dem ersten Themenblock angewendet werden.
- Bei der Identifizierung von KPIs darauf achten, dass diese eine Kaskadierung aufweisen, um durch deren Kombination detaillierte Rückschlüsse treffen zu können.

- Folgende drei KPIs sollten in jedem Fall berücksichtigt werden:
  - Wie viele BesucherInnen finden den Weg in den Online Shop?
  - Wie viele BesucherInnen legen Produkte in den Warenkorb?
  - Wie viele BesucherInnen schließen einen Bestellvorgang ab?
- Bei Handlungen, die auf den Ergebnissen der KPIs basieren, immer das Gesamtbild im Auge behalten.
- Zusätzlich zu den klassischen KPIs wird empfohlen, in regelmäßigen Abständen einen A/B Splittest durchzuführen.

### 4.3 Themenblock 3: Funktionales Design

In diesem Themenblock werden folgende Fragen beantwortet:

- Gibt es Digital Analytics Lösungen, auf die sich die Zielgruppe einschränken sollte?
- Welche Basisfunktionen sollten fixer Bestandteil einer Digital Analytics Lösung sein?
- Worin unterscheiden sich die in Deutschland und Österreich weit verbreitetsten Digital Analytics Lösungen?

Zusätzlich zu diesen drei Fragen, wird ebenso eine Einführung und Installationsanleitung für die Analytics Lösungen Google Analytics und Piwik vorgestellt. Dieser Teil des Handlungsleitfadens wurde auf Basis der Auswertungen von Themenblock 3 erstellt und war nicht Teil des Interviews. Die Entscheidung, diese beiden Analytics Lösungen näher vorzustellen, resultierte zum einen aus den Erfahrungen und Antworten der ExpertInnen und zum anderen aus der aktuellen Marktsituation. Den LeserInnen soll hier neben den Informationen rund um die das Thema Digital Analytics auch eine Einführung in die Implementation selbst gegeben werden.

Heutzutage ist eine klare Unterscheidung zwischen Digital Analytics Lösungen nicht mehr möglich, da sie sich in der Grundausstattung der angebotenen Funktionen kaum unterscheiden. Laut ExpertInnen unterschieden sich in den Anfangszeiten von Digital Analytics diese Tools im Hinblick auf die angebotenen Funktionen stark. Die Analytics Lösung Webtrekk galt als Tool, das speziell für den Einsatz im E-Commerce gedacht war. Google Analytics war bekannt für seine breitgefächerten Einsatzmöglichkeiten. Dies ist laut ExpertInnen heute nicht mehr der Fall, da alle Tools ihre Funktionen stark angeglichen haben bzw. es möglich ist, entweder durch Anpassung oder Upgrades die gewünschten Anforderungen nachzurüsten. Aus diesem Grund liegen die großen Unterschiede nicht zwischen den unterschiedlichen HerstellerInnen, sondern zwischen den Gratis- und Bezahlversionen. Im Zuge der ExpertInneninterviews trat zu Tage, dass diese den KMUs mit großer Mehrheit Google Analytics und Piwik empfehlen. Nachfolgend werden beide Tools näher vorgestellt, um der Zielgruppe einen Überblick zu geben und herauszustreichen, weshalb sie für die ExpertInnen die erste Wahl für die Zielgruppe darstellen.

**Aus Themenblock 3 ergeben sich somit folgende konkrete Handlungsempfehlungen:**

- Identifikation der Funktionalitäten, die zur Erreichung der gesteckten Ziele benötigt werden.
- Abgleich der identifizierten Funktionalitäten mit jenen der am Markt erhältlichen Analytics Lösung.
- Kosten-Nutzen-Vergleich der Analytics Lösung und Abstimmung mit dem verfügbaren Budget.
- Implementation der gewählten Analytics Lösung.

### **4.3.1 Einführung Google Analytics**

Google Analytics entstand im März 2005 durch die Übernahme der Firma Urchin. Urchins erste Web Analyse Software wurde zur Logfile Analyse eingesetzt, welche im Jahr 1997 veröffentlicht wurde. In den Jahren nach der Übernahme 2005 erweiterte Google das Featureset von Google Analytics stetig mit weiteren Funktionen, sodass es im Jahr 2013 bereits über 25 Millionen eröffnete Google Analytics Accounts gab. Ein kuriose Detail hier ist, dass Urchin bereits im Jahr 1995 gegründet wurde und somit einige Jahre älter ist als Google selbst. In Österreich setzen 85 Prozent aller Websites, die Analytics nutzen, Google Analytics ein. Durch die ständige Weiterentwicklung ist die BenutzerInnen-Oberfläche bereits in 40 Sprachen verfügbar. Für potentielle NutzerInnen ist es möglich, sich zwischen zwei Versionen – einer Free- und einer Bezahlversion (Google Analytics 360) – zu entscheiden (Aden, 2013).

Da die Auswahl auf Letztere für einen Großteil der Zielgruppe nicht zielführend wäre, behandelt dieser Leitfaden nur Google Analytics in der Gratis-Version. Der Grund, weshalb Google hier seinen KundInnen entgegenkommt und überhaupt eine kostenlose Version anbietet, erscheint erst bei näherer Betrachtung logisch. Bevor Google mit der eigenen Analytics Lösung auf dem Markt ging, gab es bereits Google AdWords, das neben der Suchmaschine eines der wichtigsten Google-Produkte ist. Das Prinzip von AdWords beruht auf kleinen (Text-)Anzeigen, welche den NutzerInnen neben dem eigentlichen Suchergebnis angezeigt werden. Da diese Anzeigen durch den Suchbegriff sehr kontextbezogen sind, befinden NutzerInnen diese nicht als störend. Die WerberInnen bei Google Adwords profitierten zwar von der platzierten Werbung bei den Suchergebnissen, dennoch konnte der Umsatz nicht erhöht werden bzw. verschlechterte sich die Konversion sogar. An dieser Stelle befand sich Google nun in einer Zwickmühle: Zum einen profitierte man von der Platzierung der Werbung maßgeblich, zum anderen wurde aber befürchtet, dass NutzerInnen aufgrund von schlechten Erfahrungen beim Besuch der Anzeigen, die Suchmaschine selbst nicht mehr verwenden würden. Um diesem Problem entgegenzusteuern, stellte Google den Werbenden nun Google Analytics zur Verfügung, um Transparenz in die Metriken der jeweiligen Seiten zu bringen und so die UserInnen-Experience für BesucherInnen zu erhöhen. Aus diesem Blickwinkel betrachtet war es für Google eine Win-Win Situation (Aden, 2013).

In der kostenlosen Version von Google Analytics werden ab 10.000.000 Hits (siehe Kapitel 2.2.4) pro Monat keine Daten mehr gespeichert bzw. wird man von Google kontaktiert, um auf die Premium Version umzusteigen. Die gesammelten Daten lassen erst nach 24 Stunden ab

Speicherung auswerten. Aus diesem Grund ist eine Liveauswertung nicht möglich. Gegen zu komplexe Datenauswertungen in der Gratisversion wappnet sich Google mit Sampling von Daten. Das bedeutet, dass bei rechenintensiven Vorgängen Google nicht mehr mit 100 Prozent der gespeicherten Daten arbeitet, sondern nur mit z.B. 30 Prozent und die restlichen 70 Prozent sind gesampelte Daten. Gesampelte Daten sind Datensätze, die aus den originalen Datensätzen hochgerechnet werden. Eine ungefähre Abschätzung, welche Datenmengen gesammelt werden und wie zeitnah nach Sammlung der Daten Reports benötigt werden, ist somit empfehlenswert (Aden, 2013).

Trotz dieser Einschränkungen ist laut ExpertInnen die Gratis-Version von Google Analytics für einen Großteil der KMUs ausreichend, da in den meisten Fällen nicht mit derart großen Datenmengen gearbeitet wird und auch Live Analysen zu Beginn nicht notwendig sind. Neben diesen Punkten zeichnet sich Google Analytics durch ein großes Netz an Google PartnerInnen aus, die Support-Dienstleistungen anbieten. Auch stellt Google Dokumentation bereit, die vor allem für den Start eine nützliche und vor allem kostenlose Informationsgrundlage bieten. Die Basisfunktionen sind hierbei auf EinsteigerInnen ausgelegt und benötigen nicht unbedingt professionelle Hilfe.

#### 4.3.1.1. Installation von Google Analytics

1. Für die erstmalige Installation von Google Analytics ist eine gültige Google Email-Adresse Voraussetzung. Sofern kein Gmail-Account vorhanden ist, muss dieser bei [www.google.at/accounts](http://www.google.at/accounts) erstellt werden.
2. Nachdem der Account erstellt wurde, gelangt man über <https://www.google.at/analytics> zur entsprechenden Schaltfläche, über die das Analytics Konto erstellt werden kann. Siehe dazu Abbildung 25:



Abbildung 25: Erster Schritt zur Kontoerstellung

- Wie in Abbildung 26 zu sehen ist, kann im darauffolgenden Fenster die Anmeldung durch einen Klick auf den Button „anmelden“ bestätigt werden:

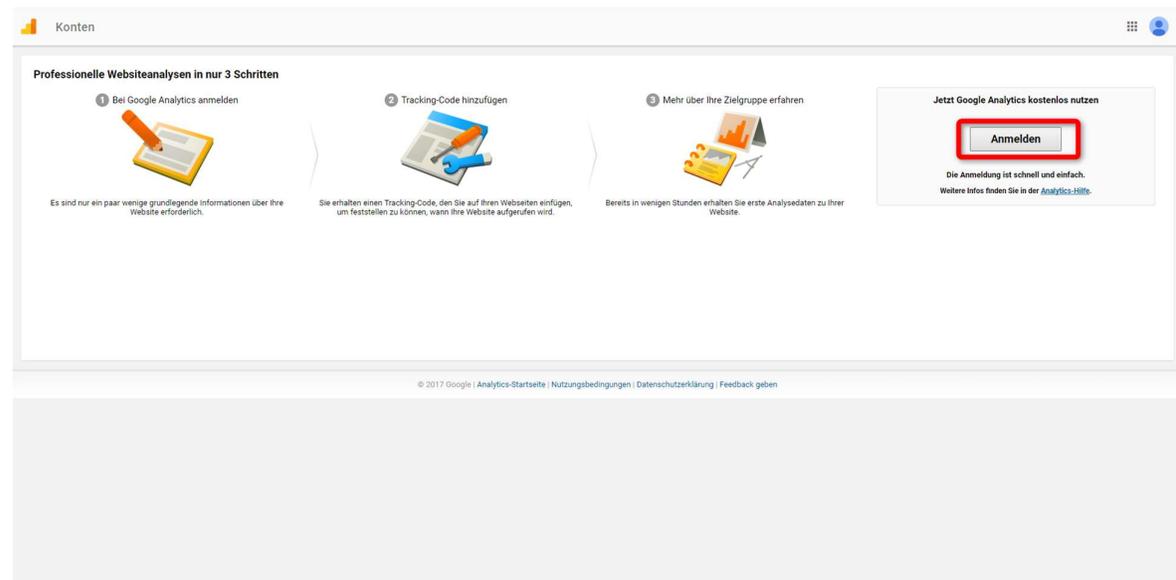


Abbildung 26: Schritt zwei im Erstellungsprozess

- Im nächsten Schritt muss festgelegt werden, ob eine Website oder eine Mobile App getracked werden soll. Zusätzlich muss ein Kontoname deklariert werden, Namen und URL der Property angegeben werden sowie die Branche und die entsprechende Zeitzone. Zu sehen sind diese Angaben in Abbildung 27:

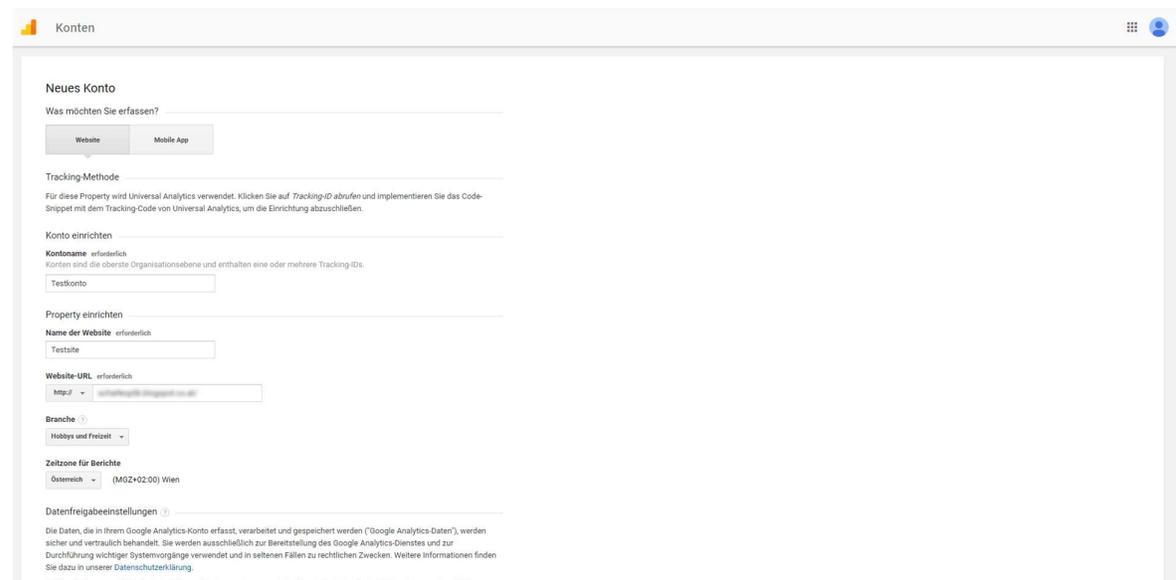


Abbildung 27: Angabe der benötigten Dateien

- Sind alle Einstellungen richtig gesetzt worden, kann am Ende der Seite der Tracking Code abgerufen werden. Es ist empfehlenswert, zu überprüfen, ob die aktuellen Datenschutzbestimmungen von Google Analytics im Einklang sind mit den gesetzlichen Bestimmungen des Landes, das die Website betreibt. Näheres dazu findet sich im Anschluss an dieses Kapitel.
- Wie in Abbildung 28 zu sehen, befindet man sich nun bereits in der Google Analytics Konsole, wo im nächsten Schritt der generierte Tracking Code von Google in die Webpräsenz eingefügt werden muss. Befindet sich die Webpräsenz auf *WordPress*, *GoDaddy* oder *Wix*, so muss in diesem Fall den Anleitungen der Hosters zur Einbindung Folge geleistet werden.

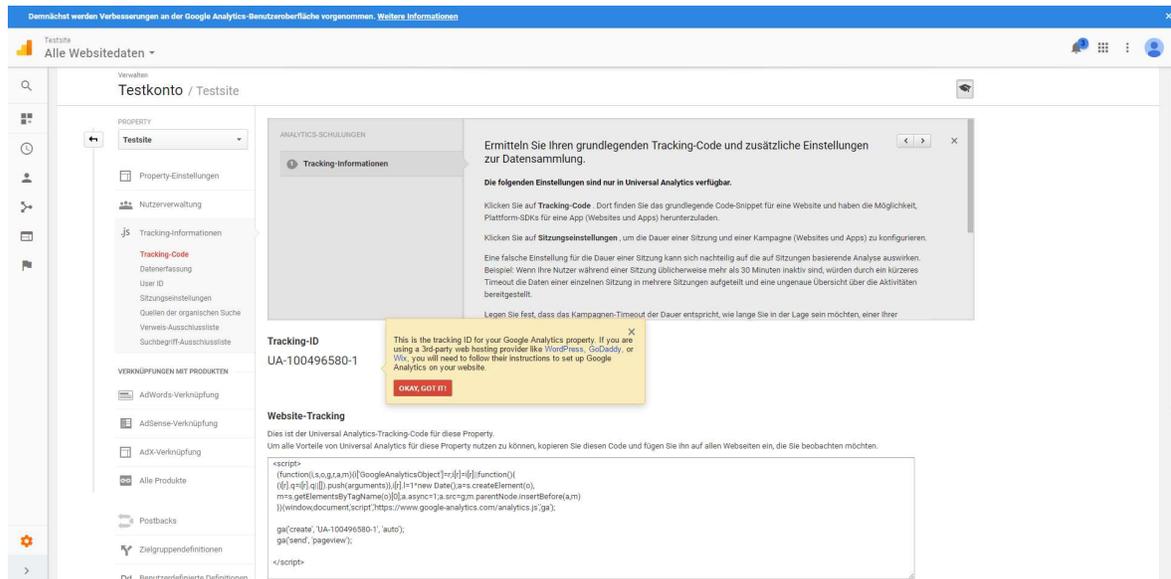


Abbildung 28: Generierter Tracking Code von Google

- Zur Einführung der einzelnen Funktionen bietet Google unter den Hauptnavigationen jeweils eine Einführung mit einem Video an. Abbildung 29 zeigt ein Beispielbild:

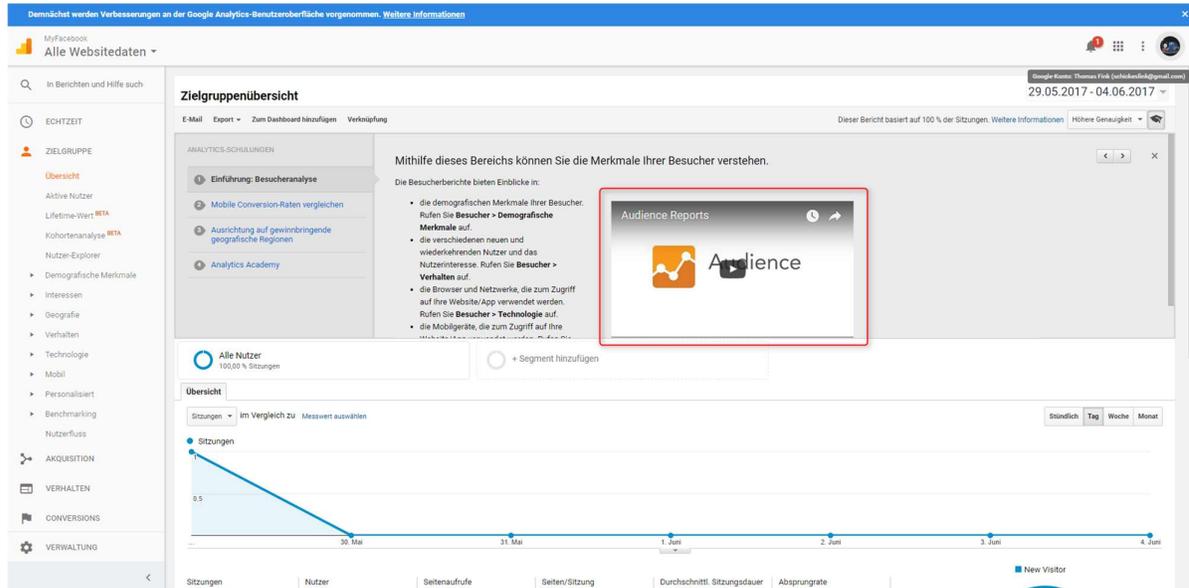


Abbildung 29: Google Konsole mit Einführungsvidéos

## 4.3.1.2. Datenschutz in Google Analytics

Wie bereits bei Punkt fünf in der Installationsanleitung erwähnt, bietet Google Analytics den NutzerInnen Möglichkeiten an, die Datenschutzbestimmungen, die auch Google verwendet, für die eigene Webseite zu nützen und diese per Cut & Paste in die AGBs zu übernehmen. Ob diese dann im Endeffekt im Einklang mit den Bestimmungen im jeweiligen Land sind, in dem die Website betrieben wird, muss von den BetreiberInnen der Webseite überprüft werden. Sollte das Unternehmen mit höchst sensiblen Daten zu tun haben, gilt es darüber hinaus zu evaluieren, ob diese nicht auf eigenen Servern, in sogenannten Inhouse-Lösungen besser aufgehoben sind. Auch wird von einem Großteil der ExpertInnen davon ausgegangen, dass die Datenschutzrichtlinien in Europa noch weiter verschärft werden. Im gleichen Atemzug wird aber auch angemerkt, dass Google mit Sicherheit ein besonderes Interesse daran hat, den europäischen Richtlinien weiterhin zu genügen, um so auf dem Analytics Markt keine Marktanteile zu verlieren. Welche Bestimmungen bei Google Analytics in jedem Fall beachtet werden müssen, wird in den darauffolgenden Ausführungen beleuchtet.

### Anonymisieren der IP-Adresse

Da der von Google vorgegebene Tracking-Code die IP-Adresse nicht anonymisiert, wurde eine Code-Erweiterungen mit dem Namen *anonymizeIp* von Google zur Verfügung gestellt. Durch die Nutzung dieser Erweiterung werden die letzten 8 Bits der IP-Adresse gelöscht. Eine grobe Lokalisierung ist zwar dennoch möglich, jedoch können UserInnen nicht mehr auf den einzelnen Rechner zurückverfolgt werden (Datenschutzbeauftragter Info, 2017). Die Anonymisierung der IP

wird erreicht, in dem im Tracking-Code *anonymizeIp* vor dem Senden des Events wie in Listing 6 deklariert wird:

```
<script>

(function(i,s,o,g,r,a,m){i['GoogleAnalyticsObject']=r;i[r]=i[r]||function(
){
(i[r].q=i[r].q||[]).push(arguments)},i[r].l=1*new
Date();a=s.createElement(o),
m=s.getElementsByTagName(o)[0];a.async=1;a.src=g;m.parentNode.insertBefore
(a,m)
})(window,document,'script','//www.google-
analytics.com/analytics.js','ga');

ga('create', 'UA-XXXXXXX-X', 'website.de');
ga('set', 'anonymizeIp', true);
ga('send', 'pageview');

</script>
```

Listing 6: Anonymisierung der IP-Adresse in Google-Analytics (Datenschutzbeauftragter Info, 2017)

### Ergänzung der Datenschutzerklärung auf der Website

Zusätzlich zu den bisherigen Anpassungen muss ebenso die Datenschutzerklärung angepasst werden. Hierbei kann folgender Hinweis von Google selbst, mit Anpassungen von Datenschutzbeauftragter Info (2017) übernommen werden:

*Diese Website benutzt Google Analytics, einen Webanalysedienst der Google Inc. („Google“). Google Analytics verwendet sog. „Cookies“, Textdateien, die auf Ihrem Computer gespeichert werden und die eine Analyse der Benutzung der Website durch Sie ermöglichen. Die durch das Cookie erzeugten Informationen über Ihre Benutzung dieser Website werden in der Regel an einen Server von Google in den USA übertragen und dort gespeichert. Im Falle der Aktivierung der IP-Anonymisierung auf dieser Website, wird Ihre IP-Adresse von Google jedoch innerhalb von Mitgliedstaaten der Europäischen Union oder in anderen Vertragsstaaten des Abkommens über den Europäischen Wirtschaftsraum zuvor gekürzt. Nur in Ausnahmefällen wird die volle IP-Adresse an einen Server von Google in den USA übertragen und dort gekürzt. Im Auftrag des Betreibers dieser Website wird Google diese Informationen benutzen, um Ihre Nutzung der Website auszuwerten, um Reports über die Websiteaktivitäten zusammenzustellen und um weitere mit der Websitenutzung und der Internetnutzung verbundene Dienstleistungen gegenüber dem Websitebetreiber zu erbringen. Die im Rahmen von Google Analytics von Ihrem Browser übermittelte IP-Adresse wird nicht mit anderen Daten von Google zusammengeführt. Sie können die Speicherung der Cookies durch eine entsprechende Einstellung Ihrer Browser-Software verhindern; wir weisen Sie jedoch darauf hin, dass Sie in diesem Fall gegebenenfalls nicht sämtliche Funktionen dieser Website vollumfänglich werden nutzen können. Sie können darüber hinaus die Erfassung der durch das Cookie erzeugten und auf Ihre Nutzung der Website bezogenen Daten (inkl. Ihrer IP-Adresse) an Google sowie die Verarbeitung dieser Daten durch Google verhindern, indem Sie das unter dem folgenden Link (<http://tools.google.com/dlpage/gaoptout?hl=de>) verfügbare Browser-Plugin herunterladen und installieren.*

*Sie können die Erfassung durch Google Analytics verhindern, indem Sie auf folgenden Link klicken. Es wird ein Opt-Out-Cookie gesetzt, das die zukünftige Erfassung Ihrer Daten beim*

*Besuch dieser Website verhindert:*  
<a href="javascript:gaOptout()">Google Analytics deaktivieren</a>

*Nähere Informationen zu Nutzungsbedingungen und Datenschutz finden Sie unter <http://www.google.com/analytics/terms/de.html> bzw. unter <https://www.google.de/intl/de/policies/>. Wir weisen Sie darauf hin, dass auf dieser Website Google Analytics um den Code „anonymizelp“ erweitert wurde, um eine anonymisierte Erfassung von IP-Adressen (sog. IP-Masking) zu gewährleisten (Datenschutzbeauftragter Info, 2017).*

## Widerspruchsrecht

BesucherInnen der Website muss die Möglichkeit geboten werden, gegen die Erfassung von Nutzungsdaten ein Widerspruchsrecht einzulegen. Dies muss auf folgende zwei Arten geschehen:

- Link zum Deaktivierungs-Add-On für Browser
- Setzen eines Opt-out-Cookies

Das Add-On für Browser verhindert beim Aufruf einer Webseite die Ausführung von Google Analytics. Als Webseiten-BetreiberIn ist man dazu verpflichtet, den Link zum Add-On auf der Webseite zu veröffentlichen.

Die andere Notwendigkeit ist, durch klicken eines Links, die Möglichkeit zur Verfügung zu stellen, für BesucherInnen ein Opt-Out-Cookie zu setzen. Durch dieses kann gewährleistet werden, dass das Tracking nicht ausgeführt wird. Zur Implementierung dieser Option muss folgendes Skript, welches in Listing 7 zu sehen ist, vor dem Google Analytics Script eingefügt werden:

```
<script>
var gaProperty = 'UA-XXXXXXX-X';
var disableStr = 'ga-disable-' + gaProperty;
if (document.cookie.indexOf(disableStr + '=true') > -1) {
  window[disableStr] = true;
}
function gaOptout() {
  document.cookie = disableStr + '=true; expires=Thu, 31 Dec 2099 23:59:59
UTC; path=/';
  window[disableStr] = true;
}
</script>
```

Listing 7: Einfügen der Opt-Out-Möglichkeit

## Vertrag zur Auftragsdatenverarbeitung

Da Google im rechtlichen Sinne als AuftragnehmerIn von WebseitenbetreiberInnen handelt, empfiehlt es sich dementsprechend auch, eine Vereinbarung zur Auftragsdatenverarbeitung abzuschließen. Hierbei wird von Google ein vorgefertigtes Dokument angeboten:

<https://goo.gl/7HrtHK>.

### **4.3.1.3. Pro und Kontra Google Analytics**

#### Pro

- Zahlreiche Informationen in Form von Dokumenten und Tutorials online und kostenfrei erhältlich.
- Aus- und Weiterbildungsangebot ausgesprochen vielfältig.
- Quasi das „Sorglos Paket“, bei dem sämtliche Hardware zur Verfügung gestellt wird.
- Großes Angebot an Agenturen die Support anbieten.
- Möglichkeit zur Verknüpfung mit anderen Google Diensten wie Google AdWords.
- Telefonsupport sobald bei Google AdWords kostenpflichtige Werbung geschaltet wird.
- Bereits in der kostenlosen Version große Auswahl an Funktionen.

#### Kontra

- Die erfassten Daten sind nicht im eigenen Verantwortungsbereich gesammelt.
- Laufende Änderungen der Datenschutzrichtlinien, die somit immer beobachtet werden sollten.

### **4.3.2 Einführung in Piwik**

Neben dem Platzhirsch Google Analytics gibt es mit Piwik eine weitere Möglichkeit, mit der Digital Analytics in der Webpräsenz eingesetzt werden kann. So wie Google Analytics ist auch Piwik kostenlos verfügbar. Im Gegensatz zu Google Analytics handelt es sich bei Piwik aber um eine Open Source Software. Das bedeutet, dass man nicht nur in den Quellcode blicken kann, sondern diesen auch an die eigenen Bedürfnisse im Rahmen der GNU (General Public License) anpassen kann. Hinsichtlich der datenschutzrechtlichen Aspekte resultiert ein großer Unterschied durch den Umstand, dass Piwik auf den eigens zur Verfügung gestellten Servern installiert werden kann und somit auch die gesammelten Daten nicht auf Fremdservern gespeichert werden. Aus diesem Grund fallen zumindest Probleme aufgrund der Weitergabe von persönlichen Daten an Dritte weg. Dies bedeutet wiederum, dass Piwik selbst installiert und gewartet werden muss. So wie bei Google Analytics muss auch bei Piwik ein Tracking Code in die Website eingefügt werden. Wird die Seite mit dem eingefügten Tracking Code in weiterer Folge besucht, wird der Tracking Code durchlaufen, die Daten des Besuchers/der Besucherin werden erfasst und an den Server weitergeleitet, auf dem die Piwik-Installation läuft.

#### **4.3.2.1. Installation von Piwik**

1. Vor dem Beginn der Installation müssen folgende Voraussetzungen erfüllt sein:

- Ein Webserver, auf dem die Piwik Installation durchgeführt wird, muss zur Verfügung stehen.

- Der Zugriff auf diesen Webserver muss via http oder shell erfolgen.
  - Ein FTP Client muss installiert werden.
2. Herunterladen der aktuellsten Piwik Version von [goo.gl/6zwvBx](http://goo.gl/6zwvBx)
  3. Das herunter geladene File kann nun lokal auf der Festplatte entpackt werden
  4. Mit einem FTP Client (z.B. FileZilla oder WinSCP) können die Piwik Files nun im *binary mode* auf den Webserver geladen werden. Diese können zum Beispiel in einem Unterordner im öffentlichen www Verzeichnis abgelegt werden.
  5. Sobald der Upload abgeschlossen ist, ruft man die Installation über einen Web Browser auf. Zum Beispiel über <http://www.meinewebsite.at/analytics>.
  6. Wurde der Upload korrekt abgeschlossen, wird nun ein Fenster zum Start der Installation angezeigt.
  7. Mit einem Klick auf den *Next* Button wird die Installation gestartet.
  8. Wie in Abbildung 30 zu sehen ist, wird in Schritt zwei überprüft, ob alle Voraussetzungen für die Installation von Piwik erfüllt sind. Sofern ein Kriterium nicht erfüllt wird, zeigt Piwik dies an und erläutert, welche Schritte in diesem Fall als nächstes durchzuführen sind.



Abbildung 30: Software Voraussetzungen für die Installation (Quelle: <https://piwik.org/wp-content/uploads/2008/11/2a-check-success1-700x332.png>, 4.6.2017)

9. Im nächsten Schritt müssen die Informationen zur MySQL Datenbank angegeben werden. Sollte noch keine Datenbank zur Verfügung stehen, muss diese in Eigenregie oder von Hosting AnbieterInnen erstellt werden.
10. In Schritt 6 müssen Zugangsdaten für den Super User/die Super Userin deklariert werden. Der Super User/Die Super Userin besitzt die höchsten Zugriffsrechte. Der Super User/Die

Super Userin wird benötigt, um sich erstmalig anzumelden, Rechte für weitere BenutzerInnen zu setzen und um Plugins zu aktivieren oder deaktivieren. Siehe dazu Abbildung 31:

The screenshot shows the 'Super User' creation form. On the left is a navigation menu with steps 1 through 9, where step 6 'Super User' is highlighted. The form itself has a title 'Super User' and contains the following fields and options:

|                                     |   |
|-------------------------------------|---|
| super user login                    | <input type="text"/>  |
| password                            | <input type="password"/>  |
| password (repeat)                   | <input type="password"/>  |
| email                               | <input type="text"/>  |
| <input checked="" type="checkbox"/> | email me with major Piwik upgrades and security alerts            |
| <input checked="" type="checkbox"/> | email me with community updates (new plugins, new features, etc.) |

Abbildung 31: Erstellung des Super Users (Quelle: <https://piwik.org/wp-content/uploads/2008/11/3-database1.png>, 4.6.2017)

11. In Schritt 7 muss – wie in Abbildung 32 zu sehen – nun die Webseite angegeben werden, die mit Piwik aufgezeichnet werden soll. Weitere URLs können erst hinzugefügt werden, sobald die Installation abgeschlossen ist.

The screenshot shows the 'Setup a Website' form. On the left is a navigation menu with steps 1 through 9, where step 7 'Setup a Website' is highlighted. The form itself has a title 'Setup a Website' and contains the following fields:

|                   |   |
|-------------------|---|
| website name      | <input type="text" value="Example"/>  |
| website URL       | <input type="text" value="http://"/>  |
| website time zone | <input type="text" value="Select a city"/> <input type="button" value="v"/> |

Abbildung 32: Hinzufügen der Webseite (Quelle: <https://piwik.org/wp-content/uploads/2008/11/6-setup-website1.png>, 4.6.2017)

12. Im vorletzten Schritt wird jener JavaScript Tag angezeigt, der auf jeder Seite eingebettet werden muss, die von Piwik getracked werden soll – siehe Abbildung 33. Es wird empfohlen, den Code unmittelbar vor dem schließenden `</head>` Tag einzufügen. Sofern ein Content Management System wie WordPress, Drupal oder Joomla! verwendet wird, können Add-Ons, Erweiterungen oder Plugins dafür genutzt werden.

**Piwik** # Open Source Web Analytics

- Welcome!
- System Check
- Database Setup
- Database Check
- Creating the Tables
- General Setup
- Setup a Website
- JavaScript Tag**
- Congratulations

### Tracking tags for Example

To record visitors, visits and page views in Piwik, you must add a Tracking code in all your pages. We recommend to use the standard Javascript Tracking tag.

#### Standard Javascript Tracking tag

Copy and paste the following code in all the pages you want to track with Piwik. In most websites, blogs, CMS, etc. you can edit your website templates and add this code in a "footer" file.

Here is the JavaScript Tracking tag to include on all your pages, just before the </body> tag.

```
<!-- Piwik -->
<script type="text/javascript">
var pkBaseURL = (("https:" == document.location.protocol) ? "https://localhost" : "http://localhost/piwik/piwik/");
document.write(unescape("%3Cscript src='" + pkBaseURL + "piwik.js"
type='text/javascript'%3E%3C/script%3E"));
</script><script type="text/javascript">
try {
var piwikTracker = Piwik.getTracker(pkBaseURL + "piwik.php", 1);
piwikTracker.trackPageView();
piwikTracker.enableLinkTracking();
} catch(e) {}
</script>
```

Abbildung 33: Darstellung des Tracking Codes (Quelle: <https://piwik.org/wp-content/uploads/2008/11/7-javascript-tag1-700x378.png>, 4.6.2017)

Links zu den Plugins folgender Content Management Systeme:

- Drupal Plugin: <https://goo.gl/1ocYkK>
- Joomla! Extension: <https://goo.gl/2GFd2H>
- Wordpress Plugin: <https://goo.gl/G4VLJB>

13. Sobald der Code eingefügt wurde, gelangt man mit einem Klick auf *Next* zur letzten Seite der Installation, die somit abgeschlossen ist. In Abbildung 34 ist das Dashboard von Piwik zu sehen:

**PIWIK** Dashboard Alle Websites Deutsch

Suche: VIRTUAL-DRUMS.COM 2017-06-03 ALLE BESUCHE DASHBOARD SIE SEHEN GERADE DIE DEMO VON PIWIK

**Besucher in Echtzeit**

| Datum             | BESUCHE | AKTIVEN |
|-------------------|---------|---------|
| Letzte 24 Stunden | 16      | 18      |
| Letzte 30 Minuten | 1       | 1       |

**Graph der letzten Besuche**

18 Besuche

**Besucherkarte**

18 Besuche

**Verweistypen**

| VERWEISART       | BESUCHE | EINDEUTIGE BESUCHER | SEKUNDEN |
|------------------|---------|---------------------|----------|
| Direkte Zugriffe | 10      | 10                  |          |
| Suchmaschinen    | 6       | 5                   |          |
| Websites         | 2       | 2                   |          |

**Besucherübersicht**

- 18 Besuche, 17 Eindeutige Besucher
- 2 Minuten 28s durchschnittliche Aufenthaltsdauer
- 67% abgesprungene Besucher (Verlassen der Webseite nach einer Seite)
- 1,7 Aktionen pro Besuch (Seitenansichten, Downloads, ausgehende Verweise und interne Suchen)
- 0,25s durchschnittliche Generierungszeit

**Premium Products & Services for Piwik**

Increase the conversions on your online forms, by learning everything about your users behavior and their pain points on your forms

**LEARN MORE**

**Piwik.org Blog**

Understand your visitors by seeing where they click, hover, type and scroll, and replay their actions in a video

Abbildung 34: Dashboard von Piwik (Quelle: <http://demo.piwik.org/index.php?module=CoreHome&action=index&idSite=3&period=day&date=yesterday>)

ay#?idSite=3&period=day&date=yesterday&category=Dashboard\_Dashboard&subcategory=1, 4.6.2017)

Dieses bietet in der kostenlosen Version Statistiken zu BesucherInnen, Ausstiegsseiten, Einstiegsseiten, Suche sowie Verweise an.

#### 4.3.2.2. Datenschutz in Piwik

So wie auch bei Google Analytics tragen auch bei Piwik die BenutzerInnen der Lösung die Letztverantwortung was das Thema Datenschutz betrifft. Da sich die Richtlinien dazu laufend ändern können, wird empfohlen, sich vor der Installation von Piwik über die aktuelle Situation zu informieren.

##### Anonymisieren der IP-Adresse

So wie bei Google Analytics muss auch bei Piwik dafür gesorgt werden, dass die IP-Adresse anonymisiert wird. Die Einstellungen zur Anonymisierung der IP-Adresse befindet sich unter *Administration* → *Privacy*. Hier kann in weiterer Folge entschieden werden, welcher Grad der Verschlüsselung gewählt werden soll. Grundsätzlich reicht hier, so wie auch bei Google Analytics, im rechtlichen Sinne die 8 Bit Verschlüsselung. Der Unterschied zu jener mit 18 Bit ist, dass hier die genaue geographische Ortung nicht mehr möglich ist. Abbildung 35 zeigt ein Beispielfeld der Installation:

Anonymize Visitors' IP addresses

Yes  
 No

Anonymize the last byte(s) of visitors IP addresses to comply with your local privacy laws/guidelines. Select "Yes" if you want Piwik not to track fully qualified IP-Addresses.

Select how many bytes of the visitors' IPs should be masked.

1 byte(s) - e.g. 192.168.100.xxx  
 2 byte(s) - e.g. 192.168.xxx.xxx (Recommended)  
 3 byte(s) - e.g. 192.xxx.xxx.xxx

Note: Geolocation will have approximately the same results with 1 byte anonymized. With 2 bytes or more, Geolocation will be inaccurate.

Also use the Anonymized IP addresses when enriching visits.

Yes (Recommended for privacy)  
 No

Plugins such as Geo Location via IP and Provider improve visitor's metadata. By default these plugins use the anonymized IP addresses. If you select 'No', then the non-anonymized full IP address will be used instead, resulting in less privacy but better data accuracy.

Save

Abbildung 35: Anonymisierung der IP-Adresse (Quelle: <https://piwik.org/wp-content/uploads/2013/11/privacy.png>, 1.6.2017)

## Widerspruchsrecht

Auch bei Piwik muss BesucherInnen der Webseite die Möglichkeit gegeben werden, Widerspruch gegen die Erfassung von Nutzungsdaten einzuräumen bzw. sie auf die Verwendung der Daten hinzuweisen. Bei Piwik findet sich der Code zum Setzen der Opt-Out-Cookie Option unter *Administration* → *Privacy*. Es empfiehlt sich, den Code, der in Listing 8 zu sehen ist, in die Datenschutzbestimmungen zu integrieren:

```
<iframe frameborder="no" width="600px" height="200px"
src="http://www.example.com/x_piwik/index.php?module=CoreAdmin
Home&action=optOut"></iframe>
```

Listing 8: Code um Opt-Out-Möglichkeit in Datenschutzbestimmungen zu integrieren (Datenschutzbeauftragter Info, 2017)

Es ist hier zu beachten, dass dieser Text in englischer Sprache erscheint. In Folge dessen muss dieser Text in der Datei *lang/en.php* beziehungsweise *lang/de.php* manuell angepasst werden.

### Ergänzung der Datenschutzerklärung auf der Website

In Kombination mit der Opt-Out Funktion kann in den Datenschutzhinweisen folgender Text von Datenschutzbeauftragter Info (2017) übernommen werden:

*Unsere Website verwendet Piwik, dabei handelt es sich um einen sogenannten Webanalysedienst. Piwik verwendet sog. „Cookies“, das sind Textdateien, die auf Ihrem Computer gespeichert werden und die unsererseits eine Analyse der Benutzung der Webseite ermöglichen. Zu diesem Zweck werden die durch den Cookie erzeugten Nutzungsinformationen (einschließlich Ihrer gekürzten IP-Adresse) an unseren Server übertragen und zu Nutzungsanalysezwecken gespeichert, was der Webseitenoptimierung unsererseits dient. Ihre IP-Adresse wird bei diesem Vorgang umgehend anonymisiert, so dass Sie als Nutzer für uns anonym bleiben. Die durch den Cookie erzeugten Informationen über Ihre Benutzung dieser Webseite werden nicht an Dritte weitergegeben. Sie können die Verwendung der Cookies durch eine entsprechende Einstellung Ihrer Browser Software verhindern, es kann jedoch sein, dass Sie in diesem Fall gegebenenfalls nicht sämtliche Funktionen dieser Website voll umfänglich nutzen können.*

*Wenn Sie mit der Speicherung und Auswertung dieser Daten aus Ihrem Besuch nicht einverstanden sind, dann können Sie der Speicherung und Nutzung nachfolgend per Mausklick jederzeit widersprechen. In diesem Fall wird in Ihrem Browser ein sog. Opt-Out-Cookie abgelegt, was zur Folge hat, dass Piwik keinerlei Sitzungsdaten erhebt. Achtung: Wenn Sie Ihre Cookies löschen, so hat dies zur Folge, dass auch das Opt-Out-Cookie gelöscht wird und ggf. von Ihnen erneut aktiviert werden muss (Datenschutzbeauftragter Info, 2017).*

### 4.3.2.3. Pro und Kontra Piwik

Pro:

- Kostenlose Version verfügbar.
- Open Source Erweiterungen können speziell an eigene Bedürfnisse angepasst werden.
- Bei Installation auf eigens zur Verfügung gestellten Servern vermeidet man das Problem der Weitergabe von Daten an Dritte.

Contra

- Man muss sich um eigenes Hosting kümmern.
- Im Vergleich zu Google Analytics sind weniger Agenturen darauf spezialisiert.
- Community im Vergleich zu Google Analytics sehr klein.
- In der kostenlosen Version sind nur sehr wenige Funktionen verfügbar.

#### 4.4 Themenblock 4: Datenschutzrechtliche Aspekte

In diesem Themenblock werden folgende Fragen beantwortet:

- Gibt es datenschutzrechtliche Aspekte, die bei der Auswahl einer Analytics Lösung in jedem Fall zu beachten sind?
- Wie unterscheiden sich Piwik und Google Analytics hinsichtlich des Datenschutzes?
- Kann unter diesen Gesichtspunkten eine Empfehlung für eine Analytics Lösung abgegeben werden?

Zum Thema der datenschutzrechtlichen Aspekte sei vorausgeschickt, dass eine vollständige Sicherheit diesbezüglich nur von JuristInnen gewährleistet werden kann. Im Bereich der Analytics Lösungen bietet Google Analytics seinen AnwenderInnen Auflagen an, denen Genüge getan werden muss, um sie mit den Datenschutzrichtlinien des jeweiligen Heimatlandes in Einklang zu bringen. Diese werden in Form von AGB aufbereitet und so den AnwenderInnen zur Verfügung gestellt. Hierbei muss man sich aber im Klaren sein, dass diese Auflagen nur in Form von Empfehlungen vorliegen. Es besteht also auf Seiten der UserInnen keine Pflicht, diese auch tatsächlich umzusetzen bzw. überprüft Google nicht, ob diesen AGBs Folge geleistet wird. Die Letztverantwortung liegt hier also im Bereich des Anwenders/der Anwenderin. Wichtige Punkte hierbei sind, dass im Impressum erkenntlich gemacht wird, dass per Tracking Daten gesammelt werden und diese an Google gesendet werden und dass keine Speicherung von personenbezogenen Daten stattfindet. Personenbezogene Daten können Vor- und Nachname, Geburtsdatum, Adressen und Email-Adressen von BesucherInnen einer Website sein. Ein weiterer sensibler Punkt ist die IP-Adresse. Wie in Kapitel 2.4.4 beschrieben, gilt in Österreich die IP-Adresse als personenbezogener Datensatz und darf somit nicht gespeichert werden. Google Analytics bietet hier den Service an, dass die Einstellung getroffen werden kann, die letzten drei Stellen der IP-Adresse nicht zu speichern, um so eine Identifizierung der Person zu verhindern. Darüber hinaus gilt es, den BesucherInnen ein Widerspruchsrecht gegen die Sammlung und Speicherung von Daten durch ein Opt-Out Feld zu ermöglichen.

Unternehmen, die mit besonders sensiblen Daten zu tun haben, sollten sich darüber Gedanken machen, ob es aus ihrer Sicht sinnvoll wäre, aus Sicherheitsgründen die gesammelten Daten auf eigenen Servern zu speichern und nicht auf jenen des Analytics Anbieters. Was die Entscheidung für oder gegen eine der beiden vorgestellten Analytics Lösung anbelangt, können die ExpertInnen keine klare Empfehlung abgeben. Fakt ist, dass beide Lösungen ähnliche Methoden nutzen, um die Daten von BesucherInnen einer Website zu speichern. Auch müssen bei beiden Lösungen Einstellungen wie das Anonymisieren der IP-Adresse oder das setzen der Opt-Out Funktion

manuell getroffen werden. Der einzige Unterschied ist, dass bei Piwik die Möglichkeit besteht, die Daten auf eigens zur Verfügung gestellten Servern zu speichern. Das wäre ein möglicher Grund, Piwik als bessere Lösung im Hinblick auf den Datenschutz darzustellen.

**Aus Themenblock 4 ergeben sich somit folgende konkrete Handlungsempfehlungen:**

- Im Impressum erkenntlich machen, dass Daten per Tracking gesammelt werden.
- Anonymisierung der IP-Adresse, da diese in Österreich und Deutschland als personenbezogener Datensatz gilt.
- Den BesucherInnen ein Widerspruchsrecht gegen die Sammlung und Speicherung von Daten einräumen.
- Es muss immer beachtet werden, dass die BetreiberInnen einer Webseite, welche eine Analytics Lösung im Einsatz haben, die Letztverantwortung zur Einhaltung von Datenschutzbestimmungen innehaben.
- Im Zweifelsfall wird empfohlen, JuristInnen, welche sich auf dieses Gebiet spezialisiert haben, zu konsultieren.

## 4.5 Themenblock 5: Zukunftsaussichten von Digital Analytics

In diesem Themenblock werden folgende Fragen beantwortet:

- In welche Richtung wird sich Digital Analytics weiterentwickeln?
- Wie wird sich die Marktsituation hinsichtlich der vorhandenen Analytics Lösungen entwickeln?
- Welche Herausforderungen kommen auf die AnwenderInnen zu?
- Kann der Zielgruppe uneingeschränkt die Nutzung von Digital Analytics empfohlen werden?

Eine große Herausforderung für die nächsten Jahre sehen ExpertInnen im Bereich der Datenanalyse (Data Science). Die große Daten-Sammelwut der Unternehmen weicht von Big-Data immer mehr in die Disziplin des Smart Data ab. Das bedeutet, dass nicht die Menge an Daten der essentielle Faktor einer nutzenbringenden Datenanalyse ist, sondern dass zum einen die „richtigen“ Daten gesammelt werden und diese dann in weiterer Folge auch mit den dafür geeigneten Methoden ausgewertet werden. Im gesamten Prozess macht der Aufwand der Implementation der Analytics Lösung nur 5 bis 10 Prozent aus. Die restlichen 90 bis 95 Prozent werden in Integration von KPIs, Datenanalyse und Datenauswertung investiert. Gerade für KMUs startet hier soeben eine Bewegung, die das sogenannte Data Mining auch für diese Zielgruppe zugänglich macht. Im Detail geht es darum, dass mathematische Datenmodelle, die für die Auswertung benötigt werden, auch für wesentlich kleinere Datenmengen – welche im Bereich der KMUs anfallen – anwendbar gemacht werden.

Ein weiterer wesentlicher Punkt wird die Evaluierung der Customer Journey von Webseiten BesucherInnen sein, sowie deren Kaskadierung. Die Customer Journey beschreibt sämtliche

messbaren Kontaktpunkte von NutzerInnen auf dem Weg zu einer definierten Aktion. Wichtig hierbei ist, dass alle Marketing Kanäle berücksichtigt werden. Dazu zählen verschiedenste Social-Media-Kanäle, Web- oder mobile Plattformen (BVDW e.V., 2012).

Laut ExpertInnen wird es hier immer essentieller zu wissen, welche Personengruppen über welche Kanäle die Webseite besuchen, um so noch gezieltere Marketingkampagnen durchzuführen. Hier spricht man auch von der Kaskadierung der NutzerInnen. Durch diese rückt mit Dynamic Pricing ein weiteres Thema stärker in den Mittelpunkt. Dynamic Pricing bedeutet zum Beispiel, dass durch die Informationen über Ort und Device von dem NutzerInnen eine Website besuchen, die Produktpreise personenbezogen angepasst werden können. Aus diesem Gesichtspunkt wird, wie auch im ersten Abschnitt dieses Themenblocks bereits genannt, der Einsatz von Data Scientists bzw. von DienstleisterInnen, die dies für KMUs anbieten, immer relevanter.

Google Analytics ist der aktuelle Platzhirsch am Markt der Analytics Lösungen und wird diesen Platz auch festigen bzw. vergrößern. Besonders für KMUs ist der Einstieg mit Google Analytics aufgrund seiner Basisfunktionen in der Gratisversion uneingeschränkt zu empfehlen. Gründe, um mit anderen Analytics Systemen den Einstieg zu wagen, können entweder datenschutzrechtliche Aspekte sein oder gewisse Nischen, auf die sich AnbieterInnen spezialisiert haben.

Von einem Teil der ExpertInnen wird nicht uneingeschränkt zum Einsatz von Digital Analytics geraten. Bevor dieser Schritt gewagt wird, sollte überprüft werden, ob überhaupt ein Wettbewerbsvorteil durch den Einsatz von Digital Analytics erzielt werden kann und welche digitalen Datensätze zur Auswertung überhaupt zur Verfügung stehen. Sollten diese beiden Fragen positiv beantwortet werden können, wird zu einem Einsatz von Digital Analytics dringend angeraten. Andere ExpertInnen wiederum vergleichen einen Web Auftritt ohne Digital Analytics Integration als ein Auto ohne Tacho. Man weiß zwar, dass das Auto sich bewegt, kann aber keine weiteren Informationen dazu geben bzw. auf besondere Situationen wie Geschwindigkeitsbeschränkungen reagieren. Um die Dringlichkeit für Digital Analytics im Besonderen für KMUs zu veranschaulichen, wurde als plakatives Beispiel ein Vergleich zwischen dem Alleinstellungsmerkmal bei Unternehmen vor und nach der digitalen Transformation genannt. In dieser Zeit war unter anderem der Standort der essentielle Faktor, weshalb Unternehmen positiv wirtschaften konnten oder nicht. Im Zuge der Digitalisierung ist dieser Vorteil vor allem gegenüber den „*Big Playern*“ im E-Commerce abhandengekommen, da über ein mögliches Geschäft nicht nur mehr die geographische Entfernung in Kilometer oder Meter entscheidet, sondern nur ein paar Klicks.

**Aus Themenblock 5 ergeben sich somit folgende konkrete Handlungsempfehlungen:**

- Die Entwicklung von Digital Analytics im Auge behalten, um sich auch in Zukunft einen Wettbewerbsvorteil gegenüber KonkurrentInnen zu sichern.
- Prüfen, ob die im Unternehmen generierten Daten für Data Science relevant sind.
- Debatten rund um die datenschutzrechtlichen Aspekte verfolgen und bei Änderungen entsprechende Handlungen setzen.

## 5 RESÜMEE

Das Ziel dieser Masterarbeit war es, einen Handlungsleitfaden zur Integration von Analytics Lösungen für Klein- und Mittelunternehmen im Bereich E-Commerce zu erstellen. Der Handlungsfadens deckt dabei die initiale Phase ab, bei der die Anforderungen an das Unternehmen und an die Analytics Lösung dargestellt werden sollen. Weiters wird der Prozess zur Findung von KPIs bzw. welche KPIs für die Zielgruppe laut ExpertInnen in jedem Fall interessant sind, beleuchtet. Welche Anforderungen eine Analytics Lösung im Hinblick auf die datenschutzrechtlichen Aspekte erfüllen soll und wie sich dieses Fachgebiets in Zukunft entwickeln wird, sollte ebenfalls in Rahmen dieser Arbeit behandelt werden.

Die Ergebnisse, die im Zuge der Untersuchung zu Tage kamen, sollen der Zielgruppe der KMUs Hilfestellung bei der Integration von Digital Analytics in ihrem Unternehmen bieten. Die für den Handlungsleitfaden befragten ExpertInnen weisen dabei alle mehrjährige Erfahrungen in diesem Bereich auf. Bei der Erstellung des Handlungsleitfadens wurde darauf geachtet, dass sich die Reihenfolge der darin enthaltenen Information mit der von den ExpertInnen vorgeschlagenen Vorgangsweise bei der Integration von Digital Analytics deckt.

Das erste Themengebiet befasste sich demnach mit den Basisanforderungen an das Unternehmen. Der Zielgruppe wird dabei vermittelt, ob bzw. welche Anforderungen an das Unternehmen gestellt werden, um eine effiziente Integration von Digital Analytics im Unternehmen zu gewährleisten. Das Augenmerk wurde dabei auf Anforderungen der IT-Infrastruktur, technischem Knowhow und dem Zielfindungsprozess gelegt. Im Zuge der Befragung stellte sich heraus, dass dies immer in Abhängigkeit der ausgewählten Analytics Lösung steht und somit nicht pauschal beantwortet werden kann. Wird etwa als Analytics Lösung Google Analytics gewählt, so bestehen keine Anforderung an die IT-Infrastruktur im Unternehmen, da diese hier von Google zur Verfügung gestellt wird. Auch der Faktor Knowhow spielt bei Google Analytics keine große Rolle, da hier ein überaus vielfältiges Angebot an frei verfügbarer Dokumentation und zertifizierten Agenturen besteht. Neben diesen beiden Punkten wurde von den ExpertInnen immer wieder betont, dass es für das Unternehmen essentiell ist, sich über die eigenen Ziele im digitalen Geschäft bewusst zu sein. Dies erstreckt sich konkret von den eigenen Zielen, die das Unternehmen erreichen will bis hin zu den Erwartungen, die man an Digital Analytics stellt. Dieser Prozess kann entweder eigenständig mit Hilfe des im Handlungsleitfaden vorgestellten OKR-Modells durchgeführt werden oder mit Hilfe von externen DienstleisterInnen wie UnternehmensberaterInnen. Ebenso wird von den ExpertInnen ausdrücklich empfohlen, von Anfang an die Kosten und den dadurch generierten Nutzen im Auge zu behalten.

Im darauffolgenden Block wurde das Thema rund um KPIs behandelt. Im speziellen ging es darum, wie Unternehmen den Prozess zur Findung von KPIs gestalten sollen, welche KPIs für die Zielgruppe in jedem Fall relevant sind und welchen Fehlern die ExpertInnen in diesem

Prozess immer wieder begegnen. Im Prozess zur Findung von KPIs empfehlen ExpertInnen den selben Ansatz wie zur Findung von Unternehmenszielen – entweder eigenständig mit Hilfe des OKR-Modells oder durch Beratung mit externen DienstleisterInnen. Bei der Implementation dieser KPIs verweisen die ExpertInnen auf den agilen Ansatz. Dies bedeutet, dass die KPIs immer nur einzeln implementiert werden und danach einer zweiwöchigen Implementationsphase unterzogen werden. Erst wenn die Informationen, die aus dem KPI gewonnen werden können, den gewünschten Nutzen bringen, kann mit dem nächsten fortgefahren werden. Werden mehrere KPIs auf einmal implementiert, besteht die Gefahr, dass aus den gewonnenen Informationen falsche Schlüsse abgeleitet werden.

Im dritten Themenblock wurden Fragen hinsichtlich der Auswahl von Analytics Lösungen behandelt und welche Funktionen diese aufweisen sollten. Im Zuge der Befragung stellte sich heraus, dass die ExpertInnen der Zielgruppe uneingeschränkt kostenlose Analytics Lösungen wie Google Analytics und Piwik empfehlen. Aus diesem Grund werden diese beiden Analytics Lösungen im Handlungsleitfaden näher vorgestellt. Im Detail werden dabei allgemeine Informationen, Implementierung, Datenschutz und sowie Pro und Kontra der beiden Analytics Lösungen besprochen.

In Themenblock vier wurden die datenschutzrechtlichen Aspekte von Analytics Lösungen besprochen. Von den ExpertInnen wird der Zielgruppe dabei ans Herz gelegt, sich über Entwicklungen in diesem Bereich immer am Laufenden zu halten, um nicht mit dem Gesetz in Konfrontation zu geraten. Es wird dabei nochmals auf die in Themenblock drei vorgestellten Methoden, die bei den Analytics Lösungen Google Analytics und Piwik beachtet werden müssen, hingewiesen. Laut ExpertInnen ist damit zu rechnen, dass sich die Bestimmungen zum Datenschutz zumindest auf EU-Ebene weiter verschärfen werden. In Anbetracht dessen wird aber auch darauf hingewiesen, dass die AnbieterInnen von Analytics Lösungen danach trachten werden, diesen auch weiterhin Genüge zu tun, um nicht Boden am hart umkämpften Markt zu verlieren.

Im letzten Themenblock weisen die ExpertInnen nochmals auf den Nutzen durch den Einsatz von Analytics Lösungen hin. In Zukunft sehen die ExpertInnen die größte Herausforderung in der effizienten und nutzbringenden Auswertung von Datensätzen. Begriffe wie Smart Data und Data Science werden hier eine immer größere Rolle spielen. Weiters stellte sich bei den ExpertInneninterviews heraus, dass bei Fragen zu möglichen Analytics Lösungen in allen Fällen Google Analytics als populärste Möglichkeit genannt wurde.

In einer weiterführenden Arbeit wäre es interessant, sich noch stärker in das Gebiet Small und Smart Data für KMU zu vertiefen. Im Besonderen wäre ein Blick auf Hard- und Software Lösungen im Cloud- sowie Open-Source-Bereich spannend und gewiss auch für diese Zielgruppe von Nutzen. Im Speziellen wäre hier zu untersuchen, was sich in diesem Bereich die letzten Jahre getan hat und in welche Richtung sich diese Gebiete in Zukunft entwickeln werden. Hinsichtlich der Befragung zu dieser Thematik wären vor allem Personen in der Rolle eines Data-Scientists erstrebenswert.

## ABKÜRZUNGSVERZEICHNIS

|            |                                      |
|------------|--------------------------------------|
| AJAX       | Asynchronous Javascript and XML      |
| B2A        | Business-to-Administration           |
| B2B        | Business-to-Business                 |
| B2C        | Business-to-Customer                 |
| BDaaS      | Big Data as a Service                |
| C2A        | Consumer-to-Administration           |
| C2C        | Customer-to-Customer                 |
| CEP        | Complex Event Processing             |
| CMS        | Content Management System            |
| DACH       | Deutschland, Österreich und Schweiz  |
| DMS        | Dokumenten Management System         |
| DN         | Data Node                            |
| DSL        | Digital Subscriber Line              |
| E-Business | Electronic Business                  |
| E-Commerce | Electronic Commerce                  |
| ECM        | Enterprise Content Management System |
| ETL        | Extract Transform Load               |
| GIGO       | Garbage-In-Garbage-Out               |
| HDFS       | Hadoop Distributed File System       |
| JSON       | Javascript Object Notation           |
| KMU        | Klein- und Mittelunternehmen         |
| KPI        | Key Performance Indicators           |
| LTE        | Long Term Evolution                  |
| NN         | Name Node                            |
| NoSQL      | Not Only SQL                         |
| QL         | Query Language                       |
| ROI        | Return on Investment                 |
| SAN        | Storage Area Network                 |
| SNN        | Secondary Name Node                  |
| SQL        | Sequential Query Language            |
| XML        | Extensible Markup Language           |

## ABBILDUNGSVERZEICHNIS

|   |   |
|---|---|
| Abbildung 1: Kostendegressionseffekte bei der Verteilung digitaler Information (Wirtz, 2001).....   | 12  |
| Abbildung 2: Transaktionsphasen beim Verkauf über das Internet (Mülder, 2002).....  | 13  |
| Abbildung 3: Brutto-Jahresumsatz im Internet-Einzelhandel in € Mrd (Gittenberger & Vogl, 2014).....   | 14  |
| Abbildung 4: Jahresumsatz Internethandel weltweit in Milliarden Dollar (eMarketer Inc., 2016).....  | 15  |
| Abbildung 5: Oberfläche von Google Analytics.....   | 17  |
| Abbildung 6: Testing zweier Varianten: Welche spricht mehr BesucherInnen an? .....  | 18  |
| Abbildung 7: Beispiel einer Online-Umfrage bei Amazon.com (Quelle:<br><a href="https://gesterling.files.wordpress.com/2007/05/amazon-survey.jpg">https://gesterling.files.wordpress.com/2007/05/amazon-survey.jpg</a> , 1.7.2017) ..... | 19  |
| Abbildung 8: Funktionsbereiche von Digital-Analytics (Hassler, 2017).....   | 19  |
| Abbildung 9: Clientseitige Datensammlung mittels Page Tagging (eigene Darstellung in Anlehnung an<br>Hassler (2017).....  | 24  |
| Abbildung 10: Beim Aufruf von orf.at werden mehrere Hits erzeugt (orf.at, 2017) .....   | 25  |
| Abbildung 11: Automatische Weiterleitung am Ende eines Artikels (süddeutsche.de, 2017) .....  | 26  |
| Abbildung 12: Beispiel eines Hinweises zur Cookie-Verwendung von derstandard.at.....  | 28  |
| Abbildung 13: Welche Rolle spielen Daten in Ihrem Unternehmen (n=543) (Bange, Grosser, &<br>Janoschek, 2015) .....  | 32  |
| Abbildung 14: Big Data Initiativen gelistet nach Sparten (n=526) (Bange, Grosser, & Janoschek, 2015)  | 33  |
| Abbildung 15: Einsatz von Big Data in den Jahren, 2013, 2014 und 2015 (n=526) (Bange, Grosser, &<br>Janoschek, 2015) .....  | 33  |
| Abbildung 16: Welche Probleme/Herausforderungen wollen Sie mit Big Data Technologien bewältigen?<br>(n=431) (Bange, Grosser, & Janoschek, 2015).....  | 34  |
| Abbildung 17: Investition in Big Data (n=526) (Bange, Grosser, & Janoschek, 2015) .....   | 35  |
| Abbildung 18: Big Data als Ergänzung und Konkurrenz zur traditionellen IT (Bundesverband<br>Informationswirtschaft, 2012) .....   | 37  |
| Abbildung 19: Analyseansätze für traditionelle Systeme und Big Data Systeme (Bundesverband<br>Informationswirtschaft, 2012) .....   | 40  |
| Abbildung 20: Integrierte Anwendungslandschaft mit traditionellen Systemen und Big Data Lösungen<br>(Bundesverband Informationswirtschaft, 2012).....   | 41  |
| Abbildung 21: Aufteilung der Blöcke (WAGmob, 2013) .....  | 49  |
| Abbildung 22 HDFS und Replikation (WAGmob, 2013).....   | 50  |
| Abbildung 23: Das Interview als soziale Interaktion. Der Interviewer sollte ein kompetentes<br>Gesprächsklima bieten (Mieg & Brunner, 2001). .....  | <b>Fehler! Textmarke nicht definiert.</b> |
| Abbildung 24: Kurzanleitung zur Durchführung eines ExpertInneninterviews .....  | 59  |
| Abbildung 25: Erster Schritt zur Kontoerstellung.....   | 73  |
| Abbildung 26: Schritt zwei im Erstellungsprozess.....   | 74  |
| Abbildung 27: Angabe der benötigten Dateien .....   | 74  |
| Abbildung 28: Generierter Tracking Code von Google .....  | 75  |
| Abbildung 29: Google Konsole mit Einführungsvideos .....  | 76  |

|  |    |
|--|----|
| Abbildung 30: Software Voraussetzungen für die Installation (Quelle: <a href="https://piwik.org/wp-content/uploads/2008/11/2a-check-success1-700x332.png">https://piwik.org/wp-content/uploads/2008/11/2a-check-success1-700x332.png</a> , 4.6.2017).....  | 80 |
| Abbildung 31: Erstellung des Super Users (Quelle: <a href="https://piwik.org/wp-content/uploads/2008/11/3-database1.png">https://piwik.org/wp-content/uploads/2008/11/3-database1.png</a> , 4.6.2017).....   | 81 |
| Abbildung 32: Hinzufügen der Webseite (Quelle: <a href="https://piwik.org/wp-content/uploads/2008/11/6-setup-website1.png">https://piwik.org/wp-content/uploads/2008/11/6-setup-website1.png</a> , 4.6.2017) .....   | 81 |
| Abbildung 33: Darstellung des Tracking Codes (Quelle: <a href="https://piwik.org/wp-content/uploads/2008/11/7-javascript-tag1-700x378.png">https://piwik.org/wp-content/uploads/2008/11/7-javascript-tag1-700x378.png</a> , 4.6.2017).....   | 82 |
| Abbildung 34: Dashboard von Piwik (Quelle: <a href="http://demo.piwik.org/index.php?module=CoreHome&amp;action=index&amp;idSite=3&amp;period=day&amp;date=yesterday#?idSite=3&amp;period=day&amp;date=yesterday&amp;category=Dashboard_Dashboard&amp;subcategory=1">http://demo.piwik.org/index.php?module=CoreHome&amp;action=index&amp;idSite=3&amp;period=day&amp;date=yesterday#?idSite=3&amp;period=day&amp;date=yesterday&amp;category=Dashboard_Dashboard&amp;subcategory=1</a> , 4.6.2017) ..... | 82 |
| Abbildung 35: Anonymisierung der IP-Adresse (Quelle: <a href="https://piwik.org/wp-content/uploads/2013/11/privacy.png">https://piwik.org/wp-content/uploads/2013/11/privacy.png</a> , 1.6.2017) .....   | 83 |

## TABELLENVERZEICHNIS

|  |    |
|--|----|
| Tabelle 1: Vergleich transaktionaler Systeme und Big Data (Bundesverband Informationswirtschaft, 2012)       | 38 |
| Tabelle 2 Vergleich analytischer Systeme und Big Data (Bundesverband Informationswirtschaft, 2012)           | 39 |
| Tabelle 3: Vergleich Dokumenten-Management-Systeme und Big Data (Bundesverband Informationswirtschaft, 2012) | 40 |
| Tabelle 4: Bausteine von Big Data (Bundesverband Informationswirtschaft, 2012)                               | 47 |
| Tabelle 5: Aufgaben von Map und Reduce (WAGmob, 2013)  | 49 |
| Tabelle 6: Gegenüberstellung der Grenzen von Unternehmensbezeichnungen (Wirtschaftskammer Österreich, 2016)  | 53 |
| Tabelle 7: Gegenüberstellung journalistisches Interview mit ExpertInneninterview (Mieg & Brunner, 2001)      | 57 |
| Tabelle 8: Informationen zu ExpertInnen die am Interview teilnahmen  | 64 |

## LISTINGVERZEICHNIS

|   |    |
|---|----|
| Listing 1: Inhalt eines Server-Logfiles .....   | 17 |
| Listing 2: Datum, Uhrzeit und Zeitzone in einem Logfile .....   | 20 |
| Listing 3: Aufgerufene URL einer Website .....  | 21 |
| Listing 4: IP-Adresse in einem Logfile .....  | 21 |
| Listing 5: UserInnen-Agent in einem Logfile.....  | 22 |
| Listing 6: Anonymisierung der IP-Adresse in Google-Analytics (Datenschutzbeauftragter Info, 2017).....                        | 77 |
| Listing 7: Einfügen der Opt-Out-Möglichkeit.....  | 78 |
| Listing 8: Code um Opt-Out-Möglichkeit in Datenschutzbestimmungen zu integrieren<br>(Datenschutzbeauftragter Info, 2017)..... | 84 |

## LITERATURVERZEICHNIS

- Abts, D., & Mülder, W. (2010). *Masterkurs Wirtschaftsinformatik*. Wiesbaden: Vieweg+Teubner.
- Aden, T. (2013). *Google Analytics*. München: Hanser Verlag.
- Bange, C., Grosser, T., & Janoschek, N. (18. Juni 2015). *sas.com*. Von [http://www.sas.com/content/dam/SAS/bp\\_de/doc/studie/ba-st-barc-bigdata-use-cases-de-2359583.pdf](http://www.sas.com/content/dam/SAS/bp_de/doc/studie/ba-st-barc-bigdata-use-cases-de-2359583.pdf) abgerufen am 15.9.2017
- Bitkom. (20. April 2016). Von <https://www.bitkom.org/bitkom/Publikationen> abgerufen am 20.9.2017
- Bode, J. (o.T.. o.M. 1997). Der Informationsbegriff in der Betriebswirtschaftslehre. *zfbf - Zeitschrift für betriebswirtschaftliche Forschung*, S. 449-448.
- Bundesverband Informationswirtschaft. (o.T.. o.M. 2012). *www.bitkom.org*. Von <https://www.bitkom.org/Publikationen/2012/Leitfaden/Leitfaden-Big-Data-im-Praxiseinsatz-Szenarien-Beispiele-Effekte/BITKOM-LF-big-data-2012-online1.pdf> abgerufen am 15.10.2017
- Bundesverband Informationswirtschaft. (2012). *www.bitkom.org*. Von <https://www.bitkom.org/Publikationen/2012/Leitfaden/Leitfaden-Big-Data-im-Praxiseinsatz-Szenarien-Beispiele-Effekte/BITKOM-LF-big-data-2012-online1.pdf> abgerufen am 16.8.2017
- BVDW e.V., B. D. (16. September 2012). Customer Journey - Definition und Ausprägung. Düsseldorf, Nordrhein Westfalen, Deutschland. Abgerufen am 15. 9 2016
- Chaffey, D., Mayer, R., Johnston, K., & Ellis-Chadwick, F. (2000). *Internet Marketing*. München: Financial Times.
- Datenschutzbeauftragter Info. (18. Mai 2017). <https://www.datenschutzbeauftragter-info.de>. Von <https://www.datenschutzbeauftragter-info.de/fachbeitraege/google-analytics-datenschutzkonform-einsetzen/> abgerufen am 15.2.2017
- Davenport, T., Barth, P., & Bean, R. (30. Juli 2012). *How Big Data Is Different*. Von MIT Sloan: <http://sloanreview.mit.edu/article/how-big-data-is-different/> abgerufen am 16.2.2017
- Digital Analytics Association. (o.T.. o.M. 2016). <http://www.digitalanalyticsassociation.org/>. Von <http://www.digitalanalyticsassociation.org/> abgerufen am 19.1.2017
- eMarketer Inc. (22. August 2016). *emarketer.com*. Von <https://www.emarketer.com/Article/Worldwide-Retail-Ecommerce-Sales-Will-Reach-1915-Trillion-This-Year/1014369> abgerufen 15.9.2017

- Fraunhofer IAIS. (o.T.. o.M. o.J.). *fraunhofer.de*. Von [https://www.fraunhofer.de/content/dam/zv/de/forschungsthemen/kommunikation/bigdata/Innovationspotenzialanalyse\\_Big-Data\\_Fraunhofer-IAIS.pdf](https://www.fraunhofer.de/content/dam/zv/de/forschungsthemen/kommunikation/bigdata/Innovationspotenzialanalyse_Big-Data_Fraunhofer-IAIS.pdf) abgerufen am 3.2.2017
- Gartner Inc. (Juni 2016). *Gartner*. (Gartner, Herausgeber) Abgerufen am 29. Juni 2016 von <http://www.gartner.com/it-glossary/big-data/>
- Gittenberger, E., & Vogl, B. (Mai 2014). *WKO.at*. Von [https://www.wko.at/Content.Node/branchen/oe/Download\\_Internet-Einzelhandel\\_2014.pdf](https://www.wko.at/Content.Node/branchen/oe/Download_Internet-Einzelhandel_2014.pdf) abgerufen am 4.2.2017
- Gläser, J., & Laudel. (2010). *Experteninterviews und Qualitative Inhaltsanalyse, 4. Aufl.* Wiesbaden: Springer.
- Graham, B. (31. März 2007). *Daily Mail.co.uk*. Abgerufen am 2. Juli 2016 von <http://www.dailymail.co.uk/news/article-445897/George-Orwell-Big-Brother-watching-house.html>
- Hassler, M. (30. April 2007). *computerwoche.de*. Von <http://www.computerwoche.de/a/das-einmaleins-der-web-analytics,592002,3> abgerufen am 5.9.2016
- Hassler, M. (2010). *Web Analytics - Metriken auswerten, Besucherverhalten verstehen, Website optimieren*. Heidelberg: Hüthig Jele Gehm GmbH.
- Hassler, M. (2017). *Digital und Web Analytics*. Frechen: mitp Verlags gmbH & Co. KG.
- Hilbert, M. (2012). *MartinHilbert.net*. Abgerufen am 2. Juli 2016 von [http://www.martinhilbert.net/wp-content/uploads/2015/01/Hilbert\\_Significance\\_pre-publish.pdf](http://www.martinhilbert.net/wp-content/uploads/2015/01/Hilbert_Significance_pre-publish.pdf)
- IBM. (17. März 2015). *ibm.com*. Von <http://www-03.ibm.com/press/us/en/pressrelease/46330.wss> abgerufen am 9.9.2016
- ISO/IEC IS 13250-2:2006. (2006). *Information Technology - Document Description and Processing Languages - Topic Maps - Data Model*. (International Organization for Standardization, Hrsg.) Abgerufen am 04. 05 2012 von <http://www.isotopicmaps.org/sam/sam-model/>
- John Deere GmbH & Co.KG. (9. Oktober 2016). *MyJohnDeere.com*. Von [https://myjohndeere.deere.com/wps/portal/myjd/login?TYPE=33554433&REALMOID=06-d7dceccf-e630-4221-a771-7f13b8c32cbf&GUID=&SMAUTHREASON=0&METHOD=GET&SMAGENTNAME=\\$SM\\$MIHrPcDzl64EhJyTsz0bZg9O4ozvU07YATHUkg8KhA1xe%2fzNofYyZ2ICpoBgh4jL&TARGET=\\$SM\\$https%3a%2f%2f](https://myjohndeere.deere.com/wps/portal/myjd/login?TYPE=33554433&REALMOID=06-d7dceccf-e630-4221-a771-7f13b8c32cbf&GUID=&SMAUTHREASON=0&METHOD=GET&SMAGENTNAME=$SM$MIHrPcDzl64EhJyTsz0bZg9O4ozvU07YATHUkg8KhA1xe%2fzNofYyZ2ICpoBgh4jL&TARGET=$SM$https%3a%2f%2f) abgerufen am 14.9.2016
- Kohn, F., Kutzmutz, O., & Larisch, P. (Hrsg.). (2010). *Destillate - Literatur Labor Wolfenbüttel 2010*. Wolfenbüttel: Bundesakademie für kulturelle Bildung.

- Kollmann, T. (2013). *E-Business - Grundlagen elektronischer Geschäftsprozesse in der Net Economy*. Essen: Springer Gabler.
- Kurowski, O. (2012). *NoSQL Einführung - CouchDB, MongoDB und Redis*. o.O.: entwickler.press.
- Liikannen, E. (6. Mai 2003). Empfehlung der Kommission betreffend die Definition der Kleinstunternehmen sowie der kleinen und mittleren Unternehmen. Brüssel, Belgien.
- Marr, B. (27. April 2015). *Big Data-As-A-Service Is Next Big Thing*. Von Forbes: <http://www.forbes.com/sites/bernardmarr/2015/04/27/big-data-as-a-service-is-next-big-thing/> abgerufen am 18.10.2016
- Mayer-Schönberger, V., & Cukier, K. (2013). Big Data - Die Revolution, die unser Leben verändern wird. In V. Mayer-Schönberger, & K. Cukier, *Big Data - Die Revolution, die unser Leben verändern wird* (S. 13). München: Redline Verlag.
- Meier, A., & Stormer, H. (2012). *eBusiness & eCommerce*. Berlin Heidelberg: Springer.
- Meine AG. (2012). *Projektdokumentation*. Abteilung IT. Graz: Mein Unternehmen.
- Meuser, M., & Nagel, U. (2009). Das Experteninterview – konzeptionelle Grundlagen und methodische Anlage. In S. Pickel, G. Pickel, H.-J. Lauth, & D. Jahn, *Methoden der vergleichenden Politik- und Sozialwissenschaft. Neue Entwicklungen und* (S. 465-480). Wiesbaden.
- Meuser, M., & Nagel, U. (1991). *ExpertInneninterviews - vielfach erprobt, wenig bedacht : ein Beitrag zur qualitativen Methodendiskussion*. Garz: Westdeutscher Verlag.
- Mieg, H. A., & Brunner, B. (2001). <http://www.ozonfallstudie.ethz.ch>. Von <http://www.ozonfallstudie.ethz.ch/experteninterview.pdf> abgerufen am 15.3.2017
- Mülder, W. (2002). Elektronische Marktplätze. In T. Jaspersen, R. Conrady, & W. Pepels, *Online-Marketing-Strategien: Konzeption, Technologien, Prozesse, Recht*. Neuwied.
- Page, L. (Dezember 2012). *Alphabet*. Abgerufen am 29. Juni 2016 von <https://abc.xyz/investor/founders-letters/2012/index.html>
- PricewaterhouseCoopers AG. (o.T.. o.M. 2013). *pwc.de*. Von <http://www.pwc.de/de/prozessoptimierung/big-data-so-profitieren-unternehmen-von-der-datenvielfalt.html> abgerufen am 3.9.2016
- Quinn, E. (13. April 2014). *Social Science Research Network*. Abgerufen am 2. Juli 2016 von [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1462285](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1462285)
- Qureshi, R. (Juni 2016). *Ericsson Mobility Report*. (Ericsson, Herausgeber) Abgerufen am 29. Juni 2016 von <https://www.ericsson.com/mobility-report>

- RFC 2616. (1999). *Hypertext Transfer Protocol -- HTTP/1.1*. (IETF, Hrsg.) Von <http://www.ietf.org/rfc/rfc2616.txt> abgerufen am 15.10.2016
- Schreiber, D. V., & Leser, A. (Hrsg.). (2008). *Wichtiges Werk*. Graz: Wissensverlag.
- speicherguide.de GbR. (4. April 2012). *www.speicherguide.de*. Von <http://www.speicherguide.de/management/big-data/hadoop-was-ist-es-und-was-leistet-es-15378.aspx#> abgerufen am 12.10.2016
- T3N Digital Pioneers. (8. März 2014). *T3N*. Von <http://t3n.de/news/okr-google-wunderwaffe-valley-ziele-530092/> abgerufen am 3.3.2017
- Team, T. (20. April 2016). *Forbes*. Von <http://www.forbes.com/sites/greatspeculations/2015/10/09/a-comparative-look-at-the-valuation-of-amazon-alibaba-and-ebay/#3adf7d27747e> abgerufen am 20.4.2017
- Tilmann, A. (14. Jänner 2015). *Spektrum.de*. Abgerufen am 29. Juni 2016 von <http://www.spektrum.de/news/lst-die-groesste-digitalkamera-der-welt-kann-gebaut-werden/1327283>
- Vagata, P., & Wilfong, K. (10. April 2014). *www.code.facebook.com*. Von <https://code.facebook.com/posts/229861827208629/scaling-the-facebook-data-warehouse-to-300-pb/> abgerufen am 15.9.2016
- W3 Techs. (8. Juni 2017). *W3 Techs - Web Technology Surveys*. Von [https://w3techs.com/technologies/segmentation/tld-de-/traffic\\_analysis](https://w3techs.com/technologies/segmentation/tld-de-/traffic_analysis) abgerufen am 16.9.2016
- WAGmob. (2013). *Big Data and Hadoop*. Redmond: WAGmob.
- Wartburg, R. v., Steinbacher, S., Wittmer, R., & Schütze, S. (2011). *Zitieren und Referenzieren nach APA*. Abgerufen am 20. 4 2012 von Wissenschaftliches Schreiben in der Psychologie: <http://etools.fernuni.ch/wiss-schreiben/apa/de/html/index.html>
- Wingfield, N. (9. September 2011). *wsj.com*. Abgerufen am 2. Juli 2016 von <http://www.wsj.com/articles/SB10001424053111904823804576502442835413446>
- Wirtschaftskammer Österreich . (10. Februar 2017). *www.wko.at*. Von [https://www.wko.at/Content.Node/branchen/oe/sparte\\_iuc/Werbung-und-Marktkommunikation/Zustimmungserklaerung\\_fuer\\_Speicherung\\_von\\_Kundendaten\\_bei.html](https://www.wko.at/Content.Node/branchen/oe/sparte_iuc/Werbung-und-Marktkommunikation/Zustimmungserklaerung_fuer_Speicherung_von_Kundendaten_bei.html) abgerufen am 10.2.2017
- Wirtschaftskammer Österreich. (12. September 2016). *WKO*. Von [https://www.wko.at/Content.Node/Interessenvertretung/ZahlenDatenFakten/KMU\\_Definition.html](https://www.wko.at/Content.Node/Interessenvertretung/ZahlenDatenFakten/KMU_Definition.html) abgerufen am 12.9.2016

*Wirtschaftslexikon Gabler*. (20. April 2016). Von <http://wirtschaftslexikon.gabler.de/Definition/big-data.html> abgerufen am 20.4.2016

Wirtz, B. W. (2001). *Electronic Business, 2. Auflage*. Wiesbaden: Gabler.

Yarow, J. (6. Jänner 2014). *Business Insider*. Von <http://www.businessinsider.com/googles-ranking-system-okr-2014-1?IR=T> abgerufen am 20.3.2017

Zadrozny, P., & Kodali, R. (2013). *Big Data Analytics Using Splunk: Deriving Operational Intelligence from Social Media, Machine Data, Existing Data Warehouses, and Other Real-Time Streaming Sources*. Springer Science + Business Media: New York.