



Lehrstuhl für Wirtschafts- und Betriebswissenschaften

Masterarbeit



Datenmanagement in österreichischen  
Industrieunternehmen - Umsetzung,  
Trends und Hindernisse

Martin Wallner, BSc

Oktober 2019



## Aufgabenstellung

Herrn Martin Wallner wird das Thema

### **Datenmanagement in österreichischen Industrieunternehmen – Umsetzung, Trends und Hindernisse**

zur Bearbeitung in einer Masterarbeit gestellt.

Der theoretische Teil der Arbeit soll die Thematik des Datenmanagements behandeln. Die Grundlage bilden Daten im Allgemeinen und das Datenmanagement und dessen Abgrenzung zum Informations- und Wissensmanagement im Speziellen. Das Datenmanagement soll umfänglich beschrieben werden und muss daher über die rein technische Sicht hinausgehen und die organisatorische Sicht einbeziehen. Zur vollständigen Betrachtung der Thematik zählen die aktuellen Trends die sich aus Industrie 4.0 ergeben.

Der zweite Teil der Arbeit umfasst die Ausarbeitung eines Fragebogens, womit die im theoretischen Teil erarbeiteten Inhalte erhoben werden. Die Zielgruppe bilden österreichische Industrieunternehmen, deren Entwicklung im Datenmanagement festgestellt werden soll. Die Ergebnisse sind vor dem Hintergrund, wie Größe, Umsatz und Industriesektor der Unternehmen, zu interpretieren und mit anderen Studien zu vergleichen.

A handwritten signature in blue ink, which appears to read 'Hubert Biedermann', is positioned above the printed name.

Leoben, September 2018

o.Univ.Prof. Dr. Hubert Biedermann

**EIDESSTÄTLICHE ERKLÄRUNG**

Ich erkläre an Eides statt, dass ich diese Arbeit selbständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt, und mich auch sonst keiner unerlaubten Hilfsmittel bedient habe.

Ich erkläre, dass ich die Richtlinien des Senats der Montanuniversität Leoben zu "Gute wissenschaftliche Praxis" gelesen, verstanden und befolgt habe.

Weiters erkläre ich, dass die elektronische und gedruckte Version der eingereichten wissenschaftlichen Abschlussarbeit formal und inhaltlich identisch sind.

Datum 30.10.2019



Unterschrift Verfasser/in  
Martin, Wallner

## **Gleichheitsgrundsatz**

Aus Gründen der Lesbarkeit wurde in dieser Arbeit darauf verzichtet, geschlechtsspezifische Formulierungen zu verwenden. Es wird ausdrücklich festgehalten, dass die bei Personen verwendeten maskulinen Formen für beide Geschlechter zu verstehen sind.

## Danksagung

Mein erster Dank gilt meiner Frau Barbara Wallner, die mir durch ihre Unterstützung durch die schwierigen Zeiten, die während eines Studiums auftreten, geholfen hat und die sich um unsere gemeinsame Tochter Charlotte gekümmert hat, damit ich Zeit hatte diese Arbeit zu verfassen. Meiner Tochter Charlotte danke ich für die tägliche Inspiration.

Ein weiterer Dank gilt meinen Eltern Anetta und Witold, die mir diese Ausbildung überhaupt erst ermöglicht haben. Im gleichen Zug danke ich auch meinen Schwiegereltern Anna und Stefan welche mich in ihre Familie aufgenommen haben und immer ein offenes Ohr für mich hatten.

Herrn Dr. Robert Bernerstätter danke ich für sein tolles Engagement bei der Betreuung dieser Arbeit und für die zahlreichen wertvollen Tipps.

Auch Herrn Prof. Dr. Hubert Biedermann als Institutsvorstand des Lehrstuhls Wirtschafts- und Betriebswissenschaft an der Montanuniversität Leoben möchte ich an dieser Stelle danken.

## Kurzfassung

In der heutigen Wirtschaftswelt sind Daten eine der wichtigsten Ressourcen die Unternehmen zur Verfügung stehen. Zunächst wurden in dieser Masterarbeit die wichtigsten Konzepte in Bezug auf Daten und das Datenmanagement in der Theorie erklärt. Außerdem wird auf die derzeit verfügbaren Datenanalysemethoden eingegangen und jene statistischen Instrumente aufgezeigt, auf welchen diese beruhen.

Im zweiten Teil der Arbeit wurde untersucht ob in österreichischen Industrieunternehmen auch die notwendigen Schritte eingeleitet wurden, um einen effektiven und effizienten Umgang mit Daten zu bewerkstelligen. Dazu wurde ein Fragebogen an österreichische Industrieunternehmen geschickt, welcher 141 beantwortet wurde. Nach der Auswertung der Ergebnisse konnten u.a. folgende Trends aufgezeigt werden:

- Wenige Industrieunternehmen in Österreich haben eigens definierte Stellen für den Umgang mit Daten definiert.
- Ein Großteil der Unternehmen hat zumindest in einigen Fachabteilungen ein Datenmanagement implementiert.
- In den meisten Industrieunternehmen erfolgt bereits eine Standardisierung der Auswerteperspektiven und KPIs.
- Big Data Projekte welche komplexere Datenanalyseverfahren erlauben, werden von weniger als 50% der befragten Unternehmen durchgeführt. Auch in Zukunft wird angegeben, dass dies nur ein Kernthema unter vielen IT-Schwerpunkten ist.
- Die Unternehmen erkennen, dass sie in vielen Bereichen des Datenmanagements Aufholbedarf haben.
- Es besteht ein klarer Zusammenhang zwischen dem Umsatz eines Unternehmens und der Implementierung eines unternehmensweiten Datenmanagements.
- Trotz eines eigens definierten Datenverantwortlichen stellt Kommunikation weiterhin eines der Hauptprobleme im Umgang mit Daten dar.
- Wenige Industrieunternehmen in Österreich setzen auf moderne Datenarchitekturen oder fortgeschrittene Datenanalyseverfahren.
- Österreichische Industrieunternehmen sind sich bewusst, dass in den nächsten Jahren zukunftsbezogene Analysemethoden vorherrschend sein werden.

Die Masterarbeit gibt somit einen guten Überblick über die aktuelle Lage des Datenmanagements in Österreich und zeigt auf bei welchen Punkten die österreichische Industrie gut aufgestellt ist und wo noch Aufholbedarf herrscht. Diese Ergebnisse werden in weiterer Folge auch mit anderen Studien verglichen, um Unterschiede zu verschiedenen geographischen Gebieten aufzeigen zu können.

## Abstract

In today's economic world, data are one of the most important resources that companies have. In the first chapters of this master thesis the essential concepts relating to data and data management are explained. In addition, the currently used data analysis methods are described, and the statistical instruments are shown on which these methods are based.

In the second part of this thesis it has been evaluated if the Austrian industrial companies have made the necessary steps to ensure an effective and efficient handling of data. For this purpose, a survey has been sent to Austrian industrial companies. After the evaluation of the 141 replies, the following trends can be outlined:

- Only a few of the industrial companies in Austria have specifically defined positions for data management.
- A majority of the companies have at least in a few departments an implemented data management.
- In most industrial companies a standardization of the evaluation perspective and of KPI's is done.
- Big data projects, which allow more complex data analysis, are conducted by less than 50% of the respondents. It's also stated by most companies that this topic is not considered as a key issue in the future.
- The companies realize that they have to catch up on a few data related topics.
- There is a clear correlation between the turnover of a company and the implementation of a company-wide data management.
- Even if there is a defined data controller in the company, communication is stated as one of the main problems in the handling of data.
- Very few industrial companies use modern data architectures or advanced data analysis.
- Austrian industrial companies are aware that forward-looking data analysis will be predominant in the future.

This master thesis provides a good overview of the current state of the data management of Austrian industrial companies and shows where they are well-positioned and where they need to catch up. The results of the conducted survey are also compared to other studies on this subject to point out territorial differences.

---

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b> .....	<b>1</b>
1.1	Ausgangssituation und Problemstellung .....	2
1.2	Zielsetzung .....	3
1.3	Methodische Vorgehensweise .....	4
1.4	Aufbau der Arbeit .....	5
<b>2</b>	<b>Daten</b> .....	<b>7</b>
2.1	Arten von Daten .....	9
2.2	Lebenszyklus von Daten .....	10
2.2.1	Beschaffung .....	10
2.2.2	Speicherung .....	11
2.2.3	Nutzung und Verwaltung .....	13
2.2.4	Entsorgung der Daten .....	14
<b>3</b>	<b>Einführung in das Datenmanagement</b> .....	<b>15</b>
3.1	Implementierung eines erfolgreichen Datenmanagements .....	15
3.2	Mission des Datenmanagements .....	17
3.3	Involvierte Personen .....	18
3.4	Abgrenzungen zum Informations- und Wissensmanagement .....	19
3.5	Datenqualität .....	21
3.6	Informations- und Kommunikationssysteme .....	24
3.7	Data Mining .....	26
<b>4</b>	<b>Datenmanagement in Zeiten von Industrie 4.0</b> .....	<b>29</b>
4.1	Big Data .....	29
4.2	Datenanalyseverfahren .....	31
4.2.1	Deskriptive Analyse .....	31
4.2.2	Diagnostische Analyse .....	35
4.2.3	Prädiktive Analyse .....	39
4.2.4	Präskriptive Analyse .....	42
4.3	Datenarchitekturen .....	45
4.4	Aktuelle Trends .....	46
4.4.1	Internet of Things .....	47
4.4.2	Machine Learning .....	48
4.5	Probleme und Herausforderungen .....	50

<b>5</b>	<b>Fragebogenerhebung</b> .....	<b>52</b>
5.1	Anlass und Ziele der Fragebogenerhebung .....	52
5.2	Konzept des Fragebogens .....	53
5.3	Fragen und Antwortvorgaben.....	55
5.4	Durchführung der Fragebogenerhebung .....	59
<b>6</b>	<b>Auswertung der Fragebogenerhebung</b> .....	<b>62</b>
6.1	Einleitende Fragestellungen.....	63
6.2	Fragen zur Organisation .....	68
6.3	Technische Fragestellungen .....	75
6.4	Fragen zu Big Data und Datenanalysemethoden .....	80
6.5	Diskussion zur Fragebogenerhebung.....	88
6.6	Vergleich mit anderen Studien .....	104
<b>7</b>	<b>Zusammenfassung und Ausblick</b> .....	<b>110</b>

## Abbildungsverzeichnis

Abbildung 1: Geschätztes weltweites Datenvolumen bis 2020 .....	2
Abbildung 2: Begriffshierarchie.....	7
Abbildung 3: Bestandteile von Daten als Ressource .....	10
Abbildung 4: Dateispeicherung.....	12
Abbildung 5: Block-Speicherung.....	12
Abbildung 6: Objekt-Speicherung.....	13
Abbildung 7: Total Information Quality Management Modell nach Larry English.....	23
Abbildung 8: Total Information Quality Management Modell nach Danette McGilvray .....	24
Abbildung 9: Systempyramide .....	25
Abbildung 10: Notwendige Prozessschritte für einen KDD-Prozess .....	27
Abbildung 11: Klassifikation der analytischen Methoden Deskriptive Analyse .....	31
Abbildung 12: Kontingenztafel.....	32
Abbildung 13: Histogramm .....	33
Abbildung 14: Box-Plot Diagramm.....	34
Abbildung 15: Qualitätsregelkarte.....	39
Abbildung 16: Entscheidungsbaum - Klassifikationsbaum (rechts) und Klassifikation der Daten in bestimmte Klassen (links).....	41
Abbildung 17: IoT-Kommunikationsmöglichkeiten .....	47
Abbildung 18: Unterbereiche Künstlicher Intelligenz.....	49
Abbildung 19: Aufbau der Fragebogenerhebung .....	55
Abbildung 20: Frage 1: Sektor des Unternehmens .....	63
Abbildung 21: Frage 2: Mitarbeiter des Unternehmens.....	64
Abbildung 22: Frage 3: Jahresumsatz des Unternehmens .....	65
Abbildung 23: Frage 4: Umsatzmäßiger Vergleich.....	66
Abbildung 24: Frage 5: Unternehmensbereich des Teilnehmers .....	67
Abbildung 25: Frage 6: Verantwortung .....	68
Abbildung 26: Frage 7: Implementierung.....	69
Abbildung 27: Frage 8: Implementierung Datenmanagement Vergleich .....	70
Abbildung 28: Frage 9: Mehrwert .....	71
Abbildung 29: Frage 10: Probleme .....	72
Abbildung 30: Frage 11: Operative Entscheidungen.....	73
Abbildung 31: Frage 12: Strategische Entscheidungen .....	74
Abbildung 32: Frage 13: Standardisierung .....	75
Abbildung 33: Frage 14: Technischer Schwerpunkt.....	76

---

Abbildung 34: Frage 15: Aufholbedarfsfaktoren.....	77
Abbildung 35: Frage 16: Dimensionen der Datenqualität.....	78
Abbildung 36: Frage 17: Speicherung .....	79
Abbildung 37: Frage 18: Erfahrung mit Big Data .....	80
Abbildung 38: Frage 19: Potential für Big Data Projekte.....	81
Abbildung 39: Frage 20: Verbesserungen durch Data Analytics.....	82
Abbildung 40: Frage 21: Herausforderungen auf geschäftlicher Seite .....	83
Abbildung 41: Frage 22: Herausforderungen auf technischer Seite.....	84
Abbildung 42: Frage 23: Derzeitige Analyseformen.....	85
Abbildung 43: Frage 24: Zukünftige Analysemethoden .....	86
Abbildung 44: Frage 25: Datenquellen .....	87
Abbildung 45: Umsatzstruktur von Unternehmen mit implementierten Datenmanagement .....	90
Abbildung 46: Datenmanagementstruktur von Unternehmen welche komplexere Datenanalysen betreiben .....	92
Abbildung 47: Datenarchitektur von Unternehmen welche die Vorteile von Datenmanagement erkannt haben.....	94
Abbildung 48: Menschliche Fehler als Grund für Probleme in Bezug auf Datenmanagement .....	96
Abbildung 49: Umsatzstruktur von Unternehmen welche operative Entscheidung auf Grundlage von Daten treffen.....	97
Abbildung 50: Umsatzstruktur von Unternehmen welche strategische Entscheidung auf Grundlage von Daten treffen .....	98
Abbildung 51: Zusammenhang Standardisierung Auswerteperspektiven und Durchführung von Big Data Projekten.....	99
Abbildung 52: Unternehmensstruktur von Unternehmen welche moderne Datenarchitekturen verwenden .....	100
Abbildung 53: Umsatzstruktur von Unternehmen welche Big Data Projekte durchführen.....	102
Abbildung 54: Anteil an Unternehmen welche moderne Analysemethoden verwenden .....	103
Abbildung 55: Anteil an Unternehmen welche ein unternehmensweites Datenmanagement definiert haben .....	106
Abbildung 56: Anteil an Unternehmen welche moderne Datenarchitekturen verwenden .....	108

## **Tabellenverzeichnis**

Tabelle 1: Dimensionen der Informationsqualität .....	22
Tabelle 2: Anzahl der Beantwortungen und Nicht-Beantwortungen pro Frage.....	60

## Abkürzungsverzeichnis

CAPEX	Capital expenditures
DACH	Deutschland, Österreich und Schweiz
DAMA	Data Management Association
EPK	Ereignisgesteuerte Prozesskette
ERM	Entity Relationship Modell
ETL	Extraktion, Transformation und Laden
IOT	Internet of Things
ISO	International Organization for Standardization
IT	Informationstechnologie
IuK	Informations- und Kommunikationssysteme
KDD	Knowledge Discovery in Databases
KI	Künstliche Intelligenz
KPI	Key Performance Indicators
KVP	Kontinuierlicher Verbesserungsprozess
NLP	Natural Language Processing
OLAP	On-Line Analytical Processing
OPEX	Operating expenditures
SQL	Structured Query Language
WKO	Wirtschaftskammer Österreich

# 1 Einleitung

„Wissen ist Macht“ dieses Zitat vom englischen Politiker und Philosophen Francis Bacon ist bereits über 400 Jahre alt, trotzdem ist es wohl heute wo die Digitalisierung alle Bereiche unseres Lebens durchdrungen hat, so aktuell wie nie zuvor.

Daten sind längst zu einer der wertvollsten Ressourcen unserer Zeit geworden und die Innovationen in diesen Bereichen führen zu noch nie dagewesenen Datenmengen. Durch die Verwendung von Messenger-Diensten, dem Versenden von Mails, Dokumenten oder Fotos und der Verwendung von unzähligen Apps auf Smartphones oder Smartwatches erzeugt jeder von uns mittlerweile täglich eine schier unglaubliche Menge an Daten. 2018 wurden so weltweit 2,5 Trillionen Byte<sup>1</sup> pro Tag erzeugt, das beachtliche hierbei ist nicht nur dieser gigantische Wert, sondern auch, dass 90% aller heute weltweit existierenden Daten in den letzten zwei Jahren erzeugt worden sind<sup>2</sup>. Verantwortlich dafür sind neben der ständig wachsenden Anzahl an Datenquellen, welche mit einer stetig steigenden Frequenz Daten senden auch eine Erhöhung der Rechenleistung, sowie eine erhebliche Senkung der Kosten für die Erzeugung, Speicherung und Verwaltung von Daten. Gleichzeitig hat es in den letzten Jahren erhebliche Investitionen der großen IT-Konzerne in Technologien (z.B. Cloud-Dienste) gegeben, welche die Speicherung und Verwaltung von Daten für Anwender noch weiter erleichtern. Auch Unternehmen haben es sich zum Ziel gemacht möglichst viele Daten über ihre Kunden, Maschinen oder Konkurrenten zu sammeln, um wettbewerbsfähig zu bleiben.

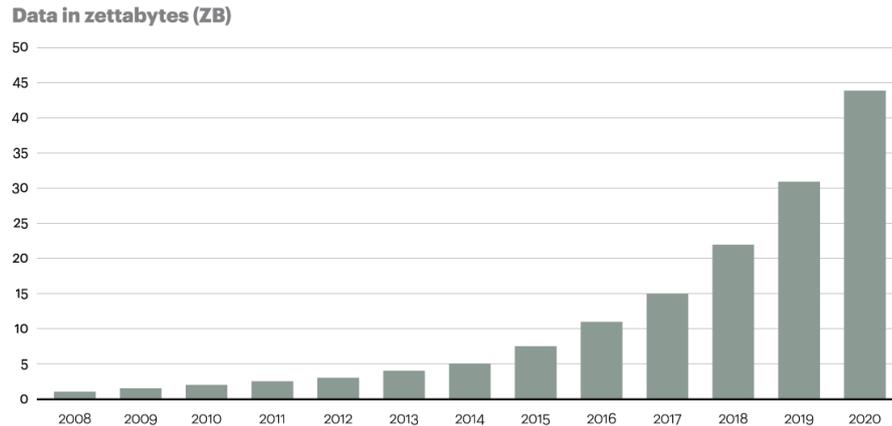
Ein Ende des stetigen Datenanstieges ist nicht in Sicht, laut Ansicht vieler Experten wie z.B. Oracle<sup>3</sup>, einem der weltweit führenden Hersteller für Datenbanksysteme, wird dieser Trend in den nächsten Jahren durch das Internet of Things (IoT) sogar noch deutlich verstärkt. Bereits in den nächsten zwei Jahren ist durch die Vernetzung von physischen und virtuellen Gegenständen so mit einer 40% jährlichen Wachstumsrate des weltweiten Datenvolumens zu rechnen (siehe Abbildung 1).

---

<sup>1</sup> Forbes, <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#5654930d60ba> (Zugriff: 30.10.2019).

<sup>2</sup> IBM, <https://www.ibm.com/blogs/insights-on-business/consumer-products/2-5-quintillion-bytes-of-data-created-every-day-how-does-cpg-retail-manage-it/> (Zugriff: 16.11.2018).

<sup>3</sup> Oracle, <https://blogs.oracle.com/datamining/mining-structured-data-and-unstructured-data-using-oracle-advanced-analytics-12c> (Zugriff: 11.12.2018).



**Abbildung 1: Geschätztes weltweites Datenvolumen bis 2020<sup>4</sup>**

Aufgrund des weltweit stark gestiegenen Datenvolumens bekommen alle davon betroffenen Bereiche in Unternehmen einen immer höheren Stellenwert. Dieser Trend ist auch innerhalb der österreichischen Industrie bemerkbar. Dabei ist das reine Sammeln von Daten, aber längst nicht mehr das primäre Ziel. Vielmehr wird in den letzten Jahren verstärkt darüber nachgedacht, wie die vorhandenen Informationen analysiert werden könnten, um einen möglichst großen Mehrwert für das Unternehmen zu erzielen.

## 1.1 Ausgangssituation und Problemstellung

Im 21. Jahrhundert haben Daten und vor allem die Informationen, die daraus gezogen werden können, den gleichen Stellenwert an Wichtigkeit erreicht wie das Personal oder die Vermögenswerte eines Unternehmens. Aus diesem Grund sollte dem Management von Daten im Unternehmensprozess eine hohe Gewichtung zugeschrieben werden. Daten sollten als Ressourcen angesehen werden, welche merklich zum Unternehmenserfolg beitragen können und Unternehmen, welche dies verschlafen, werden wohl in der modernen Wirtschaftswelt nicht überleben können. Hierbei ist ein unternehmensweites Konzept notwendig, welches die genaue Behandlung der Daten in jedem Abschnitt des Lebenszyklus regelt. Auch die Zuständigkeiten und Verantwortlichkeiten für alle im Unternehmen vorhandenen Daten müssen niedergeschrieben sein. Nur so kann eine möglichst hohe Datenqualität sichergestellt werden und in weiterer Folge ein maximaler Nutzen aus den vorhandenen Daten gezogen werden.

Der sorgsame Umgang mit den Daten ist dabei ebenso entscheidend wie die Datenanalyse. Während vor einigen Jahren noch die Analyse von bereits passiertten Ereignissen, wie den Ausfall von Maschinen, im Bereich der Datenanalyse vorherrschend war, wird nun in weitaus komplexeren Auswertungen versucht die Zukunft vorherzusagen oder sogar zu beeinflussen. So versucht z.B. Amazon durch Algorithmen die bisherigen Käufe, aber auch Suchbegriffe zu analysieren, um vorherzusagen zu können was ein Kunde sich wünschen könnte, bevor dieser Wunsch überhaupt aktiv geäußert wurde. Industrieunternehmen nutzen Betriebs- und Zustandsdaten von Maschinen, um

<sup>4</sup> Quelle: A.T.Kearney (2013), S. 2.

Ableitungen über den Verschleiß von Maschinenteilen zu treffen und um in weiterer Folge daraus Vorhersagen treffen zu können welche Teile wann getauscht werden müssen. So lassen sich ungeplante Stehzeiten gezielt verhindern und benötigte Komponenten können rechtzeitig, oft auch schon automatisch, bestellt werden und Stehzeiten der Maschine wie auch des Personals können so gezielt vermieden werden. Mit Hilfe dieser komplexen Analysen ist es Unternehmen möglich die richtigen Schlüsse aus den Daten zu ziehen und so die maximale Wertschöpfung für das Unternehmen zu erreichen.

## 1.2 Zielsetzung

Für den Theorieteil dieser Arbeit ist die Zielsetzung, alle wichtigen Begriffe zum Thema Daten und Datenmanagement zu erklären. Des Weiteren ist der Theorieteil ein wichtiger Ausgangspunkt für die Formulierung des Fragebogens. Dabei sollen auch die derzeitige Handhabung und Zuständigkeiten von Daten geklärt werden und die Abgrenzung zum Informationsmanagement aufgezeigt werden. Außerdem sollen die Entwicklung und Probleme des Datenmanagements beschrieben werden und auf derzeitige Trends durch die Industrie 4.0 eingegangen werden. Ein wichtiger Teil der Arbeit ist auch die Erklärung und die sinnvollen Applikationen der verschiedenen Datenanalysemethoden, u.a. die Deskriptive-, Diagnostische-, Prädiktive- und Präskriptive-Auswertung.

Das Ziel des Praxisteils ist die Erforschung der aktuellen Lage des Datenmanagements in österreichischen Industrieunternehmen ausgehend von den im Theorieteil ausgearbeiteten Konzepten zu liefern. Relevante Themen zu dieser Forschungsfrage sind:

- Zuständigkeiten für Daten im Unternehmen
- Implementierung von einem klar definierten Datenmanagement
- Vorteile für das Unternehmen durch ein Datenmanagement
- Probleme im Zusammenhang mit dem Datenmanagement
- Subjektiver Vergleich mit anderen Unternehmen in der Branche
- Wo liegen derzeit die Schwerpunkte
- In welchem Bereich des Datenmanagements herrscht der größte Aufholbedarf
- Technischer Stand (Speicherung, Datenquellen)
- Potential und Herausforderungen zum Thema Big Data
- Derzeit verwendete bzw. in Zukunft geplante Datenauswertungsmethoden
- Beschreibung der Bereiche wo durch eine komplexere Datenauswertung die größten Vorteile zu erwarten sind

Um diese Forschungsfrage statistisch robust beantworten zu können wird ein Fragebogen an österreichische Industrieunternehmen verschickt, welcher mindestens 40 Rückmeldungen benötigt, um aussagekräftig zu sein. Durch eine Korrelation mit den angefragten Basisdaten wie Branche, Umsatz und Mitarbeiterzahl, sollen so Unterschiede, welche sich auf die Basisdaten beziehen aufgezeigt werden. Ebenso sollen die ausgewerteten Daten mit anderen internationalen Studien bzw. Studien aus der DACH-Region verglichen werden. Daraus sollen Unterschiede in der derzeitigen

Umsetzung, aber auch die Schwerpunkte von zukünftigen Projekten aufgezeigt werden. Zudem sollen aus der Auswertung und Interpretation der Daten Trends und Hindernisse für das Datenmanagement in Österreich erkannt und beschrieben werden. Daneben sollen auch Versäumnisse und Stärken der österreichischen Industrie aufgezeigt werden.

### 1.3 Methodische Vorgehensweise

Durch die Formulierung der Zielsetzung und der Forschungsfrage wurden daraus die Keywords definiert, mit denen eine Literaturrecherche durchgeführt werden konnte. Diese wurde primär an der Bibliothek der Montanuniversität Leoben durchgeführt, gesucht wurden hierbei Publikationen in Deutsch und Englisch.

Die Keywords, nach denen die Suche zu den einzelnen Kapiteln durchgeführt wurde, lauteten:

- Kapitel 2,3: Datenmanagement (Data Management), Wissensmanagement (Knowledge Management), Wirtschaftsinformatik (Business Informatics), Informationsmanagement (Information Management), Datenqualität (Data Quality), Qualitätsmanagement (Quality Management),
- Kapitel 4: Industrie 4.0 (Industry 4.0), Big Data, Datengewinnung (Data Mining), Datenbanken (Data Warehouse), Internet der Dinge (Internet of Things), Datenanalyseverfahren (Industrial Analytics), Maschinelles Lernen (Machine Learning), Vorausschauende Instandhaltung (Predictive Maintenance) Diagnostische Analyse (Diagnostic Analysis), Präskriptive Analyse (Prescriptive Analysis), Deskriptive Analyse (Descriptive Analysis), Diagnostische Analyse (Diagnostic Analysis)

Nach dieser Recherche wurden Bücher, Veröffentlichungen und Fachartikel in grobe Kapitel unterteilt und studiert, um einen Überblick über die IST-Situation des Datenmanagements zu bekommen. Danach wurde aufgrund der gefundenen Literatur und der definierten Zielsetzungen der Arbeit ein erstes Inhaltsverzeichnis definiert und die Arbeit in folgende Kapitel unterteilt:

- Daten
- Einführung in das Datenmanagement
- Datenmanagement in Zeiten von Industrie 4.0
- Praxisteil
  - Fragebogen
  - Analyse
  - Vergleich mit anderen Studien
  - Interpretation der Ergebnisse
  - Ableitung von Trends und Hindernissen

Bei der Beschreibung der Datenanalyseverfahren wurden die im Hintergrund angewendeten statistischen Verfahren beschrieben. Dazu wurde erneut eine

Literaturrecherche durchgeführt, um hier Fachliteratur im Bereich der Statistik zitieren zu können.

Ausgehend von den beschriebenen Themen und den Schwerpunkten, welche sich aus der Literatur und der anfangs definierten Forschungsfrage und der Zielsetzung ergeben haben, wurde ein Fragebogen erstellt welcher, um die statistische Robustheit der Befragung zu gewährleisten, von mindestens 40 Industrieunternehmen aus Österreich beantwortet werden muss. Hierbei wurde nach dem technischen Stand, Zuständigkeiten, Herausforderungen, Problemen und Potenzialen gefragt. Auf Basis der so gewonnenen Ergebnisse konnte eine Interpretation der Daten und der Vergleich mit anderen Studien (International und DACH-Region) durchgeführt werden. Daraus wurden Trends und Hindernisse für den Bereich des Datenmanagements in Österreich formuliert.

## **1.4 Aufbau der Arbeit**

Im theoretischen Teil der Arbeit werden zunächst alle notwendigen Grundbegriffe zum Thema Daten geklärt. Danach wird erklärt wie aus Zeichen, Wissen entsteht und welche Arten von Daten unterschieden werden. Hierbei werden auch die Unterscheidungsmerkmale für die Klassifikation erklärt. In weiterer Folge wird der typische Lebenszyklus von Daten beschrieben und die Anforderungen und Probleme in jedem der vorhandenen Schritte erklärt.

In Kapitel 3 der Arbeit wird die Begriffsdefinition, Mission, Entwicklung und auf eine erfolgreiche Implementierung von Datenmanagement eingegangen. Dabei widmet sich ein Unterkapitel auch den Personen die notwendig sind, um ein Datenmanagement in einem Betrieb erfolgreich aufzusetzen. In weiterer Folge erfolgt eine klare Abgrenzung zum Wissensmanagement und Informationsmanagement. Ein wichtiger Punkt ist ebenfalls die genaue Definition welche Kriterien für eine gute Datenqualität sorgen.

Untersucht werden auch zukünftige Entwicklungen des Datenmanagements durch das Voranschreiten der Digitalisierung und der Industrie 4.0. Begriffe, die in diesem Kapitel näher ausgeführt werden, sind z.B. Big Data, Informations- und Kommunikationssysteme, Data Mining, Knowledge Discovery in Databases, Internet of Things und Machine Learning. Ferner wird auch die Erläuterung über die Arten von Datenanalyseverfahren wie die Deskriptive, Diagnostische, Prädiktive und die Präskriptive Analyse eingegangen.

Im Praxisteil wird ein Fragebogen an Industrieunternehmen in Österreich verschickt, welcher den derzeitigen Stand des Datenmanagements in österreichischen Industrieunternehmen erfassen soll. Behandelte Themen hierbei sind: Verantwortlichkeiten, Probleme mit Datenmanagement, technologische Schwerpunkte, Qualität von Daten, Speicherung, Potential für Big Data, Herausforderungen, Datenquellen und Analysemethoden. Ergänzend wurde nach Standardparametern wie z.B. Branche, Umsatz und Mitarbeiteranzahl gefragt um die gewonnenen Daten so besser klassifizieren zu können und so Rückschlüsse auf gewisse Trends, die sich aus diesen Parametern ergeben erkennen zu können. Durch Auswertung dieser Daten und den Vergleich mit anderen Studien (International und DACH-Region) soll der aktuelle Stand des Datenmanagements in Österreich dargestellt werden und Abweichungen zu

anderen Ländern und Bereiche in denen Aufholbedarf herrscht aufgezeigt werden. Daraus werden Trends und Hindernisse für die nächsten Jahre abgeleitet.

## 2 Daten

Viele der Begriffe im Zusammenhang mit Daten werden im Alltag sehr regellos verwendet, dabei bestehen zwischen den einzelnen Begriffen große Unterschiede und klare Definitionen. Zunächst soll in diesem Kapitel geklärt werden wie aus Zeichen, Wissen entsteht und welche unterschiedlichen Arten von Daten in Unternehmen auftreten können. Außerdem wird ein typischer Lebenszyklus von Daten angefangen von der Beschaffung bis zur sachgemäßen Entsorgung gezeigt.

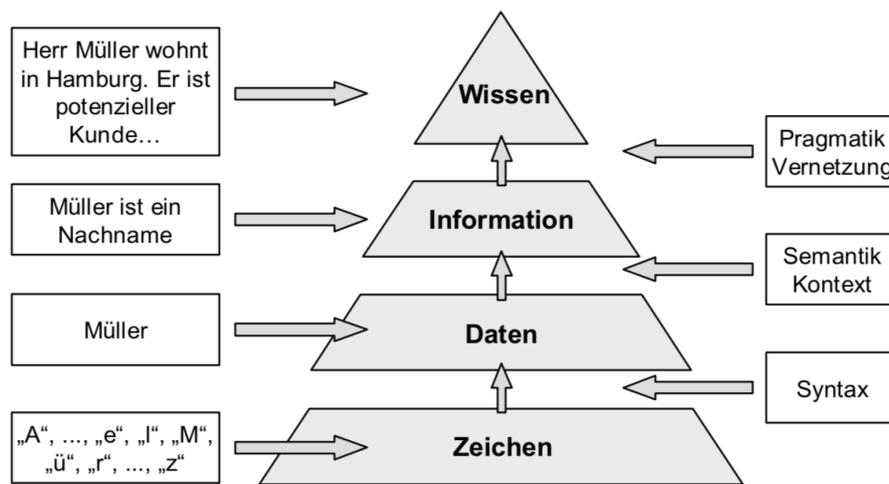


Abbildung 2: Begriffshierarchie<sup>5</sup>

Die Begriffe der Abbildung 2 werden oft in der Umgangssprache so verwendet, dass man den Eindruck gewinnen könnte sie hätten die gleiche Bedeutung. In der Fachliteratur sind diese jedoch eindeutig definiert, so stellen Zeichen die elementarste Ebene der Kommunikation dar. Dazu zählen Buchstaben, Zahlen, Satzzeichen und Leerzeichen. Für den Menschen bieten Zeichen in den meisten Fällen nicht genügend Informationen um sinnvoll genutzt werden zu können. Im Binärsystem, einem Zahlensystem mit der Basis zwei, gibt es aber nur die Zeichen „0“ und „1“. Durch aneinander Reihung dieser zwei Zeichen, lassen sich alle beliebigen Zahlenfolgen darstellen. Eine Zeichenfolge von 0 oder 1 entspricht dabei genau einem Bit, welches der grundlegende Baustein für die Informationsdarstellung in Computern oder Maschinen ist. Der Wert 1, heißt hier Strom fließt und der Wert 0, heißt dass kein Strom fließt. Aus der obigen Begriffspyramide (Abbildung 2) wird deutlich, dass erst die hintereinander Reihung von Zeichen in einer vordefinierten Syntax<sup>6</sup> für Menschen sinnvolle Daten erzeugt werden. Dabei muss die Syntax so gewählt werden, dass diese für den Menschen oder die Maschine geeignet ist. Ein Datensatz in einer Sprache, die von einem Menschen nicht verstanden wird, entspricht somit nur in der Theorie einem Mehrwert zu Zeichen. Eine genaue und treffende Definition des Begriffes „Information“

<sup>5</sup> Quelle: Bodendorf, Fr. (2005), S. 1.

<sup>6</sup> Unter Syntax versteht man vordefinierte Regelsysteme zur Kombination elementarer Zeichen

zu finden ist schon weitaus schwieriger, so wurde schon 1959 dieser Begriff als „Information ist zweckbezogenes Wissen“ formuliert<sup>7</sup>. Je nach Quelle werden hier unterschiedliche Erklärungen gefunden, wobei die Einschätzung dieses Begriffes sehr stark vom Fachbereich und vom Entstehungsjahr des Werkes abhängt. So wird dieser Begriff in der nachrichtentechnischen Branche als Unsicherheit bezeichnet, welche durch das Erscheinen des betreffenden Zeichens beseitigt wird. Innerhalb der Betriebswirtschaftslehre wird Information als eine immaterielle, aber keineswegs kostenlose Ressource verstanden. Heutzutage werden Informationen meist zu immateriellen Wirtschaftsgütern gezählt, die folgende Eigenschaften aufweisen<sup>8,9</sup>:

- Informationen können wiederverwendet werden und mehrere Besitzer haben
- Informationen können zum Nutzen des Betroffenen beitragen
- Informationen haben meist einen Wert, dieser lässt sich aber schwer am Markt ermitteln. Der Wert ist auch stark durch den Nutzen für den Besitzer definiert
- Für das Übertragen von Informationen sind gemeinsame Standards notwendig (z.B. Sprache oder Schnittstelle)

Im Zuge der weiteren Entwicklung des Informationsmanagements wurde der Begriff meist als Zwischenstufe von Daten und Wissen bezeichnet, welcher Daten in Kontext stellt und sie dadurch zu verwertbaren Informationen werden lässt. Im obigen Beispiel wird aus dem Datensatz Müller, so die Information gewonnen, dass Müller ein Nachname ist. Besonders deutlich wird der große Unterschied zwischen Daten und Information, wenn man eine Zeichenfolge von Nummern als Beispiel nimmt. Die Zahlenfolge 100125 kann dabei die Bedeutung des Datums 10.01.2025 genauso haben wie die der Länge 100125 cm. Obwohl also die Zahlenfolge von jedem Menschen verstanden wird, lässt sich die Information daraus erst mit dem richtigen Kontext extrahieren. Diese Information kann nun schon wertvoll sein, aber erst wenn einzelne Informationen vernetzt werden mit anderen Informationen und Zusammenhänge analysiert werden, entsteht daraus Wissen, welches zielgerichtet ist und in der heutigen Zeit entscheidende Wettbewerbsvorteile bringen kann. Im obigen Beispiel ist das Wissen welches aus dem Datensatz Müller gewonnen wird, dass es sich hierbei um einen potentiellen Kunden handelt. Diese Veranschaulichung zeigt aber auch, dass der Mehrwert durch das Wissen sehr abhängig vom Betrachter ist. Wissen welches für einen Betrachter wertvoll ist, ist es nicht unbedingt auch für einen anderen.<sup>10,11</sup>

---

<sup>7</sup> Wittmann, W. (1959), S. 14.

<sup>8</sup> Vgl. Bodendorf, F. (2005), S. 1 ff.

<sup>9</sup> Vgl. Krcmar, H. (2015), S. 11 ff.

<sup>10</sup> Vgl. Bodendorf, F. (2005), S. 1 ff.

<sup>11</sup> Vgl. Krcmar, H. (2015), S. 11 ff.

## 2.1 Arten von Daten

Laut der ISO 8000 <sup>12</sup> Norm sind Stammdaten als Daten definiert, welche unabhängig und fundamental für eine Organisationseinheit sind. Diese Daten lassen wenig Interpretationsspielraum für den Betrachter zu und gelten meist für das gesamte Unternehmen oder zumindest für eine einzelne Organisationseinheit. Ein weiteres Identifikationsmerkmal dieser Daten ist die Statik. Statik bedeutet in diesem Zusammenhang, dass sich diese Daten nur sehr selten ändern. Somit ist die exakte Definition welche Daten im Detail den Stammdaten zuzuordnen sind schwierig. Obwohl diese Daten unternehmensweit gelten, werden nur einzelne Attribute<sup>13</sup> einer Stammdatenklasse zentral verwaltet, da sonst das Datenvolumen und die Komplexität des zentralen Stammdatenmanagements zu hoch wäre.<sup>14</sup>

Welche Attribute einer Stammdatenklasse von einzelnen Organisationseinheiten gespeichert werden müssen, hängt zu großen Teilen von den strategischen Anforderungen dieser Einheit ab.

Unterscheidungsmerkmale, nach denen Attribute klassifiziert werden können<sup>15</sup>:

- Organisatorische Reichweite der Daten (global, lokal)
- Datentyp: Strukturierte Daten oder unstrukturierte Daten
- Ort der Metadatendefinition<sup>16</sup>: Intern oder Extern

Bewegungsdaten sind im Gegensatz zu den Stammdaten nicht statisch, sondern immer auf einen bestimmten Zeitpunkt bezogen (dynamisch). Somit müssen Daten dieser Klasse auch immer das Attribut Datum enthalten, damit sie sich fortlaufend verändern können und trotzdem zugeordnet werden kann von welchem Tag der aktuelle Datensatz ist. Als Beispiel dieser Datenart gelten viele im Betrieb aufzeichnungspflichtige Vorgänge wie z.B. Rechnung an Kunden, Rechnung von Lieferanten oder Gehaltszettel von Mitarbeitern.<sup>17</sup>

Metadaten enthalten Informationen über andere Daten, diese Informationen können sowohl technischer wie auch fachlicher Natur sein. Diese Daten sollen zur Beschreibung dienen und das Auffinden erleichtern. Beispiele sind z.B. ein Referenzschlüssel in der Datenbank oder den Namen des Autors dieser Daten.<sup>18</sup>

Daten können in verschiedenen Formen auftreten, man unterscheidet dabei Daten die eine normalisierte Form haben und so in zeilen- und spaltenorientierte Datenbanken abgelegt werden können. Diese Art von Datentyp wird als strukturierter Datensatz bezeichnet und ist leicht zugänglich und interpretierbar.

Unstrukturierte Daten sind hingegen schwer zugänglich und interpretierbar. Beispiele hierfür sind z.B. Textdateien, Videos oder Bilder. Mit herkömmlichen Datenbanken ist

---

<sup>12</sup> Norm für Datenqualität

<sup>13</sup> Eigenschaften

<sup>14</sup> Vgl. Otto, B.; Österle, H. (2016), S. 29 f.

<sup>15</sup> Vgl. Otto, B.; Österle, H. (2016), S. 30.

<sup>16</sup> Daten, die Informationen über andere Daten enthalten.

<sup>17</sup> Vgl. Spitta, Th.; Brick, M. (2008), S. 85.

<sup>18</sup> Vgl. Otto, B.; Hinderer, H. (2009), S. 4.

eine strukturierte Ablage dieser Daten nicht möglich, der einzige Weg der dauerhaften Speicherung war es ein digitales Abbild zu speichern. Der große Nachteil dieser Speicherart ist, die viel größere Datenmenge und was noch viel erheblicher ist, die Unmöglichkeit der weiteren Analyse oder Verarbeitung dieser Daten. Erst in den letzten Jahren sind zunehmend Verarbeitungsprogramme auf den Markt gekommen, die eine Text und Spracherkennung unterstützen und so die Darstellung dieser Daten in Datenbanken erlauben. Die Analyse so strukturierter Daten stellt noch immer eine Herausforderung da, aber gerade aus diesem Datentyp (Stichwort: Social Media) lassen sich wertvolle Marktkenntnisse gewinnen.<sup>19</sup>

## 2.2 Lebenszyklus von Daten

Wenn Daten als Ressource betrachtet werden, so haben diese einen Wert, eine bestimmte Qualität und einen Lebenszyklus (Abbildung 3). Dieser Zyklus hat einen Einfluss auf die Qualität und in weiter Folge wird dadurch auch der Wert bestimmt, welchen diese Daten später für das Unternehmen haben. Das genaue Management dieser Schritte ist von höchster Bedeutung, um einen möglichst hohen Wert der Daten für das Unternehmen sicherstellen zu können.<sup>20</sup>

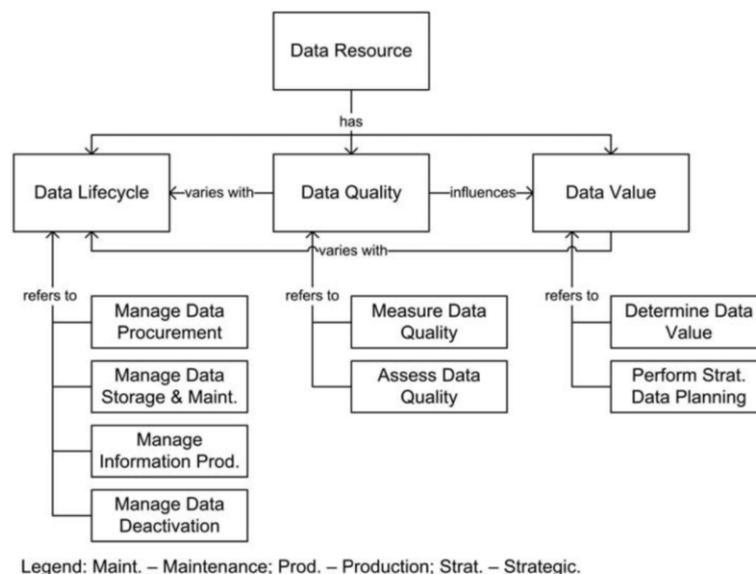


Abbildung 3: Bestandteile von Daten als Ressource<sup>21</sup>

### 2.2.1 Beschaffung

Je nachdem wie die notwendigen Daten akquiriert werden, müssen verschiedene Überlegungen angestellt werden. Vor der Auswahl der richtigen Datenbeschaffungsmethode sollte hinterfragt werden wozu die Daten notwendig sind und wie diese später weiterverwendet werden sollen. Weitere wichtige Punkte sind die

<sup>19</sup> Vgl. Otto, B; Österle, H. (2016), S. 21.

<sup>20</sup> Vgl. Otto, B. (2015), S. 237-238.

<sup>21</sup> Quelle: Otto, B. (2015), S. 238.

notwendige Qualität der Daten und wie oft und zu welchem Zeitpunkt diese verfügbar sein müssen. Zuletzt sollte auch noch überlegt werden ob es gesetzliche Richtlinien für diese Art von Daten gibt und in welchem Format diese für mich überhaupt lesbar oder analysierbar sein müssen. Im Hinblick auf die Industrie 4.0, mit welcher sich diese Arbeit ebenfalls beschäftigt, sei zu beachten, dass viele dieser Bereiche noch gar nicht untersucht wurden und man deshalb gezwungen ist auf eigene Datenerhebungen zurückzugreifen. Ebenfalls gilt dies für die Charakterisierung von eigenen Maschinen in der Industrie, hier können aber durchaus auch Daten für ein Benchmarking<sup>22</sup> zugekauft werden.<sup>23</sup>

Nachteil der eigenen Erhebung sind die vorausgesetzten Fähigkeiten zur Analyse dieser Daten und der höhere zeitliche und finanzielle Aufwand.

Möglichkeiten der Datenbeschaffung<sup>24</sup>:

- Ankauf vorhandener Daten
- Eigene Datenerhebung
- Auffinden und gegebenenfalls Weiterverarbeitung bestehender Daten
- Datentausch

### 2.2.2 Speicherung

Ab einer gewissen Datenmenge können Daten nicht ohne genauere Überlegungen abgelegt werden. Um in einem Big Data Umfeld noch auf Daten zugreifen zu können, müssen diese über Metadaten klassifiziert werden. Über diese Metadaten können diese später wiedergefunden und geordnet werden. Für diese Organisation von Daten gibt es spezielle Software, welche eine automatische Attribuierung der Daten unterstützen. Da in der Zwischenzeit ein Großteil aller Daten elektronisch abgelegt wird, ist eine geeignete Datensicherung auszuwählen, je nach strategischer Wichtigkeit der Daten. Bei Daten, welche extrem schnell wieder verfügbar sein müssen, ist außerdem ein geeignetes Wiederherstellungsverfahren auszuwählen. Welches im Fall eines plötzlichen Datenverlustes (z.B. Server nicht mehr zu erreichen), die Daten innerhalb weniger Minuten wieder verfügbar macht. Die neueste Entwicklung zum Thema Speicherung, welche aber mittlerweile, auch für Privatanwender verfügbar ist, sind Cloud-Speicher. Diese haben den Vorteil, dass so von mehreren Geräten und unabhängig von deren Position auf sehr große Datenmengen zugegriffen werden kann, ohne dass der interne Gerätespeicher mit den Datenmengen mitwachsen muss. Letztendlich sollte bei jedem Speichervorgang darauf geachtet werden ob diese Daten wirklich längerfristig notwendig sind, hierbei gilt es die Balance zu finden, die zu keiner Datenüberflutung führt, aber gleichzeitig alle wichtigen Daten speichert.<sup>25</sup>

Bei der Speicherung von Daten werden drei Arten unterschieden. Die Dateispeicherung (siehe Abbildung 4) , entspricht vom Aufbau der Speicherung wie sie auf Festplatten zu

---

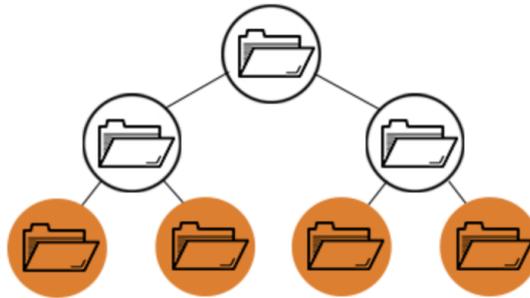
<sup>22</sup> Instrument der Wettbewerbsanalyse

<sup>23</sup> Vgl. Bodendorf, Fr. (2005), S. 4.

<sup>24</sup> Vgl. Bodendorf, Fr. (2005), S. 4.

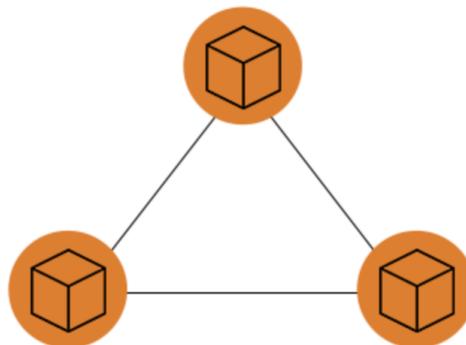
<sup>25</sup> Vgl. Bodendorf, Fr. (2005), S. 4.

finden ist. Die Daten sind Dateien und Ordnern gespeichert. Der Aufbau dieser Ordnerstrukturen am Server entspricht dabei dem Aufbau, den auch die einzelnen User sehen. Vorteile dieser Methode sind, dass sämtliche Metadaten mit abgelegt werden und die geringen Kosten.



**Abbildung 4: Dateispeicherung<sup>26</sup>**

Die Speicherung von Daten mit Block-Storage funktioniert ähnlich, nur werden die Daten in Blöcke aufgeteilt. Diese Blöcke werden dann an verschiedenen Orten separat gespeichert (siehe Abbildung 5) und beim Aufrufen, dann wieder zusammengesetzt. Der Vorteil dieser Speicherart ist die bessere Performance, da die kleineren Datenblöcke dort abgelegt werden wo es für das System am effektivsten ist und auch weniger Metadaten abgelegt werden. Eine Software übernimmt hier die Organisation der Blöcke und die Zuordnung von Daten zu Blöcken.



**Abbildung 5: Block-Speicherung<sup>27</sup>**

Die Objektspeicherung ist die kostengünstigste Variante der Speicherung und wird deshalb vor allem bei Big Data Projekten verwendet. Hierbei werden die Daten als Objekte gespeichert, welchen eine fixe ID zugeteilt wird. Auf eine Verzeichnisstruktur wird komplett verzichtet (siehe Abbildung 6). Jede dieser Objekte ist sozusagen ein eigener kleiner Datenträger, welche eine eigene Kennzeichnung hat sowie die

<sup>26</sup> Quelle: RedHat, <https://www.redhat.com/de/topics/data-storage/file-block-object-storage> (Zugriff: 20.10.2019).

<sup>27</sup> Quelle: RedHat, <https://www.redhat.com/de/topics/data-storage/file-block-object-storage> (Zugriff: 20.10.2019).

Metadaten der Daten enthält. Diese Art der Speicherung wird auch von Cloud-Anwendungen verwendet.<sup>28</sup>

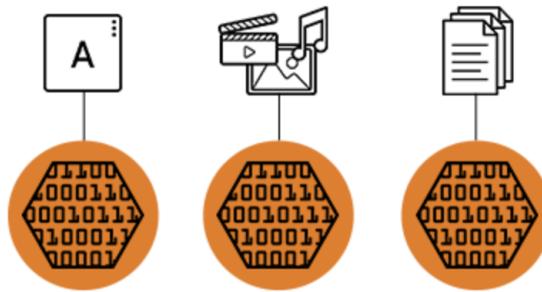


Abbildung 6: Objekt-Speicherung<sup>29</sup>

### 2.2.3 Nutzung und Verwaltung

Eine Bereitstellung von korrekten und aktuellen Daten ist eine wichtige Aufgabe des Datenmanagements (siehe Kapitel 3). Im Zuge dessen müssen aus der großen Menge an Daten, die normalerweise in einem Unternehmen zur Verfügung stehen erst die richtigen von falschen Informationen getrennt werden. Überdies sollten unwichtige Daten nicht gespeichert werden, damit diese nicht unnötig Speicherplatz verbrauchen und bei Suchanfragen nicht mitberücksichtigt werden. Bei neueren Datenarchitekturen (wie z.B. Data Lakes<sup>30</sup>) ist das Thema des Speicherplatzes zwar nur noch zweitrangig, da hier im Normalfall Daten ohne weitere Beschränkungen gespeichert werden, eine entsprechende Durchsicht der Daten und Bewertung der Qualität und Aktualität dieser ist jedoch wichtig. Ein Datenmanager sollte auch darauf achten die Daten richtig zu taggen und zu entscheiden welche Daten jetzt oder in Zukunft für den Unternehmenserfolg wichtig sein könnten. Die Bewertung der Qualität und Aktualität kann in einem Unternehmen in einigen Fällen nicht zentral erfolgen und daher sollte es in jeder Abteilung einen Content-Verantwortlichen geben, welcher für diese Bereiche verantwortlich ist. Aus diesen Daten sollte es für den Nutzer dann möglich sein, die für ihn relevanten Daten mittels Filtereinstellungen leicht zu finden. Bei besonders komplizierten oder großen Dokumenten sollten kurze Zusammenfassungen erstellt werden, welche einen groben Überblick über die wichtigsten Informationen zu diesem Dokument enthalten. So muss dieses nicht bei jeder Anfrage komplett gelesen werden, was wertvolle Arbeitszeit sparen kann und so einen Kostenvorteil für das Unternehmen bedeutet. Bei der Verteilung von Daten gibt es das Pull und Push Prinzip, beim Pull-Prinzip (to pull=ziehen) muss der Nutzer die Daten explizit abrufen, beim Push-Prinzip werden die Daten automatisch von Server zum Nutzer übermittelt. Beide Prinzipien haben ihre Daseinsberechtigung, so arbeiten z.B. Datenbanken nach dem Pull-Prinzip und Newsletter, die automatisch an Kunden geschickt werden, nach dem Push-Prinzip.

<sup>28</sup> Vgl. Tankariya, V.; Parmar B. (2017), S. 232 ff.

<sup>29</sup> Quelle: RedHat, <https://www.redhat.com/de/topics/data-storage/file-block-object-storage> (Zugriff: 20.10.2019).

<sup>30</sup> Ein System welches zur Speicherung und Verwaltung großer Datenmengen dient (siehe auch Abschnitt 4.2)

Im Zuge der Verwaltung von Daten sollte sichergestellt werden, dass Daten den richtigen Personen zur richtigen Zeit zur Verfügung stehen. So sollte es möglich sein auch von Unterwegs auf Unternehmensdaten zugreifen zu können. Zur richtigen Zeit bedeutet, dass z.B. beim Abruf der aktuellen Umsätze diese auch tatsächlich den Umsätzen bis zu diesem Zeitpunkt entsprechen. Gleichzeitig sollte gewährleistet werden, dass betriebsfremde Personen nicht auf die Daten zugreifen können und Mitarbeiter die Daten nicht beliebig verändern können. Auch sollten Mitarbeiter nur auf Daten zugreifen können, für welche sie die Vertraulichkeitsfreigabe haben. Zusätzlich sollte es genaue Angaben geben welche Daten nach Außen veröffentlicht werden dürfen ohne Urheberrechte oder Geheimhaltungsvereinbarungen zu verletzen.<sup>31</sup>

#### **2.2.4 Entsorgung der Daten**

Im Zuge der Wahrung der Aktualität sollten ältere Daten in bestimmten Zeiträumen gelöscht werden. Das gleiche sollte auch bei Daten passieren, von denen mehrere Versionen existieren. Hierbei sollte jedoch auch vorsichtig vorgegangen werden, da ältere Versionen oftmals wieder benötigt werden, falls die neue Version dieser Daten Fehler enthält oder Vergangenheitsdaten für Vergleichsmessungen notwendig sind. Die Entscheidung welche Daten im Detail gelöscht werden können, sollte dem Content-Verantwortlichen obliegen.<sup>32</sup>

Einer der wichtigsten Punkte, die ein erfolgreich implementiertes Datenmanagement erfüllen sollte, ist die Koordination der oben beschriebenen Punkte innerhalb des Unternehmens. Ein erfolgreiches Datenmanagement muss vom Top-Management unterstützt werden und die entsprechenden Personen müssen auch die Durchsetzungsgewalt im Unternehmen bekommen.

---

<sup>31</sup> Vgl. Bodendorf, Fr. (2005), S. 4 f.

<sup>32</sup> Vgl. Bodendorf, F. (2005), S. 5 f.

## 3 Einführung in das Datenmanagement

In diesem Kapitel sind die wichtigsten Schritte zur erfolgreichen Implementierung eines Datenmanagements erklärt. Die Mission eines Datenmanagement liegt nie in der Implementierung selbst, vielmehr sollte das Datenmanagement unternehmerischen Zwecken dienen und auf den Unternehmenszielen aufbauen.

### 3.1 Implementierung eines erfolgreichen Datenmanagements

Laut der Data Management Association (DAMA), einer Organisation welche Konzepte und Praktiken im Bereich des Datenmanagements entwickelt, ist Datenmanagement die Unternehmensfunktion welche sich mit der Entwicklung, Implementierung und Wartung von Regeln und Abläufen beschäftigt welche den Wert von Daten und Informationen kontrolliert, bzw. steigert und diese Daten schlussendlich zu den richtigen Personen liefert. Die genaue Zielsetzung des Datenmanagements leitet sich von den Zielen des Unternehmens ab. Allgemeine Ziele beinhalten die Senkung von Kosten, die Steigerung des Umsatzes und die bessere Abschätzung von Risiken. Hierbei handelt es sich um sehr unterschiedliche Aufgaben, welche sich dadurch auch nicht durch eine bestimmte Abteilung erfüllen lassen. Vielmehr handelt es sich bei einem erfolgreichen Datenmanagement immer um das Zusammenspiel mehrerer Abteilungen. In den letzten Jahren wurde aufgrund der stark ansteigenden Menge an Daten unternehmenseitig sehr viel in die Hardware und Software Infrastruktur investiert, auf die Struktur und die Qualität der Daten in diesem Kreislauf wurde weniger Wert gelegt. Aufgrund der zahlreichen Analysemethoden, deren Aussagequalität sehr stark von der Qualität der vorhandenen Daten abhängt, wird der Fokus derzeit stark auf die Errichtung einer funktionierenden Datenmanagementstruktur gelegt.<sup>33</sup>

Für die Implementierung eines erfolgreichen und effizienten Datenmanagements, welches zugleich kostendeckend arbeitet, sollten folgende Schritte beachtet werden<sup>34</sup>:

1. Ausgehend von einem Top-down Prinzip sollten auf Basis der Unternehmensziele festgestellt werden welche Daten wirklich benötigt werden und wie diese zu organisieren sind. Gerade in der ersten Phase wenn es darum geht welche Daten überhaupt gespeichert werden sollen und wofür diese verwendet werden sollen, sollte ein Team aus dem Management der verschiedenen Geschäftseinheiten sicherstellen, dass mit Hilfe des Datenmanagements die gesetzten Unternehmensziele erreicht werden können. Dieser Punkt beschäftigt sich sehr stark mit der groben Planung wie eine Dateninfrastruktur geschaffen werden kann, welche ein Maximum an Vorteilen für das Unternehmen bietet. Ziel ist es die Planung so zu gestalten, dass eine

---

<sup>33</sup> Vgl. Mosley, M.; Brackett, M. et al. (2009), S.4 ff.

<sup>34</sup> Vgl. Mosley, M.; Brackett, M. et al. (2009), S.18 f.

Voraussetzung geschaffen wird, dass in Zukunft Daten in hoher Qualität zur Verfügung stehen.

Fragstellungen welche hierbei geklärt werden sollten:

- Sind die Daten rechtlich vorgeschrieben.
- Welche Daten sind kritisch für den Unternehmensprozess.
- Welche Daten sind notwendig, um gezielte Entscheidungen zur Unternehmensstrategie treffen zu können.
- Sollen die Daten in einen unternehmensweiten Report integriert werden.
- Werden die Daten verwendet, um damit andere Daten zu erstellen.
- Sollen die Daten nur intern oder auch extern verbreitet werden.
- Wie können die Daten mit möglichst wenig Aufwand dokumentiert, aktualisiert und gewartet werden.
- Wie können die Daten so präsentiert werden, dass sie möglichst gezielt eingesetzt werden können.

Basierend auf der Klärung der obigen Fragen sollte ein Referenzmodell der Daten erstellt werden welches einen Überblick über die notwendigen Daten und deren Zusammenhänge bietet.

2. Nach erfolgreicher Abgrenzung der Daten sollten die notwendigen Prozesse im Unternehmen oder der jeweiligen Geschäftseinheit geklärt werden. Wichtig ist hierbei neben einer Standardisierung auch auf Besonderheiten und Ausnahmefälle Rücksicht zu nehmen. Der Entwurf der Geschäftsprozesse hat daher im Beisein von Spezialisten aus den Fachabteilungen zu erfolgen.
3. Vor der Implementierung einer Lösung ist der wirtschaftliche Nutzen (kann auch erst langfristig erkennbar sein), die technische Machbarkeit und Durchführung zu überprüfen. Bei der Implementierung von Teilsystemen (z.B. Sensoren, welche neue Daten liefern) ist zu klären wie diese Daten in die Gesamtdatenarchitektur des Unternehmens eingegliedert werden können. Außerdem ist auf eine interne und externe Zugangsbeschränkung zu achten.
4. Um den maximalen Nutzen aus den Daten gewinnen zu können sind geeignete Auswerteverfahren anzuwenden. Dazu sollte aber auf die folgenden Punkte, welche die Datenqualität sicherstellen, um die so produzierten Ergebnisse nicht zu verfälschen, geachtet werden:
  - Datengenauigkeit
  - Datenintegrität
  - Datenaktualität (von der Erfassung bis zur Auswertung)
  - Datenrelevanz
  - Klarheit der Datenstruktur

## 3.2 Mission des Datamanagements

Die Mission eines erfolgreichen Datenmanagements ist es den Bedürfnissen aller Stakeholder in Bezug auf Daten gerecht zu werden, dies beinhaltet das zur Verfügung stellen von Daten in einem gewünschten Ausmaß, mit gewünschter Qualität und gewünschter Sicherheit der Daten.

Daraus lassen sich strategische Ziele für das Datenmanagement herleiten<sup>35</sup>:

1. Informationen sollten an die Notwendigkeiten aller Stakeholder im Unternehmen angepasst werden
2. Erfassung, Speicherung und Schutz der Daten
3. Wahrung der Datenqualität
  - Datengenauigkeit
  - Datenintegrität
  - Datenaktualität (von der Erfassung bis zur Auswertung)
  - Datenrelevanz
  - Klarheit der Datenstruktur
4. Wahrung der Geheimhaltung und Zugriffssicherung vor betriebsfremden Personen
5. Maximierung der Effektivität und des Wertes der vorhandenen Daten

Ohne ein firmenweites Datenmanagement wäre es nicht bzw. nur schwer möglich Wissen zwischen den verschiedenen Anwendern zu teilen, da es keine zentrale Datenverwaltung gäbe. Erkenntnisse von einer Abteilung würden so nicht für andere Abteilung sichtbar sein. Dies kann sich auf komplexe Auswertungen, aber ebenfalls auf ein unternehmensweites Telefonverzeichnis, welches zentral verwaltet wird beziehen. Diese Probleme würden auch zu einem messbaren Wettbewerbsnachteil führen, da wichtige Daten nicht oder zu spät geteilt werden würden, was z.B. dazu führen würde, dass Bestellungen nicht dokumentiert sind und Rechnungen dazu in weiterer Folge nicht gestellt werden. Auch bei einer verspäteten Weitergabe der Informationen könnten, die heutzutage teilweise sehr zeitkritischen Aufgaben nicht bewerkstelligt werden. Weitere Kosten würden auftreten und Zeit würde verloren gehen, wenn Auswertungen, die schon gemacht worden wären, wiederholt werden müssten, weil nicht auf die vorhergehenden Daten zurückgegriffen werden könnte. Alle diese Faktoren würden zu einem Verlust der Wettbewerbsfähigkeit des Unternehmens führen, weil Services an Kunden nicht zeitgerecht und mit der notwendigen Qualität durchgeführt werden könnten.<sup>36</sup>

---

<sup>35</sup> Vgl. Mosley, M. et al. (2009), S. 18 f.

<sup>36</sup> Vgl. Gordon, K. (2013), S. 57 ff.

### 3.3 Involvierte Personen

Ein erfolgreiches Datenmanagement benötigt die Zusammenarbeit mehrerer Abteilungen. Wichtig ist die Einigung aller involvierten Parteien auf eine gemeinsame Sprache, damit sichergestellt werden kann, dass Vorgänge korrekt dargestellt werden können und sich so auf die notwendigen Informationsbedürfnisse einlassen lassen. Um in weiterer Folge eine qualifizierte Aussage über eine verbesserte Dateninfrastruktur treffen zu können, sollte die gesamte „Architektur“ des Unternehmens analysiert werden bzw. der Bereich, der für das Datenmanagement wichtig ist. Dieser Schritt sollte den analysierenden Bereich als Teilbereich des Unternehmens sehen und deshalb sollte auch die Eingliederung der Daten als Eingliederung in ein größeres System verstanden werden. Außerdem sollten auch immer die geschäftlichen Komponenten wie Ziele, Strategien und Organisationsstrukturen in der Planung berücksichtigt werden. Nachdem ein Verständnis für die notwendigen Informationen und über die bestehenden Prozesse erlangt wurde, kann mit dem Design einer Dateninfrastruktur begonnen werden. Die generelle Zuständigkeit über das Datenmanagement hat normalerweise die Informationstechnologie-Abteilung zusammen mit Spezialisten aus dem Bereich des Datenmanagements. Jedoch sollten in der ersten Phase der Definition der Problemstellung auch Mitarbeiter des Managements und der Fachabteilung mitwirken. Mitarbeiter des Managements besitzen eine bessere Sicht auf die Notwendigkeit der Information zur Erreichung der Unternehmensziele und Mitarbeiter der Fachabteilungen kennen die jeweiligen Prozesse am besten. Für jede Geschäftseinheit sollte in weiterer Folge ein Datenverantwortlicher definiert werden, welcher der Hauptansprechpartner für alle Daten-Themen der Abteilung sein sollte und außerdem eine brauchbare Qualität der vorhandenen Daten sicherstellen sollte. Eine weitere Aufgabe des Datenverantwortlichen ist die Sicherstellung, dass die vorhandenen Datenressourcen den Notwendigkeiten des Unternehmens entsprechen und nur sinnvolle Daten gespeichert werden. Verantwortung hat dieser einerseits gegenüber der Geschäftsleitung, da zum einen sichergestellt werden sollte, dass diese Informationen den entscheidenden Vorteil zur Konkurrenz bringen können, andererseits sollte dieser nach innen die Interessen aller Datenersteller und Datennutzer des Unternehmens vertreten. Data Management Spezialisten hingegeben sollten den technischen Part dazu übernehmen und sicherstellen, dass die Daten effektiv genutzt werden können und dass eine IT-Infrastruktur aufgebaut wird, in welcher die Daten sicher aufgehoben sind. Diese Tätigkeit passieren in enger Kooperation mit der EDV-Abteilung des Unternehmens. Die Verantwortung der Leitung über alle diese Schritte hat der Chief Information Officer (CIO), dieser ist direkt dem Chief Executive Officer (CEO) unterstellt und auch Mitglied der Geschäftsführung. Hauptaufgabe des CIO ist es die IT-Infrastruktur so einzurichten, dass die Geschäftsprozesse bestmöglich erledigt werden können und dass diese der Unternehmensstrategie entspricht. Dazu zählt auch Früherkennung von neuen Daten für Anwendungen, Steuerungs- und Controllingtätigkeiten. Zwei erst vor kurzen neu entstandene Stellen im C-Level-Bereich in großen Unternehmen, sind die des Chief Data Officers (CDO) und die des Chief Digital Officers (CDO). Während der Chief Digital Officer eher dafür zuständig ist im Rahmen der Digitalisierung neue Geschäftsmodelle zu entwickeln, ist die Rolle des Chief Data Officers sehr stark mit dem Datenmanagement verbunden. Falls ein Unternehmen diese Stelle besetzt hat, ist dieser

für alle Prozesse rund um die Datengewinnung und Datenverarbeitung zuständig und nimmt in einem Unternehmen damit in Bezug auf Daten eine Schlüsselrolle ein, der CIO nimmt dann eher die Rolle des Leiters der IT-Abteilung ein. Die Bestellung eines Chief Data Officers durch ein Unternehmen, ist ein Beweis, dass dieses Unternehmen demjenigen die notwendige Handlungsspielräume geben, damit dieser weitreichende Entscheidungen in Bezug auf die Daten- und Informationsstrategie des Unternehmens treffen zu kann.<sup>37,38</sup>

### 3.4 Abgrenzungen zum Informations- und Wissensmanagement

Umgangssprachlich, aber auch in einigen Fachartikeln werden die Begriffe Informations- und Datenmanagement oft vermischt oder in falschem Kontext verwendet. Die Übergänge zwischen diesen Begriffen sind oftmals sogar in Fachbüchern nicht eindeutig definiert, dies ist dadurch zu erklären, dass diese sehr oft Hand in Hand arbeiten und dieselben bzw. gemeinsamen Zielen dienen. Wie in Abschnitt 3.2 dieser Arbeit erläutert beschäftigt sich das Datenmanagement mit den organisatorischen und technischen Methoden, um aus Daten den maximalen Mehrwert für das Unternehmen lukrieren zu können und so einen Teil zum Geschäftserfolg beitragen zu können. Wie schon aus dem Namen „Datenmanagement“ erkennbar, geht es hierbei um die Sicherstellung der Qualität der Ressource der Daten und die möglichst optimale Nutzung dieser durch gezielte Maßnahmen.

Das Informationsmanagement ist Bestandteil der Unternehmensführung und soll sicherstellen, dass die Unternehmensziele durch Unterstützung der Informatik erreicht werden. Dabei werden Informations- und Kommunikationssysteme aufgebaut, die es ermöglichen sollen Informationen zur Entscheidungsunterstützung zur richtigen Zeit, in der richtigen Qualität an die richtigen Leute zu liefern<sup>39</sup>.

Die Hauptbereiche mit denen sich das Informationsmanagement beschäftigt sind<sup>40</sup>:

- Aufbau einer geeigneten Informationslogistik: Die Logistik sollte sich hierbei an den Zielen und Geschäftsprozessen des Unternehmens orientieren. Dieser Aspekt wird in Zeiten immer größerer Datenmengen von stetiger Bedeutung. Da organisatorische Strukturen durch das Datenmanagement geschaffen werden sollten um die Qualität und Aktualität der Daten zu gewährleisten. Auch die geeignete Hard- und Software Infrastruktur muss vorhanden sein um definierte Informationen überhaupt produzieren oder verarbeiten zu können.
- Steuerung der Informatik: Die Informationstechnologie des Unternehmens sollte sich an den Unternehmenszielen ausrichten, sprich alle Prozesse und Abläufe müssen von IT-Seite in einer Art ausführbar sein, dass die strategischen Ziele des Unternehmens erreicht werden können. Jedoch unterliegt dieser Bereich

---

<sup>37</sup> Vgl. Mosley, M. et al. (2009), S. 5 f.

<sup>38</sup> Vgl. Krcmar, H. (2015), S. 466 f.

<sup>39</sup> Vgl. Krcmar, H. (2015), S. 1 ff.

<sup>40</sup> Vgl. Voß, S; Gutenschwager, K. (2011), S. 65 ff.

auch immer stark wirtschaftlichen Einschränkungen (Was ist wirklich notwendig?).

Das Wissen der Unternehmung wird selbstverständlich auch durch die Daten, welche dem Unternehmen vorliegen gebildet, jedoch bezeichnet Wissen die Gesamtheit der Kenntnisse und Fähigkeiten welches durch die Mitarbeiter im Unternehmen vorliegt. Dabei besteht das organisatorische Gesamtwissen aus den individuellen Wissensvorräten aller Mitarbeiter im Unternehmen.

Kernprozesse des Wissensmanagement sind dabei <sup>41</sup>:

- Wissensidentifikation: Schaffung des Bewusstseins welches Wissen im Umfeld des Unternehmens überhaupt relevant ist. Hierbei sollten die Mitarbeiter aktiv unterstützt werden welches vorhandene Wissen für den Aufbau eines effektiven Wissensmanagementsystems von Relevanz ist.
- Wissenserwerb: Entscheidung welches Wissen intern aufgebaut wird und wo es sich lohnt Kooperationen einzugehen, um so externes Know-how einzukaufen. Neben einer Grundsatzphilosophie sollte diese Entscheidung auch immer eine rationale Kosten-Nutzen Frage sein.
- Wissensentwicklung: Dieser Punkt ergänzt den vorigen Punkt Wissenserwerb, es müssen nämlich jene Bereiche der Wissensweiterentwicklung, die nicht durch einen externen Wissenserwerb abgedeckt sind, selbst im Unternehmen geschaffen werden. Hierbei ist neben der Forschung und Entwicklungstätigkeiten auch der allgemeine Umgang mit dem Ideenmanagement im Fokus, da Wissen generell in jeder Abteilung des Unternehmens entstehen kann.
- Wissensverteilung: Dieser Punkt ist einer der essentiellen Punkte des Wissensmanagement. Individuelles gesammeltes Wissen eines Mitarbeiters des Unternehmens soll der gesamten Organisation zugänglich gemacht werden. Natürlich muss hierbei geklärt werden, welcher Mitarbeiter, was und in welchem Umfang wissen muss.
- Wissensnutzung: Es soll sichergestellt werden, dass das im Unternehmen vorhandene Wissen auch wirklich im Alltag eingesetzt wird.
- Wissensbewahrung: Ist ein weiterer sehr wichtiger Punkt. Die im Unternehmen über die Zeit erworbenen Fähigkeiten und das dazu notwendige Wissen sollen für die Zukunft gesichert werden. Ziel dabei ist es, dass auch beim Ausscheiden von Kernmitarbeitern kein Wissen verloren geht.

---

<sup>41</sup> Vgl. Probst, G. et al. (2010), S. 28 ff.

### 3.5 Datenqualität

In Zeiten von stark steigenden Datenmengen ist es für Unternehmen umso wichtiger jene Daten herauszufiltern, die gewisse Qualitätskriterien erfüllen und so auch zu einer Verbesserung der Entscheidungsqualität im Unternehmen beitragen können. Einheitliche Standards zu Mindestanforderungen der Datenqualität sind schwer möglich, da es hierzu keine allgemeinen Vorgaben gibt. Laut der ISO 9000 (Qualitätsmanagementnorm) ist Qualität als Grad der Erfüllung von Kundenanforderungen definiert<sup>42</sup>. Qualität ist also keine Absolutgröße, die messbar ist und kann als Ergebnis ein Produkt, eine Information oder einen Prozess haben. Wenn wir also annehmen, dass in diesem Fall alle Informationsleser die Kunden sind, geht es darum die Qualitätsanforderungen dieser in Bezug auf die Daten im Unternehmen zu erfüllen. Die Erfüllung dieser Anforderung ist auch messbar. Wang und Strong (1996) fügten zur Definition der Informationsqualität noch hinzu, dass diese immer nach der Eignung der Informationen für den jeweiligen Einsatzzweck beurteilt werden muss, außerdem haben diese vier Kategorien definiert nach welchem die Datenqualität eingeordnet werden kann<sup>43</sup>:

- Intrinsic Datenqualität: Genauigkeit, Objektivität und Fehlerfreiheit der Daten.
- Kontextuelle Datenqualität: Sind die Daten soweit vollständig, dass diese sinnvoll genutzt werden können.
- Darstellungsbezogene Datenqualität: Sind die Daten übersichtlich, also in weiterer Folge auch leicht verständlich.
- Zugangsbezogene Datenqualität: Sind die Daten leicht auffindbar, kann auf diese leicht zugegriffen werden und können diese einfach bearbeitet werden.

---

<sup>42</sup> Vgl. Brüggemann, H; Bremer, P. (2015), S. 3.

<sup>43</sup> Vgl. Wang, R. Y.; Strong, D. M. (1996), S. 6.

Auf Basis dieser Dimension lassen sich folgende Kriterien definieren, nach denen die Informationsqualität einigermaßen objektiv beurteilt werden kann:

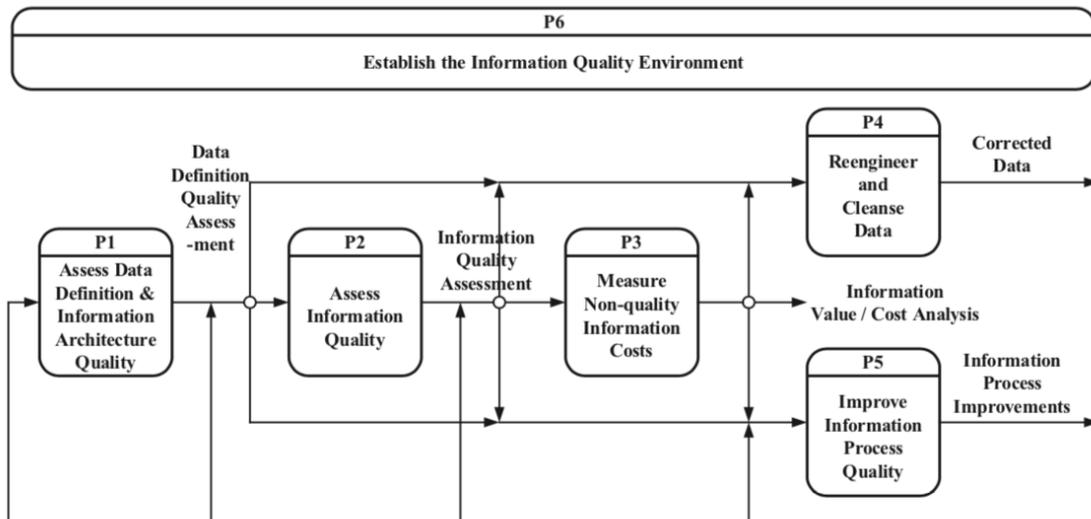
**Tabelle 1: Dimensionen der Informationsqualität** <sup>44</sup>

Qualitätskriterium	Bedeutung
Zugänglichkeit	Informationen sind für die Anwender ohne Umständlichkeit aufrufbar
Angemessener Umfang	Der gebotene Umfang der Informationen erlaubt es die Informationen für den gewünschten Zweck zu nutzen
Glaubwürdigkeit	Der Aufwand zur Überprüfung der Informationen sollte möglichst hoch sein
Vollständigkeit	Als vollständig gelten Informationen, wenn diese zum gewünschten Zeitpunkt ohne Datenlücken verfügbar sind
Übersichtlichkeit	Die Informationen sollten in einem Format abrufbar sein, dass eine möglichst einfache Ansicht oder Bearbeitung erlaubt
Einheitliche Darstellung	Die gleiche Art von Daten sollte fortlaufend in derselben Weise abgebildet werden
Bearbeitbarkeit	Informationen sollten möglichst einfach weiterbearbeitet werden können
Fehlerfreiheit	Der Unterschied zwischen der Realität und den vorhandenen Daten sollte möglichst gering sein
Objektivität	Informationen sollen wertefrei definiert sein
Relevanz	Es sollten nur Informationen dargestellt werden, die auch wirklich sinnvoll für die Organisationseinheit sind
Verlässlichkeit	Informationsquelle, Transportmedium und das verarbeitende System sollten eine möglichst hohe Vertrauenswürdigkeit aufweisen
Aktualität	Informationen sollten so schnell wie möglich beim Anwender erscheinen
Verständlichkeit	Anwender sollten die Informationen ohne weitere Erklärung nutzen können
Wertschöpfung	Die Unternehmensziele sollten durch die Informationen erreicht werden können

Um diese Kriterien sicherstellen zu können ist es anzuraten proaktive Schritte zu setzen, damit Daten, bei denen eine schlechte Datenqualität laut den Kriterien in Tabelle 1 vorliegt, nicht innerhalb der Unternehmung verbreitet werden. Für die Einführung eines geeigneten Modells zur Sicherstellung der Datenqualität und für die möglichst genaue Definition der Kriterien, wann vorhandene Daten als unbrauchbar gelten oder Nachbesserung bedürfen ist im Unternehmen der CIO verantwortlich. Für die Sicherstellung der Qualität der einzelnen Datensätze im Tagesgeschäft ist der Datenverantwortliche jeder Fachabteilung verantwortlich. Wichtig ist, dass das Thema der Datenqualität im Unternehmen ernst genommen wird und ganzheitlich betrachtet wird und nicht als ein Problem der IT-Abteilung abgetan wird.

<sup>44</sup> Quelle: Rohweder, P. et al. (2008), S. 28 (leicht modifiziert).

Zur Sicherstellung einer konsistenten Datenqualität in Unternehmen werden Modelle der Total Information Quality Management Theorie verwendet. Die zwei vorgestellten Modelle setzten hierbei schon an bevor Daten erstellt werden.<sup>45</sup>



**Abbildung 7: Total Information Quality Management Modell nach Larry English<sup>46</sup>**

Nachfolgend werden die einzelnen Schritte des Total Information Quality Management Modells nach Larry English (Abbildung 7) beschrieben welche präventiv für eine anständige Qualität der Daten sorgen sollen<sup>47</sup>:

- P1: Definition der allgemeinen Struktur der Datenarchitektur (z.B. Qualität der Datenbank) wie auch die Definitionen der Qualitätskriterien. Die Analyse erfolgt aus technischer und struktureller Sicht.
- P2: Hier werden die Daten selbst auf die Qualitätskriterien überprüft. Die Analyse erfolgt wieder aus der Perspektive der Erfüllung der gestellten Kundenanforderungen wie auch aus technischer Sicht.
- P3: Hier wird ausgerechnet und abgewogen welche Auswirkungen die Daten auf den Umsatz der Firma haben können.

Von P1, P2 und P3 ist bei Nichterreicherung der gestellten Kriterien eine Abzweigung auf P4 oder P5 vonnöten.

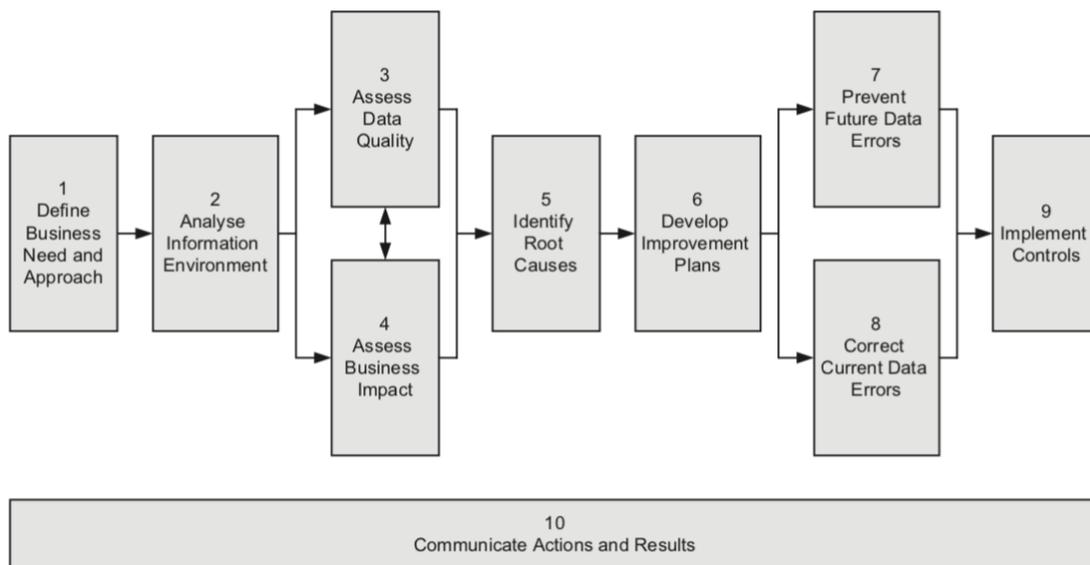
- P4: Vorhandene Datensätze werden gelöscht und neu erstellt
- P5: Vorhandene Datensätze werden ausgebessert bzw. vervollständig bis sie den Qualitätskriterien genügen
- P6: Ist ein kontinuierlicher Verbesserungsprozess (KVP) welcher für eine stetige Verbesserung der Datenqualität sorgen soll

<sup>45</sup> Gordon, K. (2013), S. 96 f.

<sup>46</sup> Quelle: Gordon, K. (2013), S. 96.

<sup>47</sup> Vgl. Gordon, K. (2013), S. 96 f.

Das Modell von Danette McGilvray (Abbildung 8) ist eher für den Aufbau eines Datenqualitätsmanagementprozesses geeignet und nicht als Modell für den kontinuierlichen Gebrauch anzusehen.



**Abbildung 8: Total Information Quality Management Modell nach Danette McGilvray<sup>48</sup>**

Es erfolgt als erstes eine Analyse welche Daten überhaupt notwendig sind und die Definition dieser. Als nächsten Schritt werden die Daten nach Qualitätskriterien und nach dem Grad der Geschäftsauswirkung untersucht. Nach der Beurteilung werden bei schlechtem Abschneiden die Probleme gesucht und Verbesserungspläne ausgearbeitet. Danach werden Schritte gesetzt damit die Datenqualität verbessert wird und in weiterer Folge diese Probleme nicht mehr auftreten. Als finalen Schritt dieses Modells werden Kontrollmechanismen eingebaut, welche die Datenqualität überprüfen.<sup>49</sup>

### 3.6 Informations- und Kommunikationssysteme

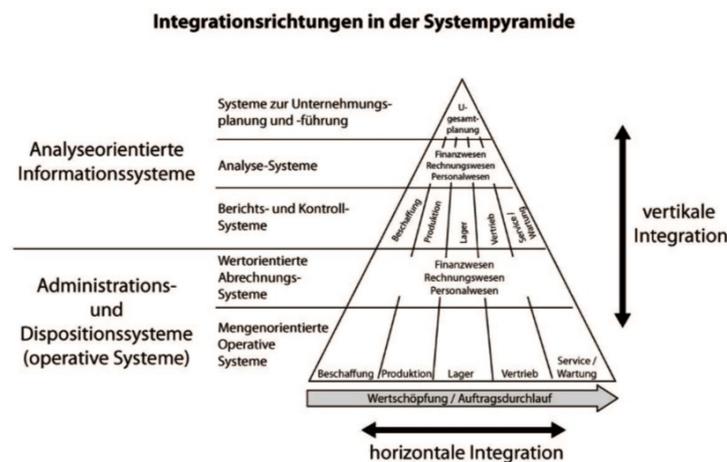
Informations- und Kommunikationssysteme (IuK) sind der Überbegriff für IT-Systeme einer Unternehmung, welche nicht nur für die Informationsbereitstellung verantwortlich sind, sondern auch ein möglichst realistisches Bild über die Beziehungen innerhalb der Unternehmung und zwischen der Unternehmung und der Umwelt ermöglichen sollen. Bei IuK-Systemen wird zwischen operativen Informationssystemen und analyseorientierten Informationssystemen unterschieden. Operative Informationssysteme haben den Sinn standardisierte Arbeitsvorgänge zu automatisieren. Die anfallenden Daten werden dann in Datenbanksystemen gespeichert und lassen sich dort verwalten. Bei den operativen Informationssystemen wird unterschieden zwischen Administrationssystemen, welche wiederkehrende Aufgaben automatisieren, bei denen es keine Entscheidungsspielräume gibt. Prognosen für die Zukunft sind mit Administrationssystemen nicht möglich. Verwendet werden solche

<sup>48</sup> Quelle: Gordon, K. (2013), S. 97.

<sup>49</sup> Vgl. Gordon, K. (2013), S. 97.

Systeme zum Beispiel bei der Erstellung von Kundenaufträgen im Verkauf oder bei der Erstellung von Bestellungen in der Einkaufsabteilung. Dispositionssysteme sollen, das Management bei der Entscheidungsfindung von routinemäßigen und kurzzeitigen Entscheidungen unterstützen. Diese haben niemals eine strategische Reichweite. Beispiel hierfür ist ein Bestellwesen, welches automatisch eine Bestellung für ein Produkt auslöst, wenn dieses unter einen festgelegten Schwellwert fällt. Durch die Verknüpfung mehrerer solcher Systeme (auch über Abteilungen hinweg) sollen Geschäftsprozesse möglichst realistisch dargestellt werden können. Die Verknüpfung der Administrationssysteme entlang der Wertschöpfungskette eines Unternehmens wird horizontale Integration bezeichnet. Beispiel für diese ist der Geschäftsprozess von der Kontaktaufnahme mit dem Kunden bis zur Zahlung des Kunden. Als vertikale Integration wird die Verknüpfung mehrerer Administrationssysteme über mehrere Hierarchieebenen hinaus verstanden, mit dem Zweck, dass analyseorientierte Informationssysteme auf weitreichendere Daten zugreifen können um so bei komplexeren Entscheidungen assistieren zu können<sup>50</sup>.

Diese Zusammenhänge werden folglich nochmals in einer Systempyramide visualisiert (Abbildung 9):



**Abbildung 9: Systempyramide<sup>51</sup>**

Ein für das jeweilige Unternehmen passendes Datenmanagement ist heutzutage unausweichlich, um die Koordination der verschiedenen Aufgaben zu ermöglichen. Die so zur Verfügung stehenden Daten werden u.a. für Datenanalysen verwendet um so den maximalen Nutzen aus den Daten generieren zu können.

<sup>50</sup> Vgl. Gabriel, R. et al. (2009), S. 2 ff.

<sup>51</sup> Quelle: Gabriel, R. et al. (2009), S. 6.

### 3.7 Data Mining

Von der Entstehung der ersten Personal Computer und der ersten Datenbanken in den 1970er Jahre waren nur sehr primitive Datenverarbeitungen und Auswertungen möglich. Mit der Entstehung von relationalen Datenbanken<sup>52</sup> und einer geeigneten Sprache (SQL) mit der es möglich war Abfragen und Manipulationen an diesen Daten zu tätigen, sind in den 1970er Datenmanagementsysteme entstanden. Diese wurden dann sowohl in den vorhandenen Analysemethoden wie auch mit den Datenbanksystemen, welche Daten zur Verfügung stellen ständig weiterentwickelt<sup>53</sup>.

Das Wort Data Mining kommt vom englischen Wort mining, was soviel wie schürfen oder abbauen bedeutet. Ursprünglich wurde dieses Wort in Zusammenhang mit dem Abbau von Gold oder Kohle verwendet (Gold or coal mining). Im Zusammenhang mit Daten wäre der Ausdruck Wissens-Gewinnung wohl passender als die direkte Übersetzung Daten-Gewinnung. Ergänzend zu diesem Ausdruck wurde der Ausdruck Knowledge Discovery in Databases (KDD) verwendet, wobei genaugenommen Data Mining ein essentieller Teilschritt dieses Prozesses ist. Ein KDD-Prozess umfasst alle Schritte, von den Vorbereitungsschritten bis zur Interpretation und Darstellung der gefundenen Daten. Ziel ist es aus einer großen Datenmenge möglichst einfach nützliche Erkenntnisse gewinnen zu können. Heutzutage wird in der Praxis der gesamte Analyseprozess auch als Data Mining bezeichnet.<sup>54</sup>

Herausforderungen für KDD-Algorithmen<sup>55</sup>:

- Große Datenmengen und Datenbanken.
- Komplexe Beziehungen der Daten.
- Datenqualität (unvollständige, fehlende oder falsche Datensätze).
- Überanpassung der Daten. Es werden zu viele erklärende Variablen gefunden, die ein Modell erstellen, welches nicht der Realität entspricht.
- Unteranpassung der Daten. Es werden relevante Variablen nicht berücksichtigt, dadurch wird ebenfalls ein Modell erstellt, welches nicht der Realität entspricht.
- Schlechte Verständlichkeit der gewonnen Muster und/oder Modelle.
- Schlechte Verständlichkeit des gewonnenen Wissens.
- Integration in andere Systeme.

---

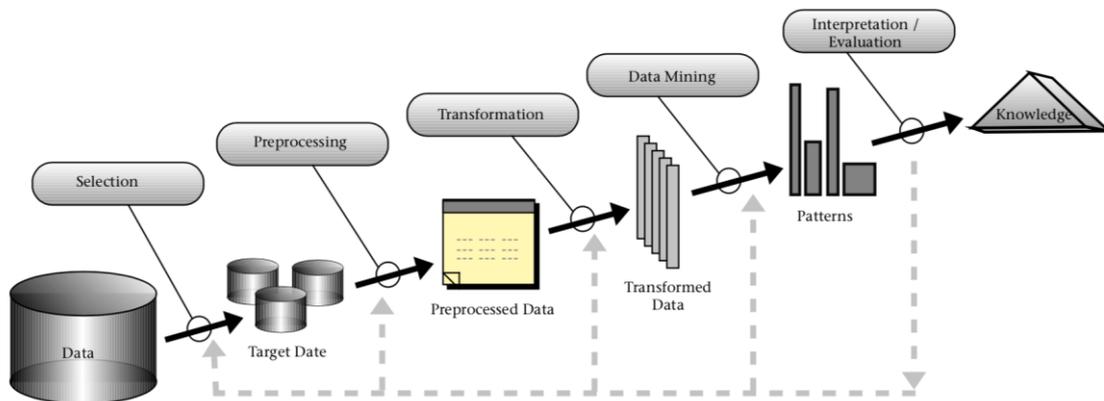
<sup>52</sup> Sammlung mehrerer Tabellen.

<sup>53</sup> Vgl. Han, J. et al. (2012), S. 3 f.

<sup>54</sup> Vgl. Han, J. et al. (2012), S. 6 ff.

<sup>55</sup> Vgl. Fayyad, U. (1996), S. 37 ff.

Modell eines KDD-Prozesses nach Fayyad:



**Abbildung 10: Notwendige Prozessschritte für einen KDD-Prozess<sup>56</sup>**

Detaillierte Beschreibung der einzelnen Stufen eines KDD-Prozesses, wie in Abbildung 10 beschrieben <sup>57;58</sup>:

1. Verständnis und Zieldefinition: Das Anwendungsgebiet bei welchem ein KDD-Prozess angewandt wird sollte genauestens verstanden werden.

Themen die für ein ganzheitliches Verständnis essentiell sind:

- Quelle der Daten (Sensoren, Datenbank oder Data Warehouse).
- Statische oder dynamische Daten (z.B. Zeitreihen).
- Datenmenge (Anzahl und Größe der Datensätze, Anzahl der Attribute, gültige Eingaben der Attribute).
- Datenqualität (fehlende, nicht vollständige oder fehlerhafte Datensätze).

Nachdem das Anwendungsgebiet vollständig verstanden wurde kann ein finales KDD Ziel definiert werden.

2. Datenauswahl: Aufbauend auf die Zielsetzung, sollte definiert werden welche Informationen die Daten enthalten müssen um das KDD-Ziel erreichen zu können. Danach sollte der IST-Informationsgehalt der derzeit vorhanden Datenquellen überprüft werden und zusätzlich nötige Datenquellen sollten identifiziert werden (z.B. Installation neuer Sensoren in der Produktion).

Vorbereitung der Daten: Der im 2.Schritt generierte Zieldatensatz, welcher alle für den KDD-Prozess notwendigen Informationen enthält, wird so vorbereitet, dass die Qualität des Datensatzes den Zielerfordernungen entspricht:

- Umgang mit fehlenden Daten oder Daten-Ausreißern (der Umgang kann z.B. einen der folgenden Schritte enthalten: Extremwerte auslassen, Werte ersetzen durch Mittelwerte, häufigsten Wert verwenden oder löschen)

<sup>56</sup> Quelle: Fayyad, U. (1996), S. 41.

<sup>57</sup> Vgl. Fayyad, U. (1996), S. 42.

<sup>58</sup> Vgl. Konen, W. (2009), S. 11 ff.

- Daten normalisieren<sup>59</sup>
  - Neu gruppieren: Falls zu viele Datensätze vorliegen und nur schwer ein Muster erkannt werden kann, müssen die Daten in vorkategorisierte Bereiche eingeordnet werden.
3. Transformation der Daten: Überprüfung der Relevanz von Variablen und Entfernung oder Zusammenführung dieser. Dieser Schritt soll die Datensätze so vorbereiten, dass ein Data Mining optimal durchgeführt werden kann.
  4. Data Mining: Auswahl der richtigen Methode entsprechend der Datensätze und des gewünschten Resultates. Funktionen welche dabei durchgeführt werden können (Auszug):
    - Klassifikation: Datensätze werden aufgrund ihrer Attribute in vorgegebene Klassen eingestuft (z.B. OK und NOK Teile in einer Produktion).
    - Regression und Vorhersage: Zu den vorhandenen Datenpunkten einer kontinuierlichen Datengröße wird ein passendes Modell gefunden mit welchem Aussagen über die Zukunft möglich sind (z.B. Vorhersage von Abnutzung von Werkzeugen).
    - Clustering: Datensätze werden nach Attributen in nicht vordefinierte Klassen eingestuft. Die Cluster zu denen zugeordnet wird, werden erst bei der Analyse gebildet (z.B. Identifikation von Fehlerursachen bei der Produktion von Autos).
  5. Interpretation: Bei diesem Schritt wird bewertet ob die Ergebnisse der Algorithmen zur Erfüllung des KDD-Zieles geeignet sind. Falls ja, werden diese dokumentiert. Falls nicht wird Schritt 5 nochmals wiederholt bis eine passende Methode gefunden wurde.
  6. Wissensbildung: Die Ergebnisse, welche in Schritt 6 dokumentiert worden sind, sollen nun so in den Geschäftsprozess integriert werden, dass ein möglichst großer Mehrwert für das Unternehmen entsteht. Auf den gewonnenen Ergebnissen kann in weiterer Folge auch aufgebaut werden und es können weiterführende Analysen und Data Mining Methoden angewandt werden.

Jeder Prozessschritt bietet hierbei eine Rückführung auf vorgelagerte Prozessschritte falls die Aufgabenstellung der Stufe nicht hinreichend erfüllt werden konnte.

Entsprechend des technologischen Fortschrittes in allen Lebensbereichen gibt es auch beim Datenmanagement neue Entwicklungen, dabei werden größere Datenmengen verarbeitet, komplexere Auswertungen der Daten durchgeführt. Dazu sind u.a. auch neue Datenarchitekturen notwendig.

---

<sup>59</sup> Vermeidung von Redundanzen.

## 4 Datenmanagement in Zeiten von Industrie 4.0

Einer der wesentlichsten Punkte der vierten industriellen Revolution ist die Digitalisierung und Vernetzung von Menschen, Dingen und Maschinen. Die Integration von neuen Systemen und die neuen Möglichkeiten, die sich daraus ergeben werden in den nächsten Jahren eine gewaltige Herausforderung, aber auch ganz neue Möglichkeiten für Unternehmen darstellen. Zusätzlich erlauben neue Analysemethoden immer komplexere Analysen zu Fragestellungen welche bis dato nicht oder nicht zufriedenstellend beantwortet werden konnten.<sup>60</sup>

### 4.1 Big Data

Obwohl der Ausdruck in der IT-Branche sehr präsent ist, ist die Herkunft nicht zu 100% dokumentiert, es wird vermutet, dass dieser Ausdruck, das erste Mal bei einem Mittagessen in der Firma Silicon Graphics Inc. Mitte der 1990 von John Mashey verwendet wurde<sup>61</sup>. Diebold hatte diesen Ausdruck im Jahr 2000 erstmals in einer Veröffentlichung verwendet<sup>62</sup>, wirklich populär wurde der Ausdruck aber erst in den 2000er Jahren, wo große IT-Firmen wie z.B. IBM in Werbeeinschaltungen auf die schier unendlichen Analysemöglichkeiten von Big Data hingewiesen haben<sup>63</sup>. Der Name gibt auch einen recht guten Eindruck, was genau damit gemeint ist, jedoch gibt es in der Fachliteratur genauere Definition welche Datenmengen wirklich als Big Data bezeichnet werden können.

So wird Big Data von Gartner, einem der führenden Forschungs- und Beratungsunternehmen in der IT-Branche folgendermaßen definiert:

„Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.“<sup>64</sup>

---

<sup>60</sup> Vgl. BMBF (2017), S. 66

<sup>61</sup> Paganoni, M. C. (2019), S. 3.

<sup>62</sup> Diebold, F. X. (2000), S. 115.

<sup>63</sup> Paganoni, M. C. (2019), S. 3.

<sup>64</sup> Gartner, <https://www.gartner.com/it-glossary/big-data/>, (Zugriff: 11.01.2019).

Die 3 “V-Faktoren” (Volume, Velocity und Variety) werden in der Fachliteratur oft verwendet um den Begriff Big Data näher einzuschränken<sup>65</sup>:

- Masse (Volume): Laut einer von IBM durchgeführten Umfrage im Jahre 2012 definieren die meisten User, Big Data als Datensätze ab einer Größe von einem Terabyte<sup>66</sup> bis einem Petabyte<sup>67</sup>. Die gleiche Studie kam ebenfalls zum Ergebnis, dass beinahe alle IT-Anwender sich einig sind, dass Datensätze in der Zukunft noch erheblich größer werden.
- Vielfalt (Variety): Daten unterschiedlicher Quellen und Formate müssen so verarbeitet werden, dass diese gemeinsam weiterbearbeitet, analysiert und gespeichert werden können.
- Geschwindigkeit (Velocity): Wichtig ist hier die Latenz<sup>68</sup> der Daten. Die Geschwindigkeit nimmt in den heutigen Zeiten einen immer höheren Stellenwert ein, da Daten oftmals noch während dem Prozess selbst analysiert werden um so noch rechtzeitig eingreifen zu können, falls der Prozess ungewünschte Ergebnisse liefert. Aber auch außerhalb der Produktionsumgebung hat die Aktualität der Daten aufgrund der Schnelllebigkeit der aktuellen Wirtschaftswelt einen sehr hohen Stellenwert, da Entscheidungen oftmals von den gelieferten Daten in ihrer aktuellsten Fassung abhängig sind.

Neben der Definition von Big Data durch die 3 V, werden mittlerweile auch andere Dimensionen als sehr wichtig in Bezug auf Big Data eingestuft. So wird der Richtigkeit (Veracity) von Daten ein sehr hoher Stellenwert eingeräumt, da Analyseverfahren in ihren Berechnungen auf eine gute Datenqualität angewiesen sind. Eine weitere Eigenschaft von Big Data ist der Wert (Value) welche die Daten für das Unternehmen liefern. Die verschiedenen Einsatzmöglichkeiten und Szenarien sind je nach Branche unterschiedlich, die Verwertbarkeit der Daten sollte aber immer im Mittelpunkt stehen.<sup>69</sup>

Neben diversen Analyseverfahren mit denen z.B. der prädiktiven Instandhaltung von Maschinen mit der die Standzeiten von Maschinen minimiert werden können oder Analysen mit welchen Kundenverhalten besser verstanden werden kann, gibt es auch weitere spannende Anwendungsmöglichkeiten, welche durch Big Data ermöglicht werden. Wie z.B. staatliche Anwendungen zur Analyse des Bürgerverhaltens, der Vernetzung von Medizindaten oder der Echtzeitauswertung von öffentlichen Verkehrsmitteln.<sup>70</sup>

---

<sup>65</sup> Vgl. Schroeck, M. et al. (2012), S. 4.

<sup>66</sup> 1 Terabyte entspricht  $10^{12}$  Byte

<sup>67</sup> 1 Petabyte entspricht  $10^{15}$  Byte

<sup>68</sup> Die Zeit vom Erstellungszeitpunkt der Daten bis diese benutzt werden können.

<sup>69</sup> Vgl. Gandomi, A; Haider, M (2015), S. 139.

<sup>70</sup> Vgl. Oussous, A. et al. (2017), S. 433.

## 4.2 Datenanalyseverfahren

Analysemethoden werden verwendet, um Daten zu kategorisieren, klassifizieren und zusammenzufassen um die Antwort auf eine bestimmte Fragestellung zu bekommen. Hierbei werden statistische Methoden verwendet, um so aus den erhobenen Daten weitreichendere Information zu gewinnen. Während diese Analysen früher genutzt wurden um Beschreibungen der Zustände aus historischen Daten zu erhalten, wird heutzutage versucht aktiv Vorhersagen über zukünftige Ereignisse zu bekommen um ungewünschte Vorgänge im Prozess zu vermeiden. Eine Klassifikation der Analysemethoden nach Wert und Schwierigkeit ist in Abbildung 11 ersichtlich.

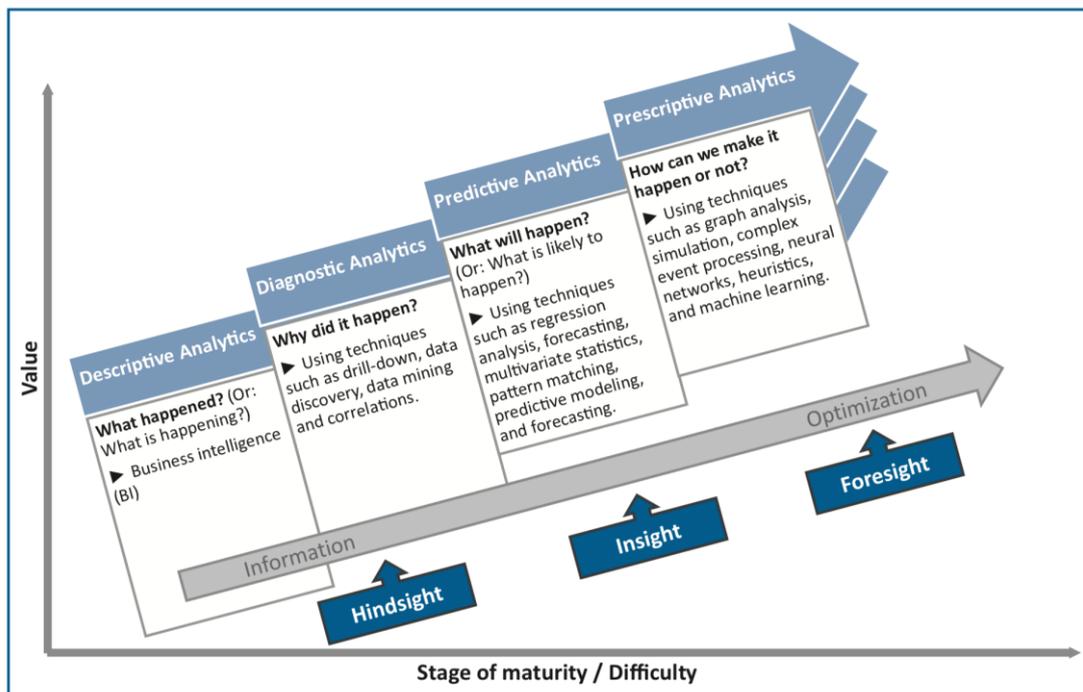


Abbildung 11: Klassifikation der analytischen Methoden Deskriptive Analyse<sup>71</sup>

### 4.2.1 Deskriptive Analyse

Die deskriptive (beschreibende) Analyse wird verwendet, um Zusammenhänge zwischen Variablen festzustellen, um dadurch ermitteln zu können was passiert ist und in weiterer Folge selbst Einschätzungen über zukünftige Daten treffen zu können. Meist wird die deskriptive Analyse auch verwendet, um Daten zu verdichten, um so Erkenntnisse aus den Einzelinformationen ziehen zu können. Das Ziel ist es einen ersten Eindruck bzw. eine Zusammenfassung über eine Datenreihe gewonnen werden kann. Dadurch können schnell Ausreißer einer Datenreihe entdeckt werden bzw. können dadurch weitere Analyseverfahren ausgewählt werden. Eine Fragestellung aus der industriellen Praxis könnte hierbei lauten: „Wie viele der produzierten Teile hatten grobe Fehler“. Um ein besseres Verständnis über die Daten zu gewinnen können diese im Zuge der deskriptiven Analyse auch visualisiert werden.

<sup>71</sup> Quelle: Romeike, F.; Eicher, A. (2016), S. 168.

Hierzu kommen oft Data-Mining-Algorithmen zum Einsatz, einige der zugrundeliegenden mathematischen und statistischen Verfahren werden untenstehend näher beschrieben.<sup>72</sup>

**Tabellen und grafische Darstellungsmethoden**

Häufigkeitstabellen

Häufigkeitstabellen eignen sich für die Darstellung von Häufigkeiten einer bestimmten Merkmalsausprägung bei einer überschaubaren Menge von Datensätzen. Bei einer höheren Anzahl an Datensätzen sollten zur Wahrung der Übersichtlichkeit Klassen gebildet werden, welche Werte von Merkmalsausprägungen zusammenfassen (z.B. 1 Klasse: Weniger als zwei Fehler, 2.Klasse: Weniger als 4 Fehler, 3.Klasse: Weniger als 6 Fehler). Zusätzlich zur absoluten Häufigkeit, welche die Anzahl an Datensätzen mit dieser Merkmalsausprägung angibt, wird öfters auch die relative Häufigkeit angegeben, welche sich aus der absoluten Häufigkeit dividiert durch die Gesamtzahl der Datensätze berechnet.

Kontingenztafeln sind eine Erweiterung von Häufigkeitstabellen, welche die Häufigkeit des Auftretens von zwei definierten Merkmalen abbilden (Abbildung 12). Die definierten Merkmale tragen die Bezeichnungen  $x_1$  bis  $x_m$  und  $y_1$  bis  $y_L$ , die Anzahl der Datensätze welche beide Merkmale in der entsprechenden Zeile und Spalte erfüllen, werden mit  $n_{11}$  bis  $n_{ML}$  angegeben.<sup>73</sup>

	Y						
X	$y_1$	$y_2$	...	$y_l$	...	$y_L$	$\Sigma$
$x_1$	$n_{11}$	$n_{12}$	...	$n_{1l}$	...	$n_{1L}$	$n_{1\bullet}$
$x_2$	$n_{21}$	$n_{22}$	...	$n_{2l}$	...	$n_{2L}$	$n_{2\bullet}$
⋮							
$x_m$	$n_{m1}$	$n_{m2}$	...	$n_{ml}$	...	$n_{mL}$	$n_{m\bullet}$
⋮							
$x_M$	$n_{M1}$	$n_{M2}$	...	$n_{Ml}$	...	$n_{ML}$	$n_{M\bullet}$
$\Sigma$	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet l}$	...	$n_{\bullet L}$	$n$

**Abbildung 12: Kontingenztafel<sup>74</sup>**

Diagramme

Histogramme (Säulendiagramme) werden verwendet, um die Häufigkeit des Auftretens einer Merkmalsausprägung darzustellen. Das zu untersuchende Merkmal wird dabei in Klassen aufgeteilt, die eine sinnvolle Visualisierung der Häufigkeitsverteilung erlauben. Hierbei wird jeder Wert einer Merkmalsausprägung durch eine senkrechte Säule dargestellt. Die Höhe der Säule entspricht der Häufigkeit des Auftretens dieser

<sup>72</sup> Vgl. Cleff, T. (2008), S. 4 f.

<sup>73</sup> Vgl. Toutenburg, H.; Heumann, C. (2008), S. 97 f.

<sup>74</sup> Quelle: Freie Universität Berlin, <https://wikis.fu-berlin.de/pages/viewpage.action?pageId=712409813>, (Zugriff: 30.01.2019).

Merkmalsausprägung (Abbildung 13). Bei nominal skalierten Variablen<sup>75</sup> werden meist Kreisdiagramme zur Darstellung verwendet. Die Größe des Kreissegmentes gibt dabei die Häufigkeit des Auftretens der Merkmalsausprägung der Variable an.<sup>76</sup>

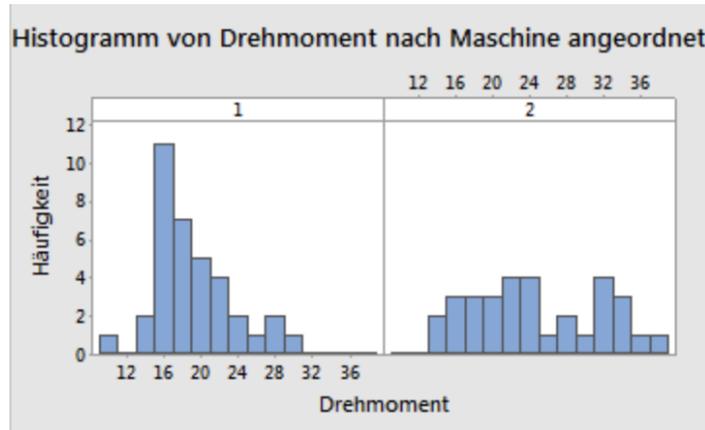


Abbildung 13: Histogramm<sup>77</sup>

### Boxplot

Boxplots (Abbildung 14) sind ein gutes und beliebtes Instrument in der Praxis um eine Visualisierung von Daten während der ersten Analyse zu erhalten. Datensätze werden durch fünf Punkte beschrieben: Minimum, untere Quartil, Median, obere Quartil und das Maximum. Dazu wird eine Box gezeichnet, welche durch die untere und obere Quartile begrenzt ist. Die obere Quartile reicht von Maximalwert bis zum Median und die untere vom Minimalwert bis Median. Der Median wird durch eine weitere Linie auf dieser Box gekennzeichnet.<sup>78</sup> An der Lage des Medians innerhalb der Box kann beurteilt werden ob die Verteilung symmetrisch verteilt ist, weniger gut geeignet ist die Darstellung von bi- oder multimodalen Verteilungen, welche mehrere Maxima aufweisen.<sup>79</sup>

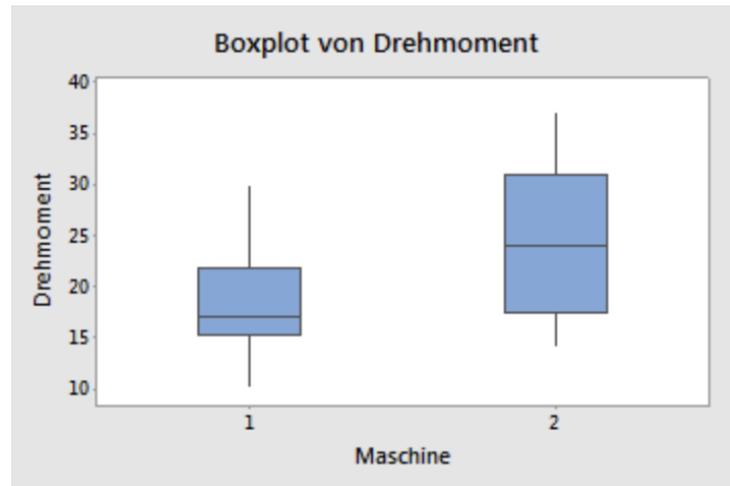
<sup>75</sup> Nominal skalierte Variablen haben keine natürliche Rangfolge, aber dennoch Merkmale, die eine Unterscheidung ermöglichen (z.B. Farben).

<sup>76</sup> Vgl. Toutenburg, H.; Heumann, C. (2008), S. 40 f.

<sup>77</sup> Quelle: Minitab (2018), <https://support.minitab.com/de-de/minitab/18/help-and-how-to/statistics/basic-statistics/how-to/display-descriptive-statistics/before-you-start/example/> (Zugriff: 30.01.2019).

<sup>78</sup> Toutenburg, H.; Heumann, C. (2008), S. 351 f.

<sup>79</sup> Vgl. LeBlanc, D (2004), S. 48 f.

Abbildung 14: Box-Plot Diagramm<sup>80</sup>

### Lage- und Skalenparameter

#### Arithmetisches Mittel

Dieser Lageparameter ist in der Praxis der am häufigsten verwendete, wenn es um die Verteilung eines quantitativen Merkmals geht. Der Mittelwert wird berechnet in dem die Werte aufsummiert und dann durch die Gesamtanzahl der Werte dividiert werden. Der Mittelwert entspricht dabei dem Schwerpunkt der Daten. Dieser Wert ist zwar sehr leicht zu interpretieren, jedoch ist er bei einer kleinen Anzahl an Datensätzen sehr empfindlich gegenüber Ausreißern, sprich ein sehr großer oder kleiner Wert kann den Mittelwert erheblich beeinflussen.<sup>81</sup>

#### Median

Der Median gibt die Mitte einer Distribution von Datensätzen an. Es liegen also somit gleich viele Wert über wie unter dem Median. Zur Berechnung werden die Werte in aufsteigender Reihenfolge aufgeschrieben, jetzt wird bei einer ungeraden Anzahl an Daten der mittlere Wert als Median notiert, bei einer geraden Anzahl an Daten wird der Durchschnitt der zwei Werte in der Mitte als Median angenommen. Der Median hat den Vorteil, dass dieser unempfindlich gegenüber Ausreißern ist.<sup>82</sup>

#### Modus

Als Modus bzw. Modalwert wird jener Wert bezeichnet, der bei den vorhandenen Datensätzen am häufigsten vorkommt. Diese Kennzahl kann auch bei Datensätzen verwendet werden, welche vorher in Klassen unterteilt worden sind. Der Modus ist ebenfalls unempfindlich gegenüber Ausreißern und kann auch bei nominal-skalierten Variablen verwendet werden.<sup>83</sup>

<sup>80</sup> Quelle: Minitab (2018), <https://support.minitab.com/de-de/minitab/18/help-and-how-to/statistics/basic-statistics/how-to/display-descriptive-statistics/before-you-start/example/> (Zugriff: 30.01.2019).

<sup>81</sup> Vgl. Toutenburg, H.; Heumann, C. (2008), S. 59 f.

<sup>82</sup> Vgl. Cleff, T. (2008), S. 50 ff.

<sup>83</sup> Vgl. Toutenburg, H.; Heumann, C. (2008), S. 50

## Verteilungs-Parameter

### Spannweite

Die Spannweite ist die Breite der Merkmalsausprägungen einer Messreihe. Diese ist durch den Minimal- und Maximalwert begrenzt. Weitere Aussagen über die Distribution abseits der Extrema sind mit dieser statistischen Größe nicht möglich. Sie ist daher auch sehr empfindlich gegenüber Ausreißern und wird meist nur bei kleinen Stichprobenumfängen genutzt.<sup>84</sup>

### Quartile

Quartile teilen die Distribution in vier gleich große Abschnitte auf. Das erste Quartil gibt dabei an, dass 25% der Daten kleiner oder gleich diesem Wert sind. Das zweite Quartil entspricht dem Median der Distribution. Das dritte Quartil gibt an das 75% der Daten kleiner oder gleich diesem Wert sind. Als Interquartilbereich wird der Abstand zwischen Q1 und Q3 bezeichnet, dieser umfasst 50% der Daten. Dieses Streuungsmaß ist nicht so empfindlich auf Ausreißer wie die Spannweite.<sup>85</sup>

### Mittlere absolute Abweichung

Gibt an wie stark die Daten um das arithmetische Mittel streuen. Zur Berechnung wird der Mittelwert aus den einzelnen Abweichungen jedes Wertes berechnet.<sup>86</sup>

### Standardabweichung

Die Standardabweichung berechnet sich aus der Wurzel der Varianz. Die Varianz wird aus der Summe der quadrierten Abstände der einzelnen Werte vom arithmetischen Mittel, dividiert durch die Anzahl der verfügbaren Daten berechnet. Durch die Quadrierung wird das gegenseitige Aufheben von negativen und positiven Abweichungen verhindert. Die Standardabweichung ist ein sehr bekannter Streuungsparameter, der jedoch stark durch Ausreißer verfälscht wird.<sup>87</sup>

### Variationskoeffizient

Ist definiert als die Standardabweichung dividiert durch den Mittelwert und beschreibt wie die Standardabweichung die Streuung eines Merkmals. Der Vorteil gegenüber der Standardabweichung ist, dass der Variationskoeffizient unabhängig von der Einheit der Daten ist. Hierdurch lassen sich Streuungen von Messungen mit anderen Messungen vergleichen, welche mit jeweils anderen Einheiten gemessen wurden.<sup>88</sup>

## 4.2.2 Diagnostische Analyse

Die diagnostische Analyse ist eng verwandt mit der deskriptiven Analyse, welche ebenfalls Vergangenheitsdaten untersucht. Während die deskriptive Analyse sich aber

---

<sup>84</sup> Vgl. Toutenburg, H.; Heumann, C. (2008), S. 82.

<sup>85</sup> Vgl. Toutenburg, H.; Heumann, C. (2008), S. 82.

<sup>86</sup> Vgl. Toutenburg, H.; Heumann, C. (2008), S. 74.

<sup>87</sup> Vgl. Cleff, T. (2008), S. 59 f.

<sup>88</sup> Vgl. Cleff, T. (2008), S. 61.

eher nur mit der Fragestellung beschäftigt: „Was ist passiert“, geht die diagnostische Analyse einen Schritt weiter und versucht rauszufinden wo, warum und wieso eine gewisse Situation eingetreten ist.<sup>89</sup>

Die Arbeitsschritte der diagnostischen Analyse lassen sich grob in drei Kategorien unterteilen<sup>90</sup>:

- Identifikation von Anomalien: Aufgrund der Auswertungen der deskriptiven Analyse sollen jene Bereiche identifiziert werden, in den Fragestellungen auftauchen, welche sich durch die vorhandenen Daten nur schwer beantworten lassen. In der industriellen Praxis wären dies z.B.: Warum läuft Maschine A, besser als Maschine B oder warum tritt bei einem bestimmten Arbeitsschritt in gewissen Abständen immer der gleiche Fehler auf.
- Analyse der Anomalien: Es müssen Daten gefunden werden, mit welchen die aufgetretenen Anomalien erklärt werden können. Oftmals reicht es hier nicht bestehende Daten des Unternehmens zu analysieren, sondern es muss auf externe Daten zugegriffen werden oder es müssen neue Daten im Unternehmen entstehen. In einer Produktionsumgebung wäre dies z.B. möglich in dem eine neue Prüfmethode zur Prozessüberwachung installiert wird.
- Erkennen von Zusammenhängen: Neben der Erklärungsfindung von Anomalien durch die Schaffung von neuen Daten, können Erklärungen auch oftmals durch das Finden von versteckten Kausalzusammenhängen gefunden werden.

Folgende Techniken werden oft für die diagnostische Analyse eingesetzt:

### Varianzanalyse

Die Varianzanalyse wird eingesetzt, wenn Mittelwerte aus mehr als 2 Stichproben auf Unterschiede und Signifikanz zu überprüfen sind. Die Varianzanalyse betrachtet hierbei die Wirkung einer oder mehrerer Einflussvariablen (Faktoren) auf eine oder mehrere Zielvariablen. Für die Durchführung der Varianzanalyse wird die Gesamtvarianz aufgeteilt in die Varianz zwischen den Gruppen (Abweichung der Gruppenmittelwerte von Gesamtmittelwert aller Gruppen) und die Varianz innerhalb von Gruppen (Abweichung der Einzelmittelwerte einer Gruppe). Die Varianz innerhalb der Gruppen wird als Fehlervarianz bezeichnet, weil diese sich meist schwerer erklären lässt und die Varianz zwischen den Gruppen wird als erklärbare Varianz bezeichnet, welche sich durch die Faktoren erklären lässt. Diese beiden Varianzen werden dann miteinander verglichen, um feststellen zu können ob zwischen den Stichproben Gesetzmäßigkeiten bestehen. Wenn sich die Varianzen innerhalb der Gruppe und diese zwischen den Gruppen stark unterscheiden, kann angenommen werden das in den Gruppen unterschiedliche Gesetzmäßigkeiten vorherrschen (z.B. unterschiedliche Bedingungen bei der Durchführung der Messungen).<sup>91</sup>

---

<sup>89</sup> Vgl. Sampaio, P.; Saraiva, P. (2016), S. 27.

<sup>90</sup> Vgl. IBM, <https://www.ibm.com/blogs/business-analytics/diagnostic-analytics-101-why-did-it-happen/>, (Zugriff: 31.01.2019).

<sup>91</sup> Vgl. Eckey, H.-F. et al. (2002), S. 93 ff.

### Regressionsanalyse

Regressionsanalysen untersuchen ob und wie stark eine abhängige Variable von einer oder mehreren unabhängigen Variablen beeinflusst werden. Falls es einen Zusammenhang zwischen den Variablen gibt, beeinflusst der Wert der unabhängigen Variable, den Wert der abhängigen Variable(n). Dieser Zusammenhang kann als Funktion ausgedrückt werden. Die Regressionsanalyse ist nutzbar um zukünftige Werte besser abschätzen und vorhersagen zu können. Im Gegensatz zur Korrelation muss bei der Regression festgesetzt werden, welche Variable vorhergesagt werden soll.<sup>92</sup>

Die Durchführung einer Regressionsanalyse erfordert folgende Schritte<sup>93</sup>:

1. Aufbereitung der Daten: Um ein möglichst realistisches Modell entwickeln zu können, müssen die Daten möglichst vollständig und genau sein. Wichtig ist auch das alle Variablen definiert werden, die Einfluss auf das gegebene Modell haben. Je nachdem welche Daten vorhanden sind, kann so ein geeignetes Regressionsverfahren gefunden werden.
2. Anpassung des Modells: Abweichungen werden korrigiert indem Fehlerwerte und Annäherungen berechnet werden und im Regressionsmodell berücksichtigt werden.
3. Überprüfung des gefunden Modells: In diesem Schritt sollte überprüft werden wie gut das Modell den Zusammenhang zwischen den Variablen beschreibt. Zur Überprüfung ist es essentiell, dass auch wirklich alle Variablen berücksichtigt worden sind, ansonsten kann es passieren dass ein Modell gefunden wird, welches zwar die vorhandenen Werte gut beschreibt, jedoch zukünftige Ereignisse nicht präzise darstellen kann, weil hierfür die notwendigen Parameter fehlen.
4. Prognose: Wenn das Modell entsprechend validiert worden ist, kann es auch für Prognosezwecke verwendet werden, wenn Aussagen getroffen werden, die über die vorhandenen Daten hinausgehen, wird dieser Vorgang Extrapolation genannt.

### **Hypothesentest**

Der Hypothesentest wird verwendet um eine vermutete Wahrscheinlichkeit (Hypothese) auf Richtigkeit zu prüfen. Dabei wird versucht das Gegenteil der Annahme zu widerlegen. Es wird dabei von den zu Verfügung stehenden Stichproben versucht auf eine vermutete Grundgesamtheit zu schließen, eine Garantie für die Richtigkeit dieser Vermutung kann also nicht gegeben werden.<sup>94</sup>

Zur Durchführung des zweiseitigen Hypothesentests werden folgende Schritte ausgeführt<sup>95</sup>:

1. Nullhypothese  $H_0$  aufstellen, welche die Wahrscheinlichkeit angibt, die geprüft werden soll.

---

<sup>92</sup> Vgl. Cleff, T. (2008), S. 147 ff.

<sup>93</sup> Vgl. Ryte (2016), <https://de.ryte.com/wiki/Regressionsanalyse>, (Zugriff: 31.01.2019).

<sup>94</sup> Vgl. Treyer, O. A. G. (2003), S. 113.

<sup>95</sup> Vgl. Schulze, P. M.; Porath, D. (2012), S. 449 ff.

2. Festlegen des Signifikanzniveaus (Irrtumswahrscheinlichkeit). Die Irrtumswahrscheinlichkeit ist dabei als die Gegenwahrscheinlichkeit der Sicherheitswahrscheinlichkeit definiert. Sollen 90% aller Ausgänge abgedeckt sein, ist das Signifikanzniveau 5%. Daraus ergeben sich die Werte des Annahme- und des Ablehnungsbereiches.
3. Stichproben durchführen
4. Hypothese wird bestätigt oder verworfen. Es kann nicht gezeigt werden, dass die Hypothese wirklich korrekt ist, nur dass sie mit dem festgelegten Signifikanzniveau als richtig anzusehen ist.

Beim einseitigen Hypothesentest wird hingegen geprüft ob eine angenommene Wahrscheinlichkeit signifikant nach oben oder unten abweicht.

### Qualitätsregelkarte

Die Qualitätsregelkarte (Abbildung 15) wird für die Überwachung von Fertigungsprozessen verwendet. Dazu werden mehrere Stichproben am Fertigungsprozess gezogen. Die Mittelwerte dieser Stichproben werden dann in ein vordefiniertes Koordinatensystem eingetragen. Aufgrund der Prozessinformationen werden außerdem Eingriffsgrenzen definiert, die Überschreitung der oberen oder unteren Eingriffsgrenze (OEG und UGW) weist auf potentielle Probleme des Fertigungsprozesses hin. Die obere und untere Toleranzgrenze (OWG und UWG) wird nur in die Qualitätsregelkarte eingetragen, wenn Einzelwerte und keine Mittelwerte in die Prozessregelkarte eingetragen werden. Diese Grenzen geben an welche Abweichungen maximal vorliegen dürfen damit ein erzeugtes Produkt verwendbar ist. Die Prozessregelkarte dient als Frühwarnsystem, aber auch dazu den Prozess besser zu verstehen und dadurch verbessern zu können. Man unterscheidet zwischen Prozessregelkarten in denen OEG und UEG nicht fix definiert sind, sondern durch die Ergebnisse der letzten Serie definiert werden und der Annahmeregularte bei welcher der Eingriffsbereich durch die Toleranzgrenze definiert wird (meist liegt der Eingriffsbereich bei 80% der oberen und unteren Toleranzgrenze)<sup>96</sup>.

Standardmäßig definierte Eingriffskriterien, bei denen der Prozess näher untersucht werden sollte lauten<sup>97</sup>:

- Die Eingriffsgrenze wurde einmalig überschritten, durch ein gezieltes Nachregeln kann so ein Ausschuss verhindert werden.
- Feststellung von Trends: Wenn sieben aufeinanderfolgende Mittelwerte eine bestimmte Tendenz aufweisen (z.B. kontinuierlich steigend).
- Feststellen eines Runs: Wenn sieben aufeinanderfolgende Punkte oberhalb oder unterhalb der Mittelwertlinie liegen.
- Middle Third: Wenn weniger als 40% oder mehr als 90% der letzten 25 Mittelwerte im mittleren Drittel der Eingriffsgrenzen liegen.

<sup>96</sup> Vgl. Herrmann, J.; Fritz, H. (2016), S. 151.

<sup>97</sup> Vgl. Cloudt, T. (2001), [https://www.cloudt.de/pdf\\_archiv/3kregelkar.pdf](https://www.cloudt.de/pdf_archiv/3kregelkar.pdf), (01.02.2019).

- cp-Werte nicht erreicht: Die aus den letzten 25 Mittelwerten ermittelten cp und cpk Werte müssen größer als 1,33 (4x die Standardabweichung) sein. Die cp- und cpk-Werte geben dabei die Prozessfähigkeit an. Ein Prozess ist fähig, wenn er keinen Ausschuss liefert und die Ausschussrate der Prozessfähigkeitskennzahl cp mindestens 1,33 beträgt.<sup>98</sup>

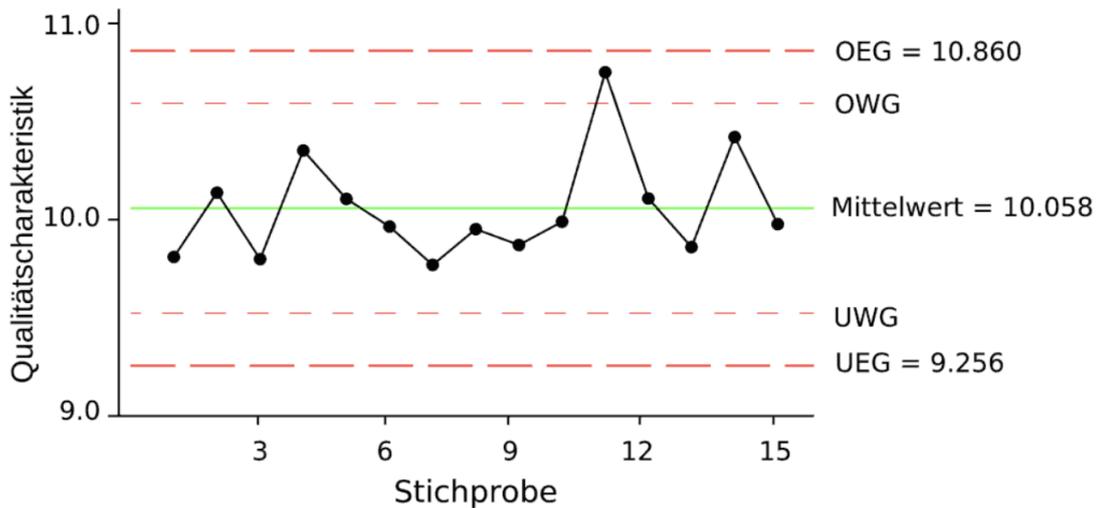


Abbildung 15: Qualitätsregelkarte<sup>99</sup>

#### 4.2.3 Prädiktive Analyse

Während bei der deskriptiven und diagnostischen Analyse Vergangenheitsdaten für die Analyse von bereits stattgefundenen Ereignissen benutzt werden (Was ist passiert und warum ist es passiert?) werden bei der prädiktiven Analyse, Modelle geschaffen anhand derer Prognosen für die Zukunft möglich sind. Hierbei werden Werkzeuge der Statistik, der Modellierung, des Data Minings, des Machine Learnings und zeitweise sogar schon künstliche Intelligenz zusammengeführt. Im Zuge dieser Analysen werden sowohl historische Daten, wie auch aktuellen Daten verarbeitet und dabei Muster und Beziehungen aus strukturierten und unstrukturierten Daten erkannt. Daraus lassen sich dann Modelle erstellen, welche es erlauben durch die Eingabe von Variablen Prognosen über bestimmte Werte zu machen. Diese Vorgehensweise ist komplexer als bei diagnostischen Modellen, bietet aber nicht nur die Möglichkeit des Reportings sondern basierend auf den neuen Daten können weitreichende Analysen, Planungen und Simulationen durchgeführt werden. Ziel ist es mit Hilfe dieser Analysen die Zukunft vorhersagen zu können (Was wird passieren?).<sup>100</sup>

<sup>98</sup> Vgl. Cloudt, T. (2001), [https://www.cloudt.de/pdf\\_archiv/3kregelkar.pdf](https://www.cloudt.de/pdf_archiv/3kregelkar.pdf), (01.02.2019).

<sup>99</sup> Quelle: Quality & Service Business GmbH, <https://www.quality.de/lexikon/qualitaetsregelkarte-qrk/>, (Zugriff: 01.02.2019).

<sup>100</sup> Vgl. Romeike, F.; Hager, P. (2016), S. 168 ff.

Die Möglichkeiten der Nutzung dieser Analysemethoden sind hier sehr weitreichend unter anderem finden diese bei folgenden Beispielen Anwendung<sup>101</sup>:

- In der industriellen Praxis können Mithilfe dieser Daten Wartungszeitpunkte von Geräten und Maschinen schon sehr präzise vorhergesagt werden. Hierbei werden durch spezielle Algorithmen sowohl aktuelle Daten der Maschinen wie auch historische Daten berücksichtigt um so den besten Zeitpunkt für eine Inspektion der Maschine festzustellen, der so knapp wie möglich vor dem wahrscheinlichen Termin des Ausfalls der Maschine liegt. Durch die Verhinderung von zu früher Wartung und eines potentiellen Maschinenausfalls können die Maschinenkosten minimiert werden.
- Bei Versicherungen und Banken werden durch spezielle Algorithmen Unregelmäßigkeiten erkannt, die auf einen Betrug hinweisen.
- Durch gezielte und definierte Qualitätskriterien kann Produktausschuss rechtzeitig erkannt werden.
- Crossselling- (Bemühen zusätzlich zum angefragten oder verkauften Produkt, weitere passende Produkte anzubieten<sup>102</sup>) und Upselling- (Bemühen, statt dem angefragten Produkt oder in weiterer Folge ein teureres Produkt zu verkaufen<sup>103</sup>) Potentielle von Kunden können erkannt und gezielt beworben werden. Dabei werden Analysen benutzt die Vorhersagen wann und auf welche Art von Werbung dieser bestimmte Kunde am besten reagiert.
- Unzufriedene Kunden werden durch gezielte Maßnahmen erkannt bevor sie den Entschluss fassen auf ein Konkurrenzprodukt zu wechseln.

Modelle der prädiktiven Analyse basieren dabei meist auf Klassifikationsmodellen (welche die Einteilung von neuen Datensätzen in Klassen erlauben) oder Regressionsmodelle (Vorhersage von bestimmten Parametern), dabei werden im Hintergrund u.a. folgende Techniken angewandt:

### **Regressionsanalyse**

Siehe Regressionsanalyse (Abschnitt 4.2.2)

In der prädiktiven Analyse werden diese Modelle verwendet um aufgrund der gefundenen Muster, Prognosen über zukünftige oder nicht vorhandene Werte erstellen zu können. Wichtig ist es hier einen Zeitrahmen der sinnvoll erscheint, je länger eine Prognose extrapoliert, umso höher ist die Schwankungsbreite dieser.<sup>104</sup>

### **Entscheidungsbäume**

Entscheidungsbäume dienen der Klassifikation von Daten anhand von definierten Merkmalen. Diese lassen sich im Normalfall leicht verständlich aufsetzen und die Entscheidungsfindung ist sehr transparent gestaltet. Jeder Zweig des

---

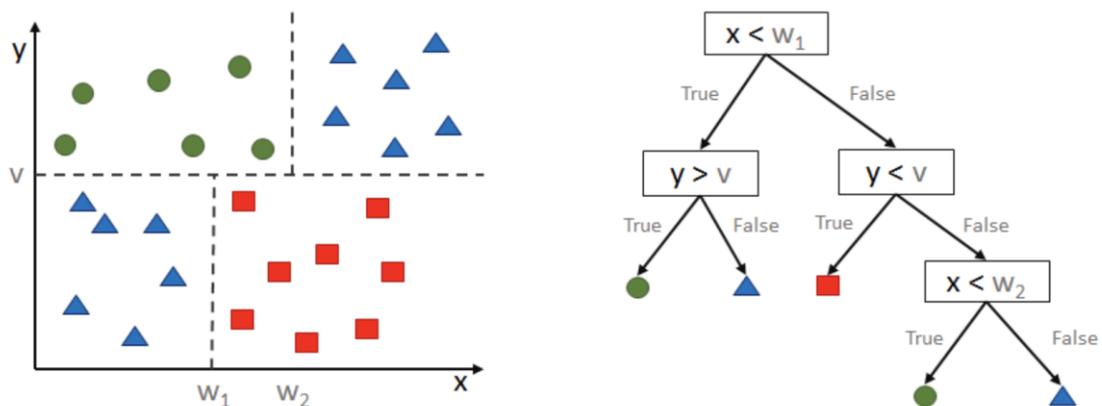
<sup>101</sup> Vgl. Krypczyk, V; Bochkor, O., <https://entwickler.de/online/development/predictive-analytics-praxis-tipps-579847089.html> (Zugriff: 04.02.2019).

<sup>102</sup> Jäggi, S; Portmann, C. (2012), S. 193

<sup>103</sup> Jäggi, S; Portmann, C. (2012), S. 193

<sup>104</sup> Vgl. SAS (2018), [https://www.sas.com/de\\_at/insights/analytics/predictive-analytics.html](https://www.sas.com/de_at/insights/analytics/predictive-analytics.html) (Zugriff: 04.02.2019).

Entscheidungsbaumes stellt eine Möglichkeit in der Entscheidung dar, meist werden binäre Entscheidungsbaume verwendet, welche nur „Wahr“ (True) oder „Falsch“ (False) als Antwortmöglichkeit haben. Diese Entscheidung kann nun aufgrund vorgegebener Parameter automatisiert beantwortet werden. Dadurch wird nicht nur eine Klassifikation der Daten möglich, sondern es gibt zu jedem Datensatz eine nachvollziehbare Begründung warum die Daten in diese Klasse eingeordnet wurden. Diese Prozedur wird solange fortgesetzt bis die Daten hinreichend klassifiziert sind, diese Klassifikation wird als Blatt des Entscheidungsbaumes bezeichnet. Entscheidungsbaume welche quantitative Variablen klassifizieren, werden Klassifikationsbaume genannt. Wenn nominal skalierte Variablen klassifiziert werden, so wird dieser Entscheidungsbaum, Regressionsbaum genannt.<sup>105</sup>



**Abbildung 16: Entscheidungsbaum - Klassifikationsbaum (rechts) und Klassifikation der Daten in bestimmte Klassen (links)<sup>106</sup>**

### Neuronale Netzwerke

Mit neuronalen Netzwerken wird versucht die Funktionsweise des menschlichen Gehirns nachzubilden. Dies stellt ein sehr modernes und anspruchsvolles Verfahren dar, das zur Modellierung von sehr komplexen Prozessen verwendet wird. Das neuronale Netzwerk besteht dabei aus einzelnen Neuronen welche einzelne Datenverarbeitungseinheiten darstellen. Diese sind schichtweise in einer Netzarchitektur angeordnet. Es gibt dabei immer eine Eingabeschicht (Input Layer), ein oder mehrere Zwischenschichten (Hidden Layer) und eine Ausgabeschicht (Output Layer). Jedes Neuron einer Schicht ist dabei immer mit allen Neuronen der nächsten Schicht verbunden. Die Anzahl der Zwischenschichten ist theoretisch unbegrenzt, jedoch führt eine weitere Schicht immer zu einer verzögerten Bearbeitung, deshalb wird hier eine entsprechende Rechenleistung benötigt. Der Output eines Neurons ist dabei immer der Input des nächsten. Je mehr Trainingsdaten verfügbar sind umso besser funktioniert das neuronale Netzwerk. Dieses Verfahren ist sehr flexibel einsetzbar, da sowohl lineare als auch nicht lineare Zusammenhänge abgebildet werden können. Neuronale Netzwerke

<sup>105</sup> Vgl. Bankhofer, U; Vogel, J. (2008), S. 273.

<sup>106</sup> Quelle: Aunkofer, B. (2017), <https://data-science-blog.com/blog/2017/02/13/entscheidungsbaumverfahren-artikelserie/>, (Zugriff: 06.02.2019).

werden meist dann verwendet, wenn keine mathematische Formel zur Erklärung des Zusammenhanges vorliegt und diese auch nicht unbedingt gesucht wird, sondern wenn es eher um eine Vorhersage zukünftiger Ereignisse geht.<sup>107</sup>

### **Bayessche Analysen**

Bei dieser Analysemethode werden objektiven Wahrscheinlichkeiten (auf Basis von relativen Häufigkeiten) mit weiteren Informationen für die Berechnung dieser Wahrscheinlichkeit kombiniert, um die Wahrscheinlichkeit für eine Hypothese anzupassen. Das Bayes-Theorem besagt dabei, dass eine gesuchte Wahrscheinlichkeit von anderen Wahrscheinlichkeiten abhängig ist. Die Neuklassifizierung der Eintreffwahrscheinlichkeit von einem gesuchten Ereignis passiert über die Kombination mehrerer gewichteter Hypothesen.<sup>108</sup>

### **Ensemble-Modellierung**

Bei dieser Methode werden mehrere ähnliche Modelle mit Testdaten trainiert. Die Ergebnisse dieser Modelle werden dann in einem Gesamtmodell kombiniert und erhöhen somit die Genauigkeit der Vorhersage im Gegensatz zu den Einzelmodellen.

<sup>109</sup>

#### **4.2.4 Präskriptive Analyse**

Die Präskriptive Analyse stellt die letzte Stufe der Datenauswertung dar, hier wird durch die Nutzung von komplexen Algorithmen die Frage beantwortet: „Wie können wir ein bestimmtes Ereignis herbeiführen“. Wie bei der prädiktiven Analyse wird versucht Vorhersagen aufgrund derzeitiger Parameter und vorliegender historischer Daten zu treffen. Im Gegensatz dazu werden aber bei der prädiktiven Analyse Zusammenhänge berücksichtigt, für die es keine historischen Daten gibt, die aber eine definierte Wahrscheinlichkeit haben, dass sie auftreten.<sup>110</sup>

Beim Praxisbeispiel aus der prädiktiven Analyse zur Vorhersage der Wartungszeitpunkte von Maschinen, wäre die Aufgabenstellung für eine Präskriptive Analyse die Erhöhung der Wartungsintervalle um 10 Prozent. Die Empfehlung einer präskriptiven Analyse für die Abnutzung eines Bohrers könnte also lauten, das Material des Bohrers zu ändern oder das Drehmoment, mit dem der Bohrer betrieben wird anzupassen. Da für keine der beiden Empfehlungen Daten vorliegen, müssen andere Quellen für diese Vorhersagen verwendet werden, um daraus die optimale Strategie empfehlen zu können. Die optimale Strategie kann dabei auch eine Kombination mehrerer Lösungsansätze sein. Natürlich lassen sich aus prädiktiven Analysen auch oft Empfehlungen ableiten, die zu einer Optimierung eines Prozesses führen.

---

<sup>107</sup> Vgl. Bourg, A. et al. (2006), S. 6 ff.

<sup>108</sup> Vgl. Gleißner, W. (2014), S. 68 ff.

<sup>109</sup> Vgl. Hudec, M. (2011), S. 6.

<sup>110</sup> Vgl. Sappelli, M. et al. (2017), S. 45 ff.

Die präskriptive Analyse besteht im Grund immer aus drei Arbeitsschritten<sup>111</sup>:

- Ausarbeitung der vorhandenen Möglichkeiten, um ein bestimmtes Ereignis herbeizuführen.
- Finden der optimalen Empfehlung aus diesen Möglichkeiten.
- Begründung der Entscheidung.

Für die Ausarbeitung dieser Strategien werden viele technische Disziplinen angewendet und die Resultate dieser miteinander verknüpft <sup>112</sup>:

- Machine Learning
- Angewandte Statistik
- Unternehmensforschung
- Verarbeitung natürlicher Sprache
- Signalverarbeitung
- Mustererkennungsalgorithmen
- Verarbeitung von Bild-Informationen
- Analyse von Graphen
- Simulationen (z.B. Monte-Carlo)
- Neuronale Netzwerke

Da diese Algorithmen für sehr komplexe Fragestellungen verwendet werden, ist die bloße Ausarbeitung einer Vorhersage nicht ausreichend. Damit diese Strategie im Geschäftsalltag zur Entscheidungsunterstützung verwendet wird ist es wichtig, dass die Methode möglichst transparent die gewählte Strategie darlegen und begründen kann. Die Entscheidung ob diese Strategie überhaupt sinnvoll ist obliegt hierbei noch immer dem Menschen, da diese Algorithmen bloß auf Grundlage der ihnen zur Verfügung stehenden Daten entscheiden, die Anpassungsfähigkeit an neue Umstände und die Kreativität die den Menschen zur Verfügung steht, kann von diesen Algorithmen noch nicht erreicht werden. Hingegen können Korrelationen in sehr großen Datenmengen erkannt werden, welche vom Menschen nur schwer zu erkennen sind, aber auch hier ist Vorsicht geboten ob es sich wirklich um einen kausalen Zusammenhang handelt, bei großen Datenmengen tritt das Phänomen der Scheinkorrelationen sehr oft auf, wobei hier in Wahrheit keine Ursache-Wirkungsbeziehung zwischen den Variablen besteht. Die Einsatzmöglichkeiten dieser Technologie sind sehr vielseitig und einige der zahlreichen Einsatzmöglichkeiten werden sich wohl erst in den nächsten Jahren bis Jahrzehnten zeigen. Die Entwicklung dieser Technologie geht einher mit der Entwicklung von Maschinen mit künstlicher Intelligenz, beide Technologien werden derzeit schon in der Praxis eingesetzt, jedoch ist der tatsächliche Einsatz und das Entwicklungsstadium ausgehend vom enormen Potential dieser Technologie noch relativ am Anfang. <sup>113</sup>

---

<sup>111</sup> Vgl. Sappelli, M. et al. (2017), S.46.

<sup>112</sup> Vgl. Poornima, S.; Pushpalatha, M. (2016), S. 472.

<sup>113</sup> Vgl. Romeike, F.; Eicher, A. (2016), S. 168 ff.

Es gibt klare Faktoren, die für die Nutzung einer präskriptiven Analyse erfüllt sein müssen, da diese ansonsten entweder nicht sinnvoll oder gar nicht einsetzbar ist <sup>114</sup>:

- Kosten-Nutzen-Verhältnis: Es sollten nur Anwendungen analysiert werden, bei welchen das für die Analyse eingesetzte Kapital auch einen dementsprechenden Nutzen bringt. Dies gilt meist für Anwendungen wo die gefundenen Strategien, einen großen und langfristigen Einfluss auf die weitere Vorgehensweise haben.
- Komplexität: Das Problem sollte multiple Variablen und Optimierungsmöglichkeiten haben.
- Ziel: Die Analyse muss ein quantitativ messbares Ziel haben.
- Beschränkungen: Gibt es zusätzliche Beschränkungen, die das ganze Projekt noch komplexer machen (z.B. bestimmte Werte können nicht verändert werden).
- Daten: Der Erfolg einer präskriptiven Analyse ist sehr von zur Verfügung gestellten Daten abhängig, nur aufgrund der Analyse wie sich das System auf bisherige Änderungen verhalten hat, ist eine Vorhersage über das Verhalten von zukünftigen Strategien möglich.

Die Anwendungsfelder dieser Analyseverfahren sind sehr vielfältig, als Beispiel sind nachstehend Anwendungsfelder genannt, welche die obigen Faktoren meist erfüllen und so einen Überblick über die verschiedenen Anwendungsarten geben sollen <sup>115</sup>:

- Erdölindustrie: Aufgrund der hohen CAPEX <sup>116</sup> und OPEX <sup>117</sup> ist, kann hier jede Art von Kosteneinsparung zu sehr hohen Profiten führen. Daten sind durch die Anwendung von Messungen und Sensor meist viele verfügbar und durch die Anwendungen der gleichen Messungen an verschiedenen Orten sind auch Vergleichswerte meist verfügbar. Potentielle Anwendungsgebiete wären z.B. die vorausschauende Instandhaltung (Predictive maintenance) von wichtigen Maschinen in Raffinerien, wo jeder ungeplante Ausfall weitreichende Folgen haben kann oder die Hilfe bei der Bewertung von potentiellen Ölbohrungen wie auch die Echtzeitüberwachung von Bohrungen.
- Speditionen: Hier werden präskriptive Analysen zum Teil auch schon eingesetzt um die Routen von LKWs unter Berücksichtigung von freien Frachträumen, gesetzlichen Stehzeiten, Rampenöffnungszeiten und Priorisierung von wichtigen Aufträgen so zu optimieren, dass die vorhandenen Ressourcen optimal eingesetzt werden und auch hier wird in Echtzeit auf Änderungen wie z.B. Unfälle, Pannen oder die Erkrankungen eines Fahrers reagiert.
- Logistik: Auch hier werden bereits präskriptive Analysen eingesetzt, um die bestmöglichen Automatisierungslösungen zu finden, um die Lagerung und Kommissionierung von Paketen möglichst zu vereinfachen und zu automatisieren. Hierbei werden nicht nur Lagerplätze effektiver genutzt und Aufträge selbständig von Robotern abgewickelt, es werden bereits Ansätze getestet wo Roboter Arbeitsaufträge zu menschlichen Arbeitern zuteilen und Vorschläge für die Optimierung liefern und so die Produktivitätsleistung erhöhen.

---

<sup>114</sup> Vgl. Sappelli, M. et al. (2017), S. 46 f.

<sup>115</sup> Vgl. Sappelli, M. et al. (2017), S. 47 ff.

<sup>116</sup> Capital expenditures – langfristige Bilanz-Aktiva

<sup>117</sup> Operating expenditures – Aufwendungen für den laufenden Betrieb

### 4.3 Datenarchitekturen

Die Datenbanken, in denen die Informationen der Administrationssysteme gespeichert sind, sind meist sehr fachspezifisch auf einen bestimmten Bereich und für bestimmte Zwecke aufgebaut. Bei der Zusammenführung von verschiedenen operativen Informationssystemen mit heterogenen Daten bedient man sich der Konzepte der Data Warehouses. Diese haben den Zweck unternehmensweit Daten zu sammeln, in eine konsistente und logische Form zu transferieren und zu verteilen, damit in weiterer Folge komplexere Analysen durchgeführt werden können und so fundiertere Entscheidungsfindungsunterstützungen durch das System möglich sind.<sup>118</sup>

Data Warehouses arbeiten dabei mit dem „Schema-on-write“, das bedeutet, dass Daten welche importiert werden einem bestimmten Schema (Datentyp, Datengröße, Beziehungen) entsprechen müssen. Dieses Vorgehen ist zeitintensiv, da die Daten für die Speicherung speziell verarbeitet werden müssen. Vor dem Speichervorgang überprüft das DWHs ob diese Strukturierung eingehalten worden ist, damit nur richtig strukturierte Daten abgespeichert werden<sup>119</sup>.

Hauptunterschiede eines Data Warehouses im Gegensatz zu operativen Datenbanken sind:

- Integrierte Datenbasis: Zusammenführung von Daten aus verschiedenen heterogenen Datenquellen.
- Nicht flüchtige Datenbasis: Auf im *Data Warehouse* befindliche Daten kann zwar zugegriffen werden und neue Daten werden ständig hinzugefügt, jedoch ist es sehr unüblich bestehende Daten in diesem zu verändern. Datensätze in Datenbanken werden in Tagesgeschäft jedoch regelmäßig gelöscht oder hinzugefügt.
- Vergleich von Daten über einen längeren Zeitraum möglich. Im Gegensatz dazu enthalten operationale Datenbanken immer nur die aktuellen Datenwerte.

Ein *Data Warehouse* besteht zwingend aus folgenden Komponenten<sup>120</sup>:

- Importkomponente: Zuständig um den Datenbestand des *Data Warehouses* am neuesten Stand zu halten, mittels einer Schnittstelle werden interne und externe Inhalte über eine ETL-Schnittstelle<sup>121</sup> importiert.
- Verwaltungskomponente: Datenbestände werden durch diese Komponente organisiert und in verschiedenen Verdichtungsstufen gespeichert. Zusätzlich zu den Daten werden auch Metadaten gespeichert, welche weiterführende Informationen (z.B. Herkunft, Verdichtungsstufe oder Regeln für die Transformation aus den Quelldaten) zu den Daten enthalten. Die Daten können entweder in einem zentralen Archivierungssystem oder in mehreren kleineren Data Marts gespeichert werden. Data Marts speichern dabei Daten eines bestimmten Bereiches des Unternehmens oder für eine bestimmte

<sup>118</sup> Vgl. Gabriel, R. et al. (2009), S. 7.

<sup>119</sup> Vgl. Freiknecht, J.; Papp, S. (2018), S. 23.

<sup>120</sup> Vgl. Leser, U. (2005), S. 11.

<sup>121</sup> Extraktion, Transformation und Laden ist eine Schnittstelle zum Datenimport.

Benutzergruppe. Durch mehrere kleinere Datenbestände lassen sich gegenüber unternehmensweiten Speicheroptionen, Kosten- und Geschwindigkeitsvorteile erzielen<sup>122</sup>.

- Zugriffskomponente: Jene Schnittstelle von der auf die Daten im *Data Warehouse* zugegriffen werden kann. Häufig wird auf diese Daten mit speziellen Programmen zugegriffen, welche die OLAP<sup>123</sup> Technologie unterstützen. Mit dieser lassen sich multidimensionale Auswertungen erstellen. Beispiel für eine multidimensionale Abfrage ist z.B. eine Auswertung aller Verkäufe (1.Dimension), eines gewissen Zeitraums (2.Dimension), eines gewissen Geschäftsbereiches (3.Dimension) und eines bestimmten Verkaufsgebietes (4.Dimension). Komplexere Abfragen lassen sich durch Data Mining Systeme durchführen.

Heutzutage werden neben DWHs auch Data Lakes als Architektur für Big Data Plattformen verwendet. Data Lakes werden vor allem bei Daten verwendet wo die Datenmengen extrem groß sind und die Daten vielen Änderungsvorgängen ausgesetzt sind. Data Lakes arbeiten mit dem Schema-on-read, es wird also nicht jeder Schreibvorgang geprüft, sondern die Daten werden im rohen bzw. unstrukturierten Zustand gespeichert. Diese Speicherung spart Zeit, da die Daten vor der Speicherung nicht strukturiert werden müssen. Die vorhandenen Daten werden erst in gewisse Schemen gebracht, wenn dies für Analysen oder für wiederkehrende Aufgaben benötigt wird. Durch das Fehlen einer vordefinierten Struktur ist jedoch die Zugänglichkeit dieser Daten für den normalen Anwender eingeschränkt. Für Analysen steht so jedoch ein großer Pool an Daten zu Verfügung, auf welche sehr schnell zugegriffen werden kann. Im Gegensatz zu DWHs benötigen Data Lakes keine spezielle Hardware und lassen sich schnell an einem geänderten Speicherbedarf anpassen. Noch flexibler ist die Verwendung von Data Lakes in der Cloud, da der benötigte Speicherplatz jederzeit angepasst werden kann, wenn der Bedarf zunimmt. Cloud-Lösungen sind meist auch kosteneffektiver, da nur der tatsächlich genutzte Speicherplatz bezahlt werden muss und zudem keine teure Hardware angeschafft werden muss.<sup>124</sup>

## 4.4 Aktuelle Trends

Die folgenden Beispiele für Trends werden schon heute teilweise angewandt, es ist aber zu bedenken, dass diese Technologien aufgrund ihrer schier unendlichen Möglichkeiten in Zukunft einen noch weitaus höheren Stellenwert in der IT-Industrie einnehmen werden. Durch die stetige Zunahme der Datenmengen wird sich auch die Verbreitung der nachfolgenden Technologien in den nächsten Jahrzehnten noch deutlich steigern und es werden aufbauend auf diesen Technologien in Zukunft komplett neue Methoden entwickelt werden.

---

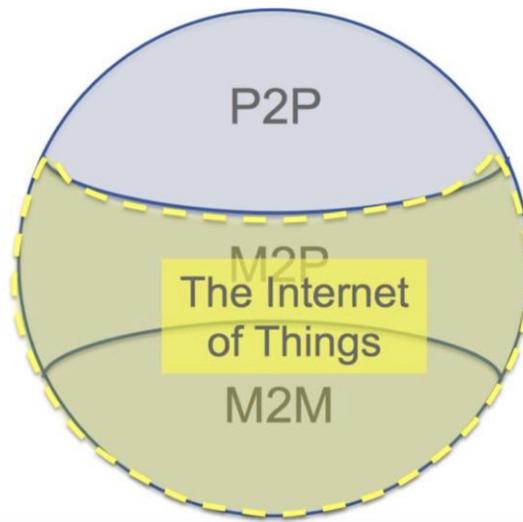
<sup>122</sup> Vgl. Laudon, K. C. (2016), S. 291 f.

<sup>123</sup> On-Line Analytical Processing ist ein Softwaretool mit welchem relativ einfach komplexere Analysen der vorhanden Daten durchgeführt werden können.

<sup>124</sup> Vgl. Gorelik, A. (2019), S. 2ff.

#### 4.4.1 Internet of Things

Das Internet of Things (IoT) oder zu Deutsch, das Internet der Dinge beschreibt die Vernetzung von jeder Art von Geräten durch die Verwendung von eingebetteter Software<sup>125</sup> und Sensoren die es ermöglichen dass diese Geräte mit anderen Geräten kommunizieren und Daten austauschen können. Embedded Software wird häufig in Geräten eingesetzt, von denen man bis jetzt nicht gewohnt war, dass sie sich vernetzen bzw. einen anderen Zweck erfüllen können als ihren ursprünglichen. Als Dinge im Sinne des Internets der Dinge, können alle physikalischen und virtuellen Objekte bezeichnet werden, welche mit einem Kommunikationsnetzwerk verbunden werden können, z.B. Fernseher, Autos, Medizintechnik, Roboter oder Produktionsmaschinen.<sup>126</sup>



**Abbildung 17: IoT-Kommunikationsmöglichkeiten<sup>127</sup>**

Die IoT-Technologie soll damit nicht nur die Kommunikation zwischen Maschinen und Personen (M2P), sondern auch die direkte Kommunikation zwischen zwei oder mehreren Maschinen ermöglichen (Abbildung 17). Wesentliche Merkmale dieser Technologie sind der Wegfall jeglicher Beschränkungen wer oder was mit wem oder was kommunizieren kann bzw. wann und wo diese Kommunikation stattfindet kann. Charakteristika dieser neuen Technologie sind<sup>128</sup>:

- **Verbindungsfähigkeit:** Jede Art von Objekt kann Teil von einer Kommunikationsinfrastruktur werden.
- **Verschiedenartigkeit:** Die Geräte, welche durch das IoT verbunden werden, können sowohl virtuell als auch physikalisch sein. Ferner können diese verschiedenen Netzwerken zugehören wie auch verschiedene Soft- und Hardwareplattformen haben.

<sup>125</sup> Software welche speziell für eine bestimmte Hardware programmiert wurde.

<sup>126</sup> Vgl. Ernst & Young (2016), S. 2.

<sup>127</sup> Quelle: Zennaro, M. (2017), S.14.

<sup>128</sup> Vgl. Zennaro, M. (2017), S. 16 ff.

- Dynamische Änderungen: Die Geräte können ihren Zustand (Ruhezustand/ Wachzustand, Verbunden/ nicht Verbunden) jederzeit und selbstständig ändern. Weiters kann sich die Anzahl an Geräten im Netzwerk auch beliebig und selbstständig ändern.
- Gigantische Größenordnung: Die Anzahl der Geräte im Internet of Things wird ein Vielfaches der Anzahl der Geräte im derzeitigen Internet betragen.

Möglich machen sollen diese Art der Kommunikation die RFID-Technologie (Radio-frequency identification system), welche das berührungslose Identifizieren und Lokalisieren von Objekten und Lebewesen durch elektromagnetische Wellen ermöglicht.

<sup>129</sup>

Ende 2020 soll es demnach so schon über 100 Mrd. vernetzter Dinge geben<sup>130</sup>. Die Einsatzmöglichkeiten dieser Technologie sind nahezu unendlich. So erlaubt diese Technologie die präzise Nachverfolgung von Waren in einer Produktion, außerdem kann durch den RFID-Chip auch automatisch der nächste Produktionsschritt eingeleitet werden oder falls notwendig automatisch Änderungen am Prozess vorgenommen werden. Andererseits kann dieser Chip auch als Diebstahlschutz oder zur automatischen Verrechnung von Waren in einem Supermarkt eingesetzt werden. Aber auch im Heimgebrauch sind Lösungen denkbar die z.B. automatisch die Fenster schließen, wenn der Bewohner das Haus verlässt oder die Beleuchtung im Haus nur in jenen Räumen stattfindet wo gerade Personen anwesend sind. Ein weiterer in Zukunft sehr wichtiger Einsatzbereich dieser Technologie liegt in der Steuerung eines Netzwerkes von selbstfahrenden Autos, welche miteinander durch RFID-Chips kommunizieren.

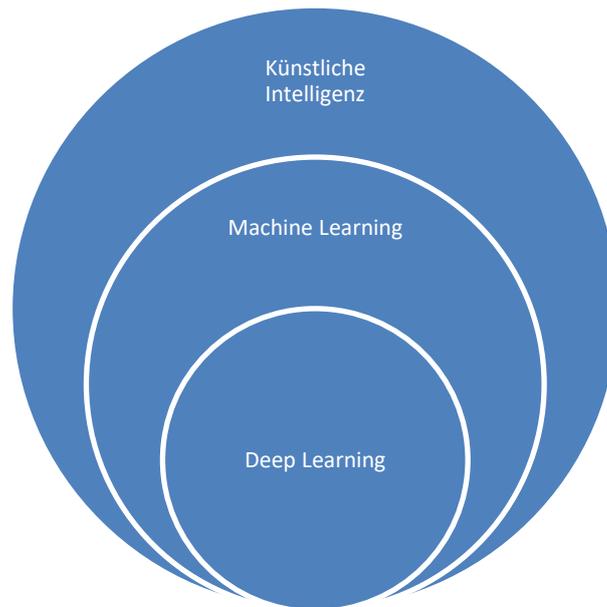
#### 4.4.2 Machine Learning

Machine Learning ist wie auch Data Mining eine Methode der Datenwissenschaft. Machine Learning Technologien verwenden oftmals Data Mining Analysen welche Korrelationen und Beziehungen in Daten finden. Der Unterschied liegt darin, dass beim Machine Learning von der gemachten Analyse automatisch Schlüsse gezogen werden und so die nächste durchgeführte Analyse noch genauer sein sollte. Ein gutes Beispiel hierfür sind die Kaufempfehlungen bei Amazon, die mit fortschreitender Zeit immer besser auf die Bedürfnisse der jeweiligen Person abgestimmt sein sollten. Im Gegensatz dazu wird das durch die Durchführung von Data Mining Prozessen gesammelte Wissen ohne die Interaktion von Menschen nicht für Folgeanalysen verwendet.

---

<sup>129</sup> Quelle: Produktion, <https://www.produktion.de/technik/was-rfid-unverzichtbar-macht-und-wo-it-gefahren-lauern-101.html> (Zugriff: 27.10.2019).

<sup>130</sup> Volker, P. A.; Hänisch, T. (2015), S. 9.



**Abbildung 18: Unterbereiche Künstlicher Intelligenz<sup>131</sup>**

Machine Learning ist eine Art von Künstlicher Intelligenz (KI) von Maschinen, welche durch einen der folgenden drei Faktoren definiert werden kann<sup>132</sup>:

- Schlussfolgerungen: Maschinen ist es möglich aufgrund der vorhandenen Daten Schlussfolgerungen zu ziehen, die es ermöglichen unvollständige Daten zu vervollständigen. Ein Beispiel für diesen KI-Faktor ist die Sprachassistenten von Google oder Apple, die bei Fragen Schlussfolgerungen ziehen können auch wenn diese auf eine bestimmte Frage des Benutzers nicht explizit trainiert worden sind.
- Natural Language Processing (NLP): Unter NLP wird die Fähigkeit von Maschinen verstanden Text oder menschliche Sprache zu interpretieren.
- Planung: Unter automatisierter Planung wird die Abweichung von einem vorprogrammierten Schema, welches die Tätigkeiten in einer vordefinierten Reihenfolge erledigt verstanden. Systeme reagieren auf die Bedingungen und weichen von einem Schema ab um ein vordefiniertes Endziel zu erreichen.

Es wird zwischen drei Kategorien von Maschine Learning unterschieden <sup>133</sup>:

- Unüberwachtes Lernen: Das Ergebnis, dass die Maschine liefern soll, ist hier nicht schon vor der Analyse bekannt. Der Algorithmus kann also nicht trainiert werden indem die Ergebnisse vorgeben werden. Ziel ist es die Daten in für Menschen sinnvolle und verwertbare Informationen umzuwandeln (oftmals passiert dies mit einer Datenvisualisierung). Die am häufigsten eingesetzten Arten des unbewachten Lernens sind die automatische Segmentierung (Clustering) und die Dimensionsreduktion, welche die Dimension der Daten reduziert, aber die wesentlichen und relevanten Informationen bewahrt. Ein

---

<sup>131</sup> Quelle: Eigene Darstellung.

<sup>132</sup> Vgl. Hurwitz, J.; Kirsch, D. (2018), S. 13 f.

<sup>133</sup> Vgl. Hurwitz, J., Kirsch, D. (2018), S. 15 ff.

Unterkategorie davon ist das teilüberwachte Lernen, bei dem für einen Teil der Eingaben schon die Ergebnisse bekannt sind anhand der Algorithmus lernen kann.

- Überwachtes Lernen: Der Algorithmus lernt hier anhand von Trainingsdaten wie das Ergebnis ausschauen soll. Es ist hierbei schon bekannt wie das Ergebnis ausschaut, der Algorithmus hat nur die Aufgabe die Lücke zwischen den Inputdaten und dem vorgegebenen Ergebnis zu schließen. Dieses Verfahren wird zur Regression und Klassifikation von Daten verwendet.
- Bestärktes Lernen: Ist eigentlich eine Sonderform des überwachten Lernens, bei dem das Ergebnis zwar nicht exakt vorgegeben ist, aber der Anwender genau weiß welcher Trend gut und welcher schlecht ist. Jedes gelieferte Ergebnis wird somit vom Benutzer bewertet ob es in die richtige Richtung geht. Die Ergebnisse, nach denen die Maschine lernt werden hier also schrittweise erarbeitet.

Wie in Abbildung 18 erkennbar ist Maschine Learning ein Unterbereich von Künstlicher Intelligenz, welcher wiederum Deep Learning als Unterbereich hat. Deep Learning hilft bei der Erkennung von Mustern in unstrukturierten Daten wie Bildern oder menschlicher Sprache. Dadurch wird es Maschinen möglich Klassifikationsaufgaben durchzuführen. Beispiel hierfür ist z.B. die Erkennung von Fußgängern bei automatisierten Autos.

## 4.5 Probleme und Herausforderungen

Die Probleme und Herausforderungen mit modernen Entwicklungen des Datenmanagements lassen sich grob in drei Gruppen aufteilen, einerseits gibt es wie bei der Einführung jeder neuen Technologie trotz bereits laufender Projekte in allen oben beschriebenen Bereichen noch offene technologische Themen in die noch Forschungstätigkeiten durchgeführt werden sollten. Durch den stetigen Anstieg der Datenmengen und den sich daraus neu entwickelnden Möglichkeiten am Markt aufzutreten wird dieser Punkt stetig vorangetrieben. Probleme gibt es hier aufgrund der Neuheit vieler dieser Technologien noch im Reifegrad hinsichtlich der Stabilität dieser Anwendungen und in der Angst vor ungeklärten Fragen bezüglich Datensicherheit. Andererseits gibt es noch die inneren und äußeren Probleme des Unternehmens selbst im Zusammenhang mit großen Investitionen in Industrie 4.0 Projekte. Im ersten Schritt muss von allen Verantwortlichen ein Verständnis für die Nutzung von Big Data aufgebaut werden, um die Vorteile und Optimierungsmöglichkeiten dieser neuen Technologien überhaupt verstehen zu können. In der zweiten Stufe braucht es einen Fokus vom Top-Management welches schlussendlich dieses Projekt auch freigeben muss, hierzu ist es von Vorteil einen belastbaren Businesscase zu haben, da in der heutigen Wirtschaftswelt Projekte ohne schnelle monetäre Vorteile sehr schwer zu rechtfertigen und zu realisieren sind. Ein Big-Data Projekt sollte daher von Anfang an das Ziel haben einen wirtschaftlichen Vorteil generieren zu können und einer Wirtschaftlichkeitsprüfung unterzogen werden. Dieser Punkt ist für jedes zu realisierende Projekt relevant, aufgrund der hohen Investitionssummen die in diesem Bereich notwendig sind ist es bei Big Data Projekten jedoch besonders schwierig einen klaren wirtschaftlichen Nutzen darzustellen, weil die Quantifizierung der Potentiale sehr komplex ist. Unternehmen welche über die

anfängliche Hürde der Investition kommen, sollten ihre Mitarbeiter schon sehr früh auf die neuen Herausforderungen mit der Technologie vorbereitet für welche komplett andere Fähigkeiten und Fertigkeiten notwendig sind wie bisher (z.B. wird der Bedarf an Softwareentwicklern und Datenanalysten in den nächsten Jahren weiterhin stark ansteigen). Äußere Herausforderungen kommen vom Staat und der Politik, welche dafür Sorge tragen müssen, dass die rechtliche Situation im Umgang mit den Daten geklärt ist. Hier sollten genaue Rahmenbedingungen zur Verwendung von externen Daten definiert werden, da sonst persönliche Daten leicht missbraucht werden können. Ein anderer wichtiger Punkt für den Staat ist der Ausbau der Basistechnologien wie z.B. die Gewährleistung eines flächendeckenden Breitbandanschlusses. Da es sich meist um sehr neuartige Technologien handelt, sollten zur Vertrauensbildung der Unternehmen auch möglichst schnell einheitliche Zertifizierungsmöglichkeiten und Standards geschaffen werden<sup>134</sup>.

Unternehmen, welche den Wandel nicht verschlafen wollen, sollten dem Datenmanagement einen hohen Stellenwert in der Unternehmenshierarchie einräumen und in die Technologien und die Verwaltung dieser investieren. Die nachfolgende Umfrage wird den aktuellen Stand des Datenmanagements in österreichischen Industrieunternehmen erarbeiten und sodass ein Vergleich mit anderen Ländern möglich ist.

---

<sup>134</sup> Vgl. PwC (2014), S. 36 ff.

## 5 Fragebogenerhebung

Dieses Kapitel beschäftigt sich mit der Forschungsfrage, der Konstruktion und der Durchführung der Umfrage, welche im Zuge dieser Masterarbeit durchgeführt wurde. Die Themenblöcke bzw. die einzelnen Fragen der Fragebogenerhebung ergeben sich aus den Theoriekapiteln oder leiten sich aus der Zielsetzung des Praxisteils ab. Im Unterpunkt der Fragebogenkonstruktion wird neben dem Aufbau des Fragebogens auch überblicksmäßig auf die Theorie zur Erstellung eines wissenschaftlichen Fragebogens eingegangen.

### 5.1 Anlass und Ziele der Fragebogenerhebung

Zum Thema Datenmanagement und Industrial Analytics gibt es schon einige detaillierte Studien, welche sich mit dem Wandel und den Einsatzmöglichkeiten der neuen Technologien beschäftigen. Größere durchgeführte Studien in diesem Bereich sind die BARC Studien Big Data Analytics<sup>135</sup> und Datenmanagement im Wandel<sup>136</sup> aus dem Jahr 2014, der Industrial Analytics Report 2016/2017<sup>137</sup>, die biMa Studie 2017/2018<sup>138</sup> und der 2018 Global Data Management Benchmark Report<sup>139</sup>. In der schnelllebigen IT-Branche spiegelt eine Umfrage aus dem Jahr 2014 jedoch nicht mehr den neuesten Stand zu diesem Thema wieder, außerdem wurden in diesen Studien die Länder aus welchen Unternehmen an diesen Umfragen teilnehmen dürfen nicht beschränkt oder auf den DACH-Raum beschränkt. Diese Vorgehensweise ist zwar gut um eine möglichst hohe Anzahl an Teilnehmern aus unterschiedlichen Branchen zu erreichen und so eine repräsentative Umfrage zu veröffentlichen, jedoch werden bei dieser Vorgangsweise nicht regionale Unterschiede in der Handhabung von Daten aufgezeigt. Unterschiede, welche sich durch den Wirtschaftssektor ergeben können, ebenfalls nicht untersucht werden. Es finden sich auch Umfragen die auf andere Länder beschränkt sind oder die das Datenmanagement in einem bestimmten Fachbereich untersuchen (z.B. Studie zum Datenmanagement bei klinischen Studien<sup>140</sup>). Die gegenständliche Umfrage soll sich nur an Industrieunternehmen in Österreich richten, um hier wirklich den Stand des Datenmanagement dieser Branche ausarbeiten zu können. Als Industrieunternehmen wurden jene Unternehmen klassifiziert, welche dem primären und sekundären Wirtschaftssektor zugehören. Also jene Unternehmen, die entweder Rohstoffe fördern oder diese weiterverarbeiten. Der tertiäre (Dienstleistungssektor) und der quartäre (Informationssektor) Wirtschaftszweig wurden für diese Umfrage nicht berücksichtigt,

---

<sup>135</sup> BARC GmbH (2014).

<sup>136</sup> BARC GmbH (2014).

<sup>137</sup> IoT Analytics (2016).

<sup>138</sup> BARC GmbH (2018).

<sup>139</sup> Experian (2018).

<sup>140</sup> Schumacher M.; Schulgen-Kristiansen, G. (2008).

darüber hinaus wurden keine Handelsunternehmen berücksichtigt. Ziel der Umfrage war es den aktuellen Stand des Datenmanagements in österreichischen Industrieunternehmen zu erfragen. Hierzu zählen sowohl die Handhabung der Daten, Zuständigkeiten, Probleme, Umgang mit Big Data wie auch die Nutzung von Industrial Analytics Methoden. Um aussagekräftige Erkenntnisse treffen zu können wurde definiert, dass der Fragebogen mindestens 40 Rückmeldungen benötigt.

## 5.2 Konzept des Fragebogens

Der Fragebogen wurde mit dem Online-Werkzeug für Internetumfragen von SurveyMonkey<sup>141</sup> erstellt. Da in der kostenlos verfügbaren Version dieses Online-Tools nur eine bestimmte Anzahl von Fragen definiert werden kann, wurde dafür die Lizenz des WBW-Instituts der Montanuniversität Leoben verwendet. Es handelt sich dabei um eine Website, auf der sehr schnell und einfach Fragebögen erstellt und ausgewertet werden können. Die dann mittels eines Hyperlinks aufgerufen werden können. Die Verbreitung dieses Fragebogens, wie auch das Aufrufen ist also sehr leicht möglich. Es handelt sich hierbei um eine standardisierte und anonyme Befragung, bei der sowohl die Fragen, wie auch die Antworten vorgegeben werden. Bei einer Diskussion oder einem Interview wäre dies anders, diese Art der Befragung wird nicht standardisiert genannt. Neben den Vorteilen gibt es auch einige Nachteile von Onlinefragebögen im Gegensatz zu Interviews, so ist keine Diskussion möglich, die aufgrund der Antworten des Befragten eine tiefgreifendere Analyse erlauben würde oder Missverständnisse aufklären könnte. Hingegen wird durch den Wegfall eines Interviewers oder eines Diskussionsführers eine Beeinflussung vermieden.

Durch den Modus und die Anonymität kommt es auch allgemein häufiger zu Abbrüchen als bei persönlich durchgeführten Interviews. Hinzu kommt, dass aufgrund von Zeitmangel und Angst vor Rechtschreibfehlern bei offenen Fragen oft nur sehr kurze Antworten gegeben oder gar der Fragebogen komplett abgebrochen.<sup>142</sup>

Um diesen Nachteilen entgegenzuwirken wurde der Fragebogen bewusst so gestaltet, dass dieser in unter 15 Minuten beantwortet werden kann und keine offenen Fragen enthält. Auch wurden bewusst kein Fragen als Pflichtfragen definiert, da dies ebenfalls zu einer höheren Dropout-Rate führen könnte. Es wurde das voreingestellte Standarddesign von SurveyMonkey verwendet und zur besseren Übersicht des Fortschrittes wurde ein Fortschrittsbalken eingefügt, der auf jeder Seite des Fragebogens angezeigt wird. Auch persönliche Meinungen oder Hypothesen wurden bewusst vermieden. Überdies wurde versucht die Fragen und Antworten möglichst so aufzubauen, dass keine Beeinflussung des Befragten stattfindet. Die Fragen wurden deshalb so gut es geht so gestellt, dass keine Beurteilung in der Fragestellung erkennbar ist. Die Antworten sind je nach Frage entweder so gewählt, dass nur eine ausgewählt werden kann (Single Choice) oder dass mehrere Antworten möglich sind (Multiple Choice). Damit es hier zu keiner Verwirrung kommt, wurde die Möglichkeit mehrerer

---

<sup>141</sup> [www.surveymonkey.com](http://www.surveymonkey.com)

<sup>142</sup> Vgl. Berger, D. (2010), S. 156 ff.

Antworten, in der Fragestellung mit dem Zusatz „Mehrfachantwort möglich“ gekennzeichnet. Bei Fragen bei welchen dies sinnvoll erschien wurde außerdem ein Textfeld für andere bzw. sonstige Antwortmöglichkeiten hinzugefügt. Jede Frage hat die Möglichkeit, dass der Teilnehmer diese auslassen kann. Der Fragebogen kann abgeschickt werden, wenn mindestens eine Frage beantwortet wurde und der Benutzer sich durch alle Frage geklickt hat.

Am Anfang der Umfrage wurde eine Einleitungsseite verfasst, welche dem Teilnehmer den Grund und die Ziele der Umfrage näherbringen soll. Ebenfalls wird hier kurz auf die behandelten Themen und die ungefähre Dauer des Fragebogens hingewiesen. Außerdem soll diese vertrauen schüren und einen potentiellen Teilnehmer werben und dazu beitragen, dass die Dropout-Quote möglichst gering bleibt. Angegeben wurden außerdem meine Kontaktdaten. Ein Absatz hat sich auch dem Thema Datenschutz und der vertraulichen Verarbeitung der Daten gewidmet und darauf hingewiesen, dass Fragebogen komplett anonym abläuft und bis auf den Unternehmensbereich des Teilnehmers keine persönlichen Daten erfragt werden. Überdies wurde auf die Einhaltung der Datenschutzgrundverordnung (DGSVO) hingewiesen und auf die Datenschutzrichtlinie von Survey Monkey<sup>143</sup>. Es wurde außerdem auf die Möglichkeit hingewiesen, dass die Ergebnisse dieser Umfrage, nach der Durchführung per Mail zugeschickt werden können. Durch Klick auf den Button „Weiter“ hat der Teilnehmer diese Richtlinien akzeptiert und startet die Umfrage. Der vollständige Einleitungstext befindet sich in der Anlage. Auf der letzten Seite des Fragebogens wird dem Teilnehmer, dann für die Teilnahme gedankt und der Teilnehmer über die erfolgreiche Teilnahme informiert.

Insgesamt besteht der Fragebogen aus insgesamt 25 Fragen. Frage 1-5 sind einleitende Fragen, die dazu dienen den Befragten trotz gegebener Anonymität klassifizieren zu können. Die Fragen 6-13 beschäftigen sich mit dem organisatorischen Teil des Datenmanagements, während die Fragen 14-17 die technische Handhabung abfragen. Die Fragen 18 bis 25 beschäftigen sich mit Big Data und dem aktuellen Stand und zukünftige Projekte im Bereich von Data Analytics. Der komplette Fragebogen befindet sich im Anhang A1.

---

<sup>143</sup> Survey Monkey, <https://de.surveymonkey.com/mp/legal/privacy-policy/> (Zugriff: 10.08.2019).

### 5.3 Fragen und Antwortvorgaben

Der Fragebogen ist in fünf Themenbereiche aufgeteilt (Abbildung 19). Im nachfolgenden Kapitel werden die einzelnen Fragen beschrieben, der Bezug zu den jeweiligen Kapiteln in der Theorie gezeigt und wie diese zur Beantwortung der Zielsetzungen des Fragebogens beitragen.

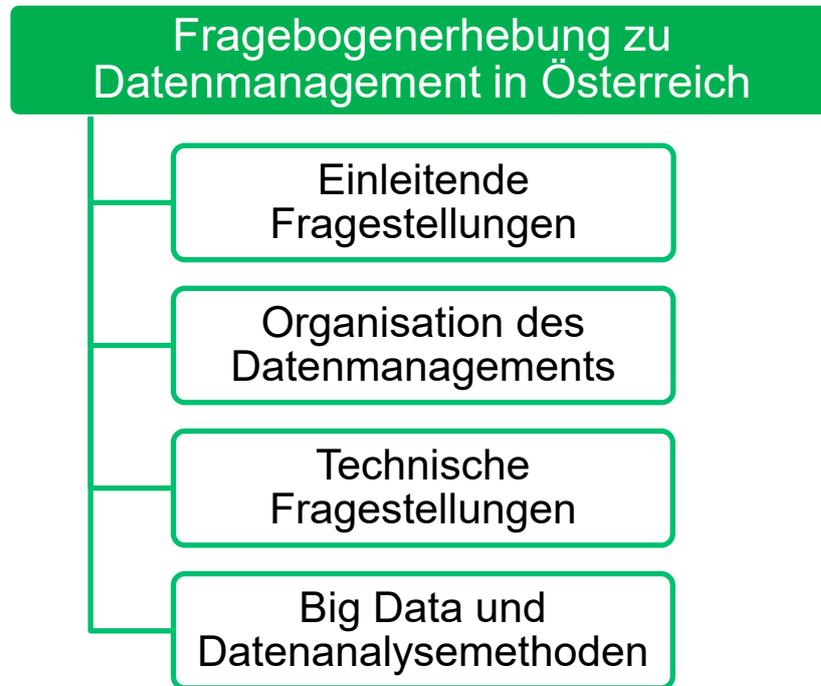


Abbildung 19: Aufbau der Fragebogenerhebung<sup>144</sup>

- Einleitende Fragestellungen

Diese sollen Rückschlüsse darauf erlauben wie sich bestimmte Unternehmenskennzahlen wie z.B. die Branche oder Firmengröße auf die Beantwortung der restlichen Fragen und auf das Gesamtergebnis der Umfrage auswirken. Aus diesem Grund wurde die Befragung mit fünf dieser Fragen gestartet. Eine potentielle Korrelation könnte lauten, dass größere Unternehmen schon komplexere Datenanalysemethoden verwenden oder, dass in bestimmten Branchen, der Umgang mit Daten sehr genau geregelt ist, während es im allgemeinen Durchschnitt eine andere Antwort gab.

In Frage 1 wurde nach dem Sektor des Unternehmens gefragt, die Antwortmöglichkeiten ergeben sich aus den definierten Industriesektoren der Wirtschaftskammer Österreich (WKO)<sup>145</sup>. Es wurde zusätzlich ein Textfeld vorgegeben um Unternehmen welche in keinen der angegeben Sektoren passen, trotzdem eine Teilnahme zu ermöglichen. Falls sich bei der Auswertung ergibt, dass der Sektor falsch angegeben wurde, erfolgt eine manuelle Zuweisung zum richtigen Sektor. Bei Unternehmen welche als Nicht-Industrieunternehmen

<sup>144</sup> Quelle: Eigene Darstellung.

<sup>145</sup> WKO, <https://www.wko.at/branchen/Branchenauswahl.html> (Zugriff: 08.03.2019).

klassifiziert werden (z.B. Dienstleistungsunternehmen) werden im Nachhinein entfernt. Es werden nachfolgend auch alle weiteren Antworten dieses Teilnehmers entfernt, um so das Umfrageergebnis nicht zu verfälschen.

Bei Frage 2 ist die Anzahl der Mitarbeiter an allen Standorten des jeweiligen Unternehmens in Österreich anzugeben. Es wurde hier bewusst nicht nach allen Mitarbeitern des Unternehmens gefragt, welche weltweit bei diesem Unternehmen beschäftigt sind, da dies das Ergebnis massiv verfälschen könnte, falls z.B. ein sonst großer Konzern in Wien nur ein kleines Büro führt. Dieselbe Vorgehensweise wurde auch bei der Frage nach dem Umsatz (Frage 3) angewendet. Frage 4 ist eine Selbsteinschätzung ob, dass Unternehmen umsatzmäßig besser oder schlechter abgeschnitten hat als der direkte Wettbewerb, hierdurch sollen Aussagen getroffen werden ob eine Abhängigkeit zwischen dem Umsatz und einem guten Umgang mit dem Thema Datenmanagement festgestellt werden kann. Frage 5 betrachtet die Herkunft der Studienteilnehmer nach Unternehmensbereichen.

- Organisation des Datenmanagements

In diesem Teil des Fragebogens wird die interne Organisation des Datenmanagements abgefragt, die genaue Definition der Zuständigkeiten und der Datenqualität etc. ist wie in Abschnitt 3.3 und 3.5 beschrieben ein zentraler Bestandteil eines funktionierenden Datenmanagements. Frage 6 beschäftigt sich mit der Zuständigkeit, hierbei wurden auch Stellen als Auswahlmöglichkeit angegeben, welche sicherlich in kleineren Unternehmen nicht definiert sind (z.B. CIO und CDO), außerdem steht auch hier ein Textfeld für die Angabe einer nicht erwähnten Antwort zur Verfügung. Frage 7 beschäftigt sich mit der Implementierung eines unternehmensweiten Datenmanagement, beschrieben wird die erfolgreiche Implementierung in Abschnitt 3.1 der Theorie. Damit es für den Teilnehmer klar ist, was in diesem Fall konkret mit Datenmanagement gemeint ist, wurde vor dieser Frage folgende Definition eingefügt:

Datenmanagement ist eine Unternehmensfunktion welche sich mit der Entwicklung, Implementierung und Wartung von Regeln und Abläufen beschäftigt, welche den Wert von Daten und Informationen kontrolliert, bzw. steigert und diese Daten schlussendlich zu den richtigen Personen liefert.<sup>146</sup>

Interessant ist hierbei vor allem wie viele Unternehmen dieses bereits in allen Fachbereichen implementiert haben. Auch hier wird nach der persönlichen Einschätzung gefragt ob das Unternehmen bezugnehmend auf die letzte Frage besser, gleich oder schlechter als der direkte Wettbewerb dasteht (Frage 8). Im Anschluss an diese Frage wird nach dem/den Faktor/en gefragt bei welchem das Unternehmen den höchsten Mehrwert durch die Implementierung des Datenmanagements verspürt. Diese Frage soll das persönliche Empfinden klären und dieses mit Meinungen aus der Literatur vergleichen, wo Datenmanagement eigentlich den größten Effekt in Unternehmen zeigen sollte.

---

<sup>146</sup> Vgl. Mosley, M.; Brackett, M. et al. (2009), S.4 ff.

Hier wurden die Antwortmöglichkeiten bewusst in der Anzahl beschränkt, weil sonst mit sehr vielen unterschiedlichen Antworten zu rechnen wäre, welche sich nur schwer vergleichen lassen. Die Missionen des Datenmanagements sind in Abschnitt 3.1 der Theorie beschrieben. Die nächste Frage (Frage 10) beschäftigt sich mit den Problemen in Zusammenhang mit Datenmanagement, hier wurden als Antwortmöglichkeiten menschliche und technische Faktoren berücksichtigt, welche sich aus Kapitel 3 ableiten lassen.

Um herauszufinden in wie weit operative und strategische Entscheidungen auf Grundlage von Daten getroffen werden wurde dies in Frage 11 und 12 behandelt. Als Auswahlmöglichkeit gab es bei dieser Frage nur JA oder NEIN, um hier ein klares Bekenntnis der Teilnehmer zu forcieren. Ziel dieses Themenbereiches ist es zu klären in wie weit ein Datenmanagement bereits in den Unternehmensprozess integriert worden ist, wie die Zufriedenheit der Teilnehmer mit diesem ist und wo die größten Vorteile bzw. Nachteile von diesem gesehen werden.

- Technische Fragestellungen

Die erste Frage dieses Fragebogenabschnitts beschäftigt sich mit der Standardisierung von Auswerteperspektiven und Key Performance Indicators (KPIs). Dieses Punkt ist ein zentrales Thema des Datenmanagements, da die Sicherung der Datenqualität sowohl organisatorisch (siehe organisatorische Fragestellungen des Fragebogens) wie auch technisch sicherlich zu den größten Herausforderungen zählt. Um interne wie auch externe Daten überhaupt nutzbar zu machen, müssen diese im Einklang mit den Dimensionen der Datenqualität (Tabelle 1) sein und daher zwingend standardisiert werden. Dieser Schritt erfordert neben einer technischen Lösung auch organisatorische Maßnahmen, wie z.B. Berichte gestaltet sein müssen, oder wie die Qualität von einzelnen Daten festgestellt werden muss. Hierbei muss aber ebenfalls zwingend auf die Führung der Flexibilität und Agilität geachtet werden damit komplexere Datenanalysen überhaupt in der geforderten Geschwindigkeit möglich sind. Wie schon im Abschnitt zum Thema Big Data erwähnt, muss die Datenqualität auch bei sehr schnell verbreiteten Informationen, welche eigenständig und weitgehend unabhängig von der IT-Abteilung verbreitet werden, unbedingt sichergestellt werden. Das Self-Service Thema wird ebenfalls bei der nächsten Frage thematisiert, wo nach dem derzeitigen Schwerpunkt im Unternehmen gefragt wird. Hier kann dieser Punkt unter anderem auch ausgewählt werden, da dieses Thema sehr breit diskutiert werden kann, wurde dieser Punkt bewusst so gestaltet, dass abseits der fix vorgegebenen Antwortmöglichkeiten, welche eher technisch orientiert sind, ebenfalls andere Antworten gegeben werden können. Die Antwortmöglichkeiten dieser wie auch der nächsten Frage ergeben sich aus den aktuellen Trends des Datenmanagements welche in Kapitel 4 beschrieben sind. Die hier angegebenen Schwerpunkte sind für die Zielsetzung des Theorieteils sehr wichtig, weil diese sowohl über derzeitige Schwächen wie auch zukünftige Schwerpunkte der Industrieunternehmen in Österreich Aufschluss geben können. Bei Frage 15 wird nach den Faktoren gefragt, wo der Teilnehmer den größten Aufholbedarf im Unternehmen sieht, um hier gewisse Trends ableiten zu können, wurde diese Frage bewusst ohne die Möglichkeit einer

sonstigen Antwort gestaltet. Eine Mehrfachauswahl war jedoch möglich. Diese Frage soll auch Aufschluss darüber geben, ob derzeitige Schwächen der Unternehmung erkannt worden sind und bei Frage 14 als derzeitige Schwerpunkte definiert worden sind. Mit Frage 16 wird ebenfalls versucht einen Aufholbedarf, diesmal jedoch zum Thema Datenqualität herauszufinden. Auch hier wurden die Antwortmöglichkeiten beschränkt, als Antwortmöglichkeiten stehen die in Abschnitt 3.5 behandelten Dimensionen der Datenqualität zur Auswahl. Die letzte Frage dieses Themenblocks behandelt die verwendete Datenarchitektur, als Grundlage hierfür dienen die in Kapitel 4 behandelten modernen Architekturarten. Interessant wird bei der Auswertung dieser Frage, ob ein direkter Zusammenhang zwischen der Unternehmensgröße, der Speicherung der Daten im Unternehmen und den angewendeten Datenanalysemethoden besteht.

- Schwerpunkt Big Data und Datenanalysemethoden

In den beiden folgenden Fragen soll erfasst werden, welchen Stellenwert Big Data Projekte, wie sie in Kapitel 4 beschrieben werden, bereits in Unternehmen haben, hier wird vor allem die Annahme zu Überprüfen sein, ob die Branchenführer tatsächlich führend sind und in welchen Bereichen Unternehmen, das meiste Potential für Big Data Projekte sehen. Da dieses Thema in einigen Feldern erst kürzlich eingeführt, wird es somit interessant sein ob hier Trends abgeleitet werden können, wo Zukunftsmärkte liegen und ob die Technologie vermutlich schon bald verstärkt eingesetzt wird.

Frage 20 behandelt das Thema der Datenanalyse in einem höheren Detailgrad, es werden die Faktoren abgefragt, bei welchen die in Kapitel 4 der Theorie beschriebenen Datenanalyse laut Meinung des Teilnehmers den größten Vorteil bringen sollen. Die Antwortmöglichkeiten sind hierbei konkrete Vorteile für das Unternehmen wie zum Beispiel eine Umsatzsteigerung.

Wie in der Theorie beschrieben scheitern Big Data Projekte in Unternehmen derzeit noch sehr oft an den finanziellen Belastungen, der Datenqualität, den infrastrukturellen Grundlagen und dem hohen Bedarf an menschlichen Ressourcen bei der Planung. Die Hindernisse können hierbei sehr vielfältig sein und sowohl von organisatorischer wie auch technischer Natur sein. Die Fragen 21 und 22 sollen genau diese ermitteln.

Die nächsten beiden Fragen beschäftigen sich mit dem Thema der Datenanalyse, da diese Themen derzeit sehr eindringlich in der Fachpresse besprochen werden, ist das Ergebnis dieser Fragen sehr aufschlussreich für die Beantwortung der Zielsetzungen des Praxisteils welche Analysemethoden derzeit in Österreich wirklich schon genutzt werden und welche nicht, wie sich hier der Trend verhält und ob österreichische Industrieunternehmen bei den Analysemethoden eher Vorreiter oder Verfolger sind und bei komplexeren Analysen eher noch abwarten. Bei der letzten Frage werden die Datenquellen abgefragt, welche für Data Analytics Projekte genutzt wurden, es wird hier interessant sein auszuwerten in wie weit schon moderne Datenquellen wie RFID oder Social-Media-Daten genutzt werden und wie österreichische Unternehmen hier im Vergleich zu anderen Ländern dastehen.

## 5.4 Durchführung der Fragebogenerhebung

Nachdem die Fragen korrekturgelesen wurden, sind diese in den SurveyMonkey Account des WBW-Institutes der Montanuniversität Leoben eingegeben worden. Nachdem die Umfrage online war, wurde diese an drei Testpersonen für einen Testlauf verschickt. Diese haben die Umfrage mit unterschiedlichen Betriebssystemen getestet (Mobiltelefon mit iOS, Computer mit Windows, Computer mit macOS). Diese vorher instruierten Personen hatten den Auftrag den Fragebogen auf mögliche Fehler wie zum Beispiel: technische Schwierigkeiten, unklare Fragestellungen, mangelhafte Antwortoptionen, Rechtsschreib und Grammatikfehler zu prüfen. Es wurden nach diesem Testlauf keine weiteren Änderungen mehr durchgeführt. Die Antworten dieser Testläufe wurden vor dem eigentlichen Start der Umfrage gelöscht.

Seitens des WBW-Instituts wurde für die Umfrage ein E-Mail Verteiler zur Verfügung gestellt. Dieser wurde nach dem Erhalt noch sortiert, damit auch wirklich nur Industrieunternehmen an dieser Umfrage teilnehmen. Es wurden deshalb alle Dienstleister und Handelsbetriebe entfernt. Um die Teilnehmerzahl zu erhöhen wurde außerdem eine Recherche auf der Seite der WKO durchgeführt, bei welcher öffentlich verfügbare E-Mail Adressen von Industrieunternehmen rausgesucht wurden. Mit einigen E-Mail Adressen von Bekannten, welche in Industrieunternehmen tätig sind, konnten so insgesamt 463 E-Mail Adressen gesammelt werden. Ungefähr 90% davon wurden persönlich angeschrieben, weil hier der Name in der E-Mail Adresse enthalten war. Aufgrund des recht hohen Wertes an E-Mails mit persönlicher Anrede, wurde mit einer höheren Rücklauf Quote gerechnet, als in vergleichbaren Umfragen. Die Rücklaufquote bei Online-Umfragen liegt typischerweise bei ca. 10%.<sup>147</sup>

Der Link zur Umfrage wurde am 23.06.2019 verschickt. Eine weitere Erinnerungsmail wurde am 30.06.2019 an alle Personen verschickt, welche in dieser Zeit nicht geantwortet haben und keine Abwesenheitsnotiz hinterlassen haben. Diese Vorgehensweise sollte den Rücklauf nochmals steigern, da typischerweise eine Beantwortung der meisten Online-Fragebögen innerhalb weniger Tage erfolgt.<sup>148</sup>

Der Anteil jener Adressen, auf welche eine Abwesenheitsnotiz gekommen ist, war bedingt durch die Urlaubszeit recht hoch (ca. 7%). Hingegen war der Anteil jener, bei welchen die E-Mail Adresse nicht mehr gültig war, sehr niedrig (ca. 2%).

Am 07.07.2019 wurde die Umfrage geschlossen, weil genügend Ergebnisse vorlagen, um ein aussagekräftiges Ergebnis zu bekommen.

Am Ende gab es 148 Beantwortungen, bei einer Abschlussquote von 76 %. Eine Auflistung der Teilnehmer nach IP-Adressen befindet sich im Anhang A2. Die Rücklaufquote lag bei ca. 32 % und ist wie schon vermutet höher als der durchschnittliche Wert laut Literatur. Die Beendigungsquote bei Umfragen schwankt sehr stark, auch abhängig von der Umfrageform deshalb ist es hier schwer einen durchschnittlichen Wert anzugeben. Der erzielte Wert von 76 % bewegt sich aber klar in

---

<sup>147</sup> Quelle: Question Pro, <https://www.questionpro.de/responsequote-ruecklaufquote-online-umfrage/> (Zugriff: 21.08.2019).

<sup>148</sup> Vgl. Gräf (2010), S.58.

der Norm die zwischen 15 und 80 % liegt<sup>149</sup>. Dies ist wohl auch dem Format des Fragebogens ohne Pflichtfragen geschuldet.

Von den erhaltenen Beantwortungen wurden die Antworten von 7 Teilnehmern gelöscht, weil diese nicht von Industrieunternehmen stammen. Es gibt also effektiv 141 gültige Beantwortungen.

Nachstehend ist die jeweilige Anzahl der Beantwortungen und Nicht-Beantwortungen pro Frage angegeben:

**Tabelle 2: Anzahl der Beantwortungen und Nicht-Beantwortungen pro Frage<sup>150</sup>**

Frage	Beantwortungen	Nicht-Beantwortungen
1	137	4
2	136	5
3	130	11
4	130	11
5	137	4
6	117	24
7	115	26
8	114	27
9	111	30
10	111	30
11	111	30
12	112	29
13	112	29
14	110	31
15	110	31
16	105	36
17	106	35
18	106	35
19	102	39
20	105	36
21	104	37
22	105	37
23	101	40
24	101	40
25	101	40

<sup>149</sup> Vgl. Gräf (2010), S.65.

<sup>150</sup> Quelle: Eigene Darstellung.

Die tendenziell steigende Anzahl an Nicht-Beantwortungen mit fortlaufenden Fragenummern ist wohl dem steigenden technischen Anspruch der Fragen geschuldet. Jedoch konnten bei allen Fragen weitaus mehr Antworten gesammelt werden als den definierten Minimumwert von 40 Antworten. Es kann somit angenommen werden, dass genug Antworten verfügbar sind, um repräsentative Ergebnisse zu erhalten, welche in weiterer Folge zur Beantwortung der Forschungsfrage dienen.

## 6 Auswertung der Fragebogenerhebung

Im folgenden Kapitel werden die Ergebnisse der Umfrage ausgewertet. Die Gliederung dieser Auswertung erfolgt konform mit dem vier Themenblöcken des Fragebogens (Einleitende Fragestellung, Organisation, Technische Fragestellungen, Big Data und Datenanalysemethoden). Ziel ist es aus den Antworten Kenntnisse zu erlangen, die es erlauben Hypothesen zu den Ergebnissen aufzustellen und zur Beantwortung der Forschungsfrage herangezogen werden können.

Zunächst wurde die von SurveyMonkey erstellte CSV-Datei, um jene Antworten bereinigt, die von Teilnehmern stammen, welche nicht als Industrieunternehmen bezeichnet werden können. Danach mussten aufgrund der veränderten Teilnehmerzahl die Prozentsätze der Antworten neu berechnet und in weiterer Folge auch die Graphen neu erstellt werden. Falls einer der Beiträge in einem extra Feld wie z.B. bei „Sonstiges“ genannt worden ist, aber zu einer der vorgegebenen Antworten passt, dann wurde dieser Beitrag zur passenden Antwort verschoben. Die Prozentsätze sind immer relativ zur Anzahl der Leute berechnet, welche diese Frage beantwortet haben, Teilnehmer welche die Frage nicht beantwortet haben werden somit in der Berechnung nicht beachtet werden. Im Sinne einer ansprechenden Ergebnispräsentation die Ergebnisse in Graphen präsentiert. Auf eine exakte Nennung aller Ergebnisse wird verzichtet, einzelne interessante Werte werden jedoch herausgenommen und näher beschrieben.

## 6.1 Einleitende Fragestellungen

### Frage 1:

Zum Beginn der Befragung wurden die Teilnehmer gefragt in welcher Branche ihr Unternehmen hauptsächlich aktiv ist (Abbildung 20). Der Fragebogen wurde bewusst an Branchen aus allen Industriesektoren verschickt, um ein möglichst aussagekräftiges Gesamtergebnis über alle Industriesektoren hinweg zu bekommen. Dies ist auch größtenteils gelungen da der Mittelwert der Antworten bei 5,88% liegt. Der einzige Ausreißer ist die Metalltechnische Industrie, aus diesem Sektor gibt es 34 Teilnehmer, welche 24,82% der Gesamtteilnehmer der gegenständlichen Umfrage ausmachen.

Da dieser Sektor jedoch für ca. ein Viertel der gesamten Industrieleistung Österreichs bezogen auf die Unternehmensparameter Umsatz, Produktionswert und Bruttowertschöpfung verantwortlich ist und aus über 150 Großunternehmen besteht, ist die deutlich höhere Anzahl an Teilnehmer dadurch gut erklärbar.<sup>151</sup>

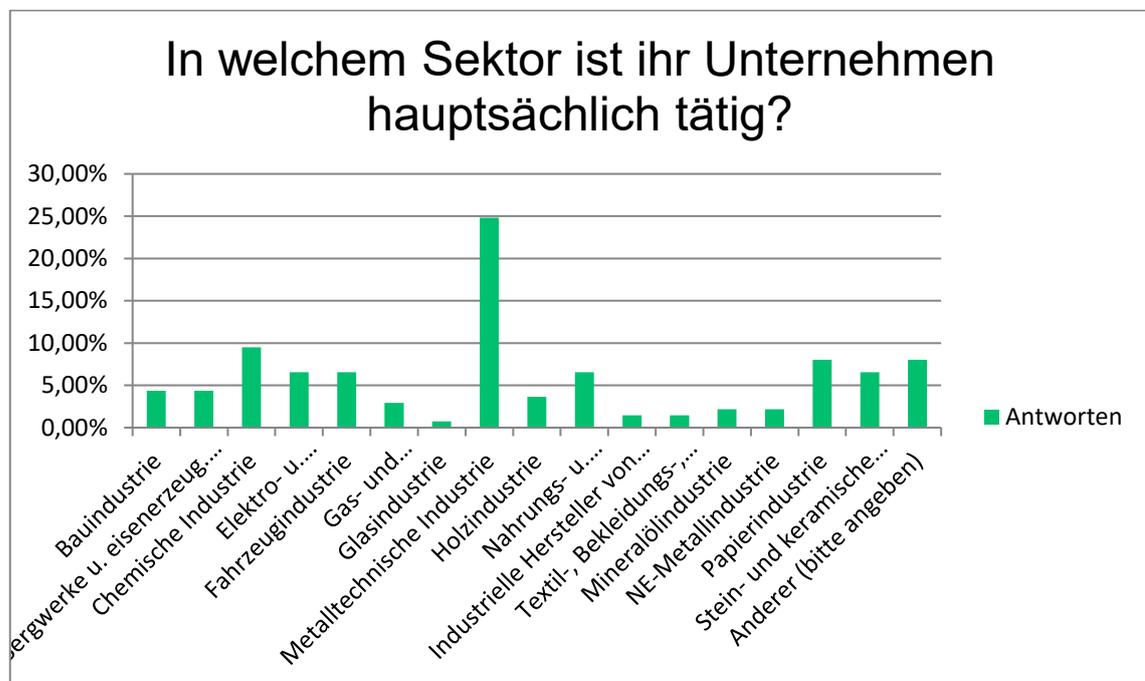


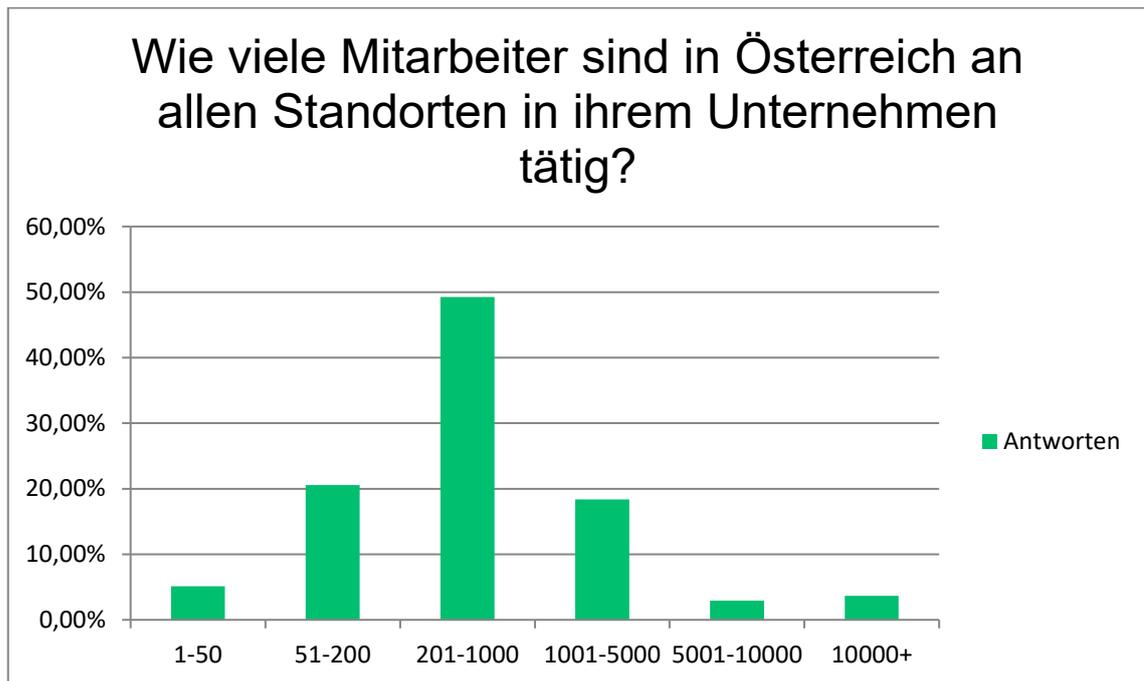
Abbildung 20: Frage 1: Sektor des Unternehmens<sup>152</sup>

<sup>151</sup> Quelle: Statista, <https://de.statista.com/themen/2179/metallindustrie-in-oesterreich/> (Zugriff: 30.08.2019).

<sup>152</sup> Quelle: Eigene Darstellung.

**Frage 2:**

Ergänzend dazu wurden die Teilnehmer gefragt wie viele Mitarbeiter an den Standorten in Österreich arbeiten (Abbildung 21). Die Fragestellung wurde bewusst so gewählt, dass nur die Mitarbeiter aus Österreich berücksichtigt werden, da es sonst zu Verfälschungen kommen könnte, zum Beispiel wenn ein großer Konzern nur eine kleine Niederlassung in Österreich betreibt. Die meisten Teilnehmer (49,26%) waren aus Unternehmen mit 201-1000 Mitarbeitern. Somit größtenteils aus Großunternehmen, welche laut WKO mindestens 250 Mitarbeiter beschäftigen müssen.<sup>153</sup> Gefolgt von Teilnehmern aus Unternehmen mit 51-200 (20,59%) und Teilnehmer aus Unternehmen mit 1001-5000 Mitarbeitern (18,38%).



**Abbildung 21: Frage 2: Mitarbeiter des Unternehmens<sup>154</sup>**

<sup>153</sup> Quelle: WKO, <https://news.wko.at/news/oesterreich/wirtschaftskraft-kmu2018.pdf> (Zugriff: 30.08.2019).

<sup>154</sup> Quelle: Eigene Darstellung.

**Frage 3:**

Auch die Frage nach dem Jahresumsatz wurde nur auf den Wert bezogen, welcher von den Unternehmen in Österreich erzielt worden ist. Auch bei dieser Fragestellung zeigt sich das größtenteils Großunternehmen (Umsatz >50 Millionen<sup>155</sup>) bei dieser Umfrage mitgemacht haben (Abbildung 22).

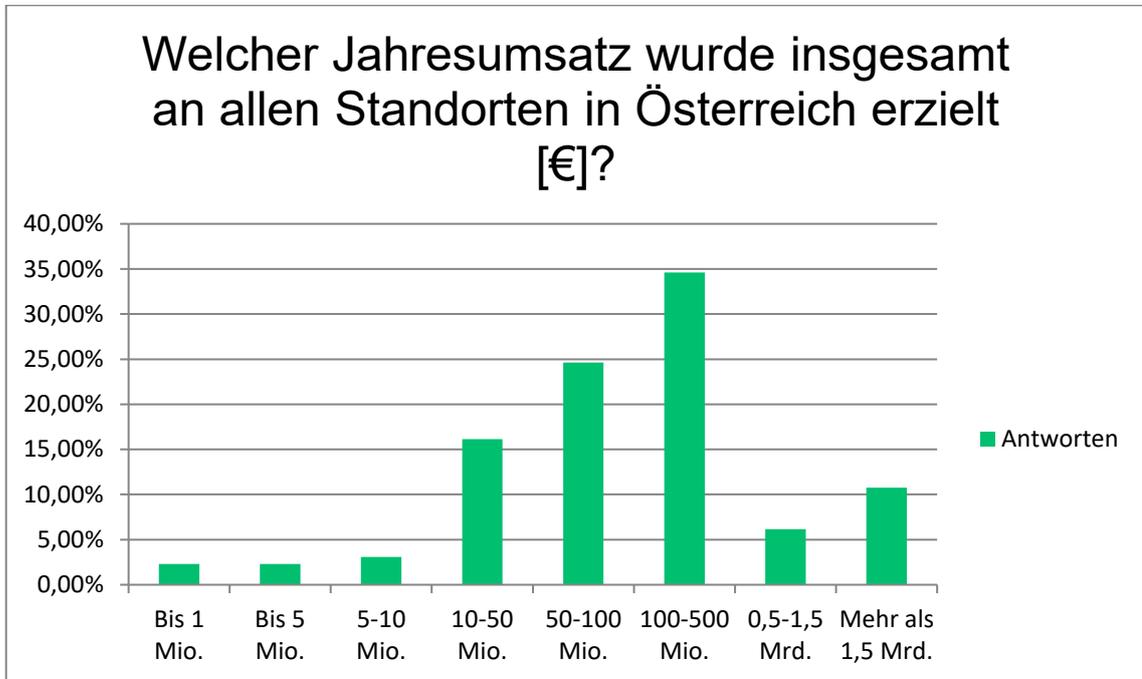


Abbildung 22: Frage 3: Jahresumsatz des Unternehmens<sup>156</sup>

<sup>155</sup> Quelle: [https://www.ffg.at/recht-finanzen/rechtliches\\_service\\_KMU](https://www.ffg.at/recht-finanzen/rechtliches_service_KMU) (Zugriff: 30.08.2019).

<sup>156</sup> Quelle: Eigene Darstellung.

**Frage 4:**

84,62% glauben, dass ihr Unternehmen mit dem oben angegebenen Umsatz besser abgeschnitten hat, als der direkte Wettbewerb, somit glauben nur 15,38% der Teilnehmer, dass ihr Unternehmen umsatzmäßig schlechter dasteht als der direkte Wettbewerb (Abbildung 23). Dieses Vertrauen ins eigene Unternehmen kann wohl auch damit erklärt werden, dass größtenteils Mitarbeiter von Großunternehmen an dieser Umfrage teilgenommen haben.

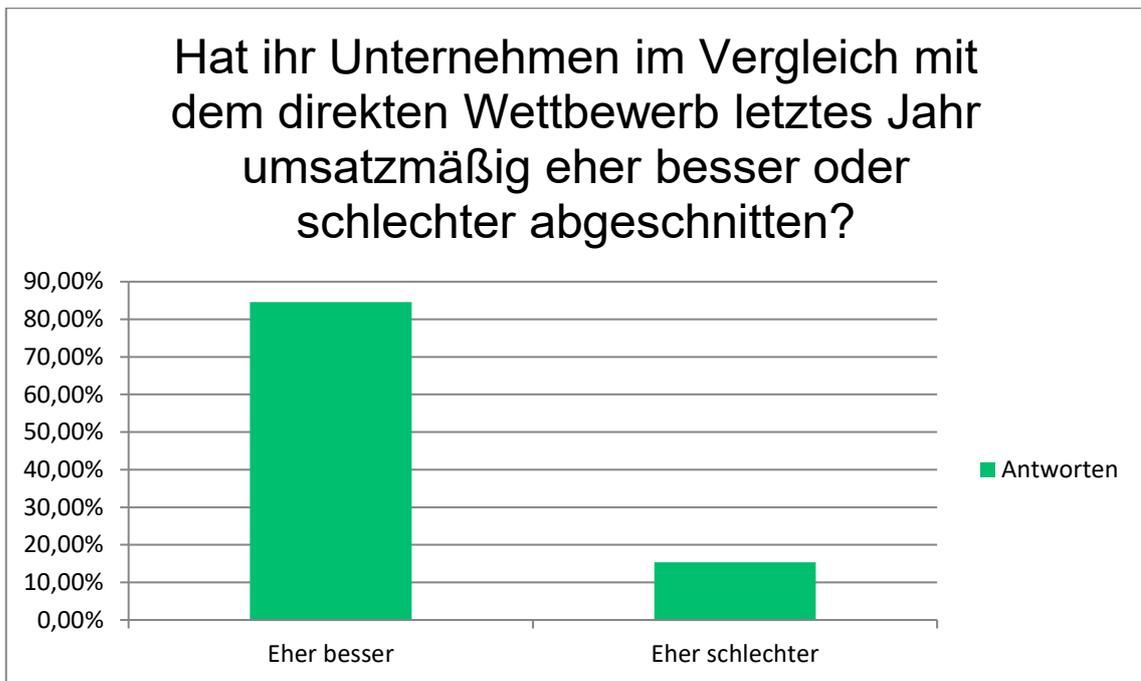


Abbildung 23: Frage 4: Umsatzmäßiger Vergleich<sup>157</sup>

<sup>157</sup> Quelle: Eigene Darstellung.

**Frage 5:**

Bei der Frage, nach dem Unternehmensbereich (Abbildung 24) zeigt sich eine klare Tendenz von Teilnehmern aus den Bereichen Geschäftsführung (20,44%), IT (18,98%) und Produktion (16,79%). Bei der Antwortmöglichkeit „Anderer“ welche in einem Textfeld eingetragen werden konnte, haben die Teilnehmer zumeist ihre offizielle Berufsbezeichnung im Unternehmen eingetragen. Größtenteils stammen diese Mitarbeiter aus dem technischen Bereich.



**Abbildung 24: Frage 5: Unternehmensbereich des Teilnehmers<sup>158</sup>**

<sup>158</sup> Quelle: Eigene Darstellung.

## 6.2 Fragen zur Organisation

### Frage 6:

Zu Beginn der Kernfragen dieser Umfrage, wurde nach dem Verantwortlichen für Daten innerhalb des Unternehmens gefragt (Abbildung 25). Die Topantwort unter den abgegebenen Antworten war mit 43,59% die IT-Abteilung. Bei 23,39% ist jede Fachabteilung selbst für die Daten verantwortlich und bei 13,68% gibt es einen eigens definierten Datenverantwortlichen. Alle anderen auswählbaren Antworten wurden von 6% oder weniger ausgewählt. Unter den anderen Antworten, die in einem Textfeld angegeben werden konnten, wurde entweder der CFO genannt oder es wurde angegeben, dass es verschiedene Verantwortliche für unterschiedliche Daten gibt.

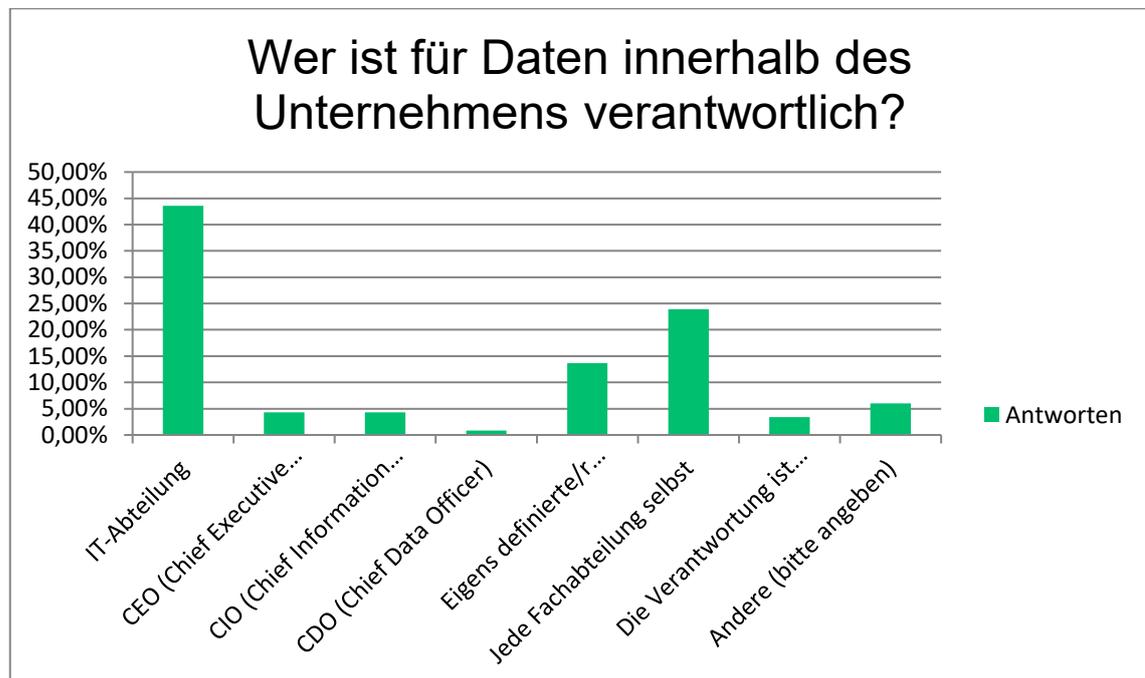


Abbildung 25: Frage 6: Verantwortung<sup>159</sup>

<sup>159</sup> Quelle: Eigene Darstellung.

**Frage 7:**

Bereits 77,05% der Befragten Unternehmen haben ein Datenmanagementsystem unternehmensweit oder zumindest in einzelnen Fachbereichen integriert. Nur 3,28%, was 4 Teilnehmern entspricht, haben angegeben, dass keine fixen Standards bestehen und auch in Zukunft nicht geplant sind (Abbildung 26).

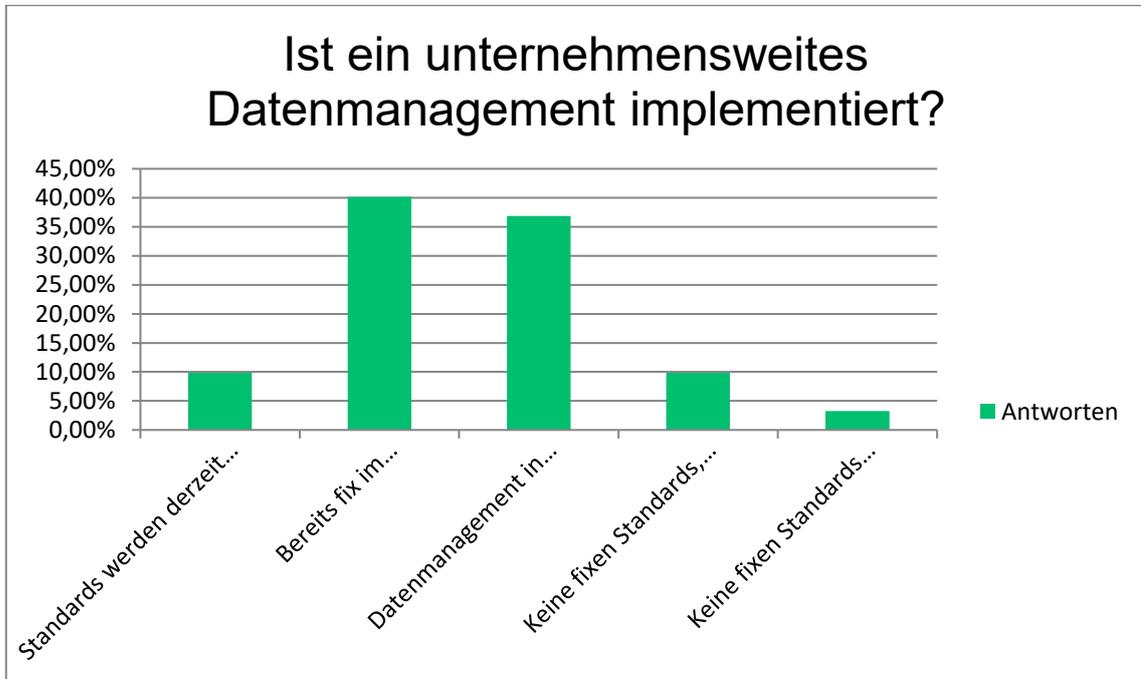
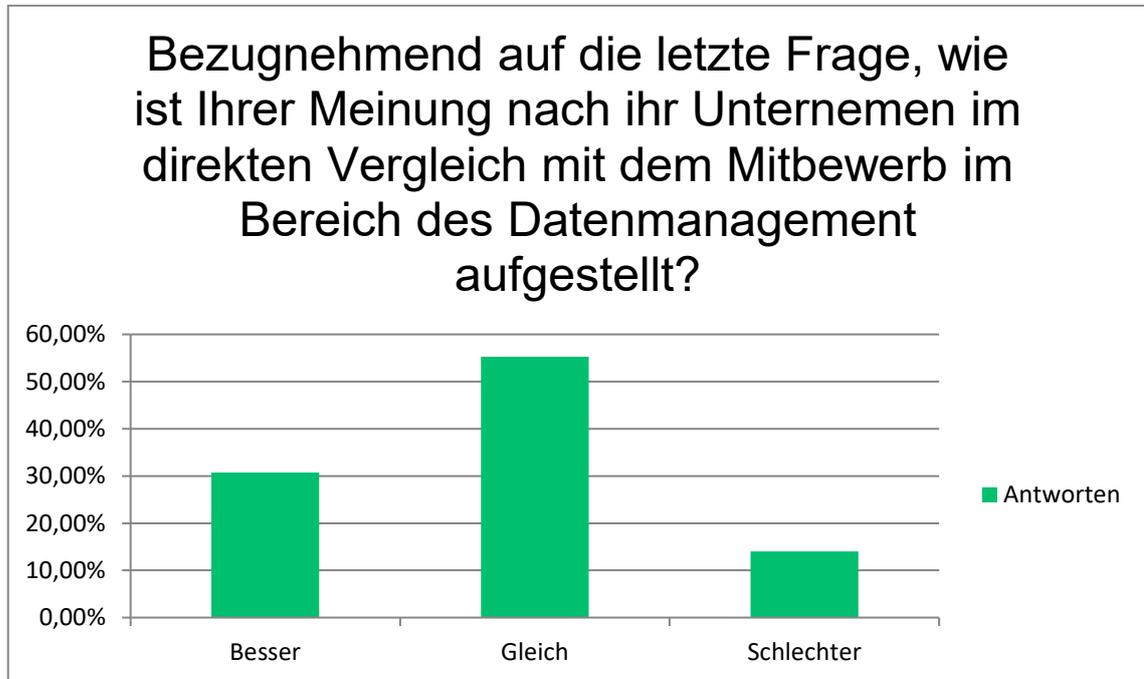


Abbildung 26: Frage 7: Implementierung<sup>160</sup>

<sup>160</sup> Quelle: Eigene Darstellung.

**Frage 8:**

Bezugnehmend auf die vorige Frage nach dem Implementierungsgrad (Abbildung 27), wurde in Frage 8 gefragt, ob die Teilnehmer, mit der gegebenen Antwort denken im Gegensatz zum direkten Mitbewerber besser oder schlechter dazustehen. Nur 14,04% (16 Teilnehmer) gaben an schlechter als der Wettbewerb aufgestellt zu sein, hingegen glauben 55,26% dass sie gleich gut organisiert sind.



**Abbildung 27: Frage 8: Implementierung Datenmanagement Vergleich<sup>161</sup>**

<sup>161</sup> Quelle: Eigene Darstellung.

**Frage 9:**

Frage 9 beschäftigt sich damit, wo die Implementierung des Datenmanagements die größten Vorteile gebracht hat (Abbildung 28). Bei dieser Frage wurde bewusst darauf verzichtet, den Teilnehmern die Möglichkeit einer eigenen Angabe in einem Textfeld zu geben, da die Antworten wohl sehr unterschiedlich ausfallen würden und eine Aussage dadurch nur sehr schwer zu treffen wäre. Auch mit den angebotenen Antwortoptionen, waren die Rückmeldungen sehr gut verteilt, so wurde jede der Optionen von durchschnittlich 36,42% angeklickt. Die TOP 3 waren dabei die Bereitstellung von Daten, Datenqualität und die möglichen Analysemethoden.

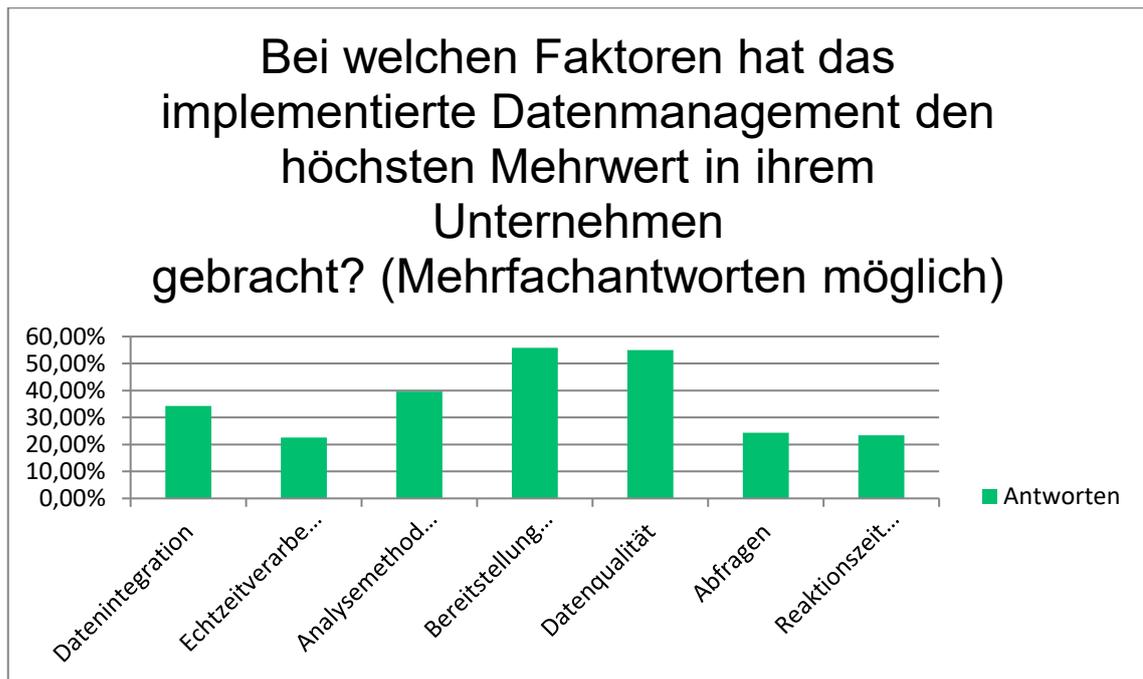
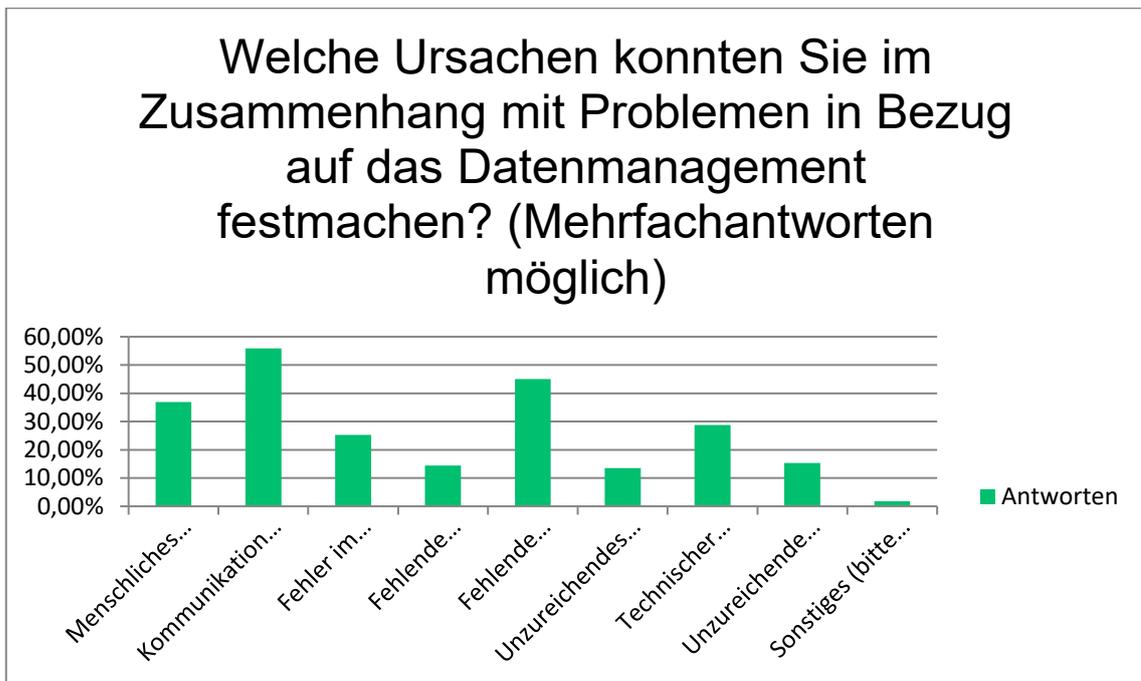


Abbildung 28: Frage 9: Mehrwert<sup>162</sup>

<sup>162</sup> Quelle: Eigene Darstellung.

**Frage 10:**

Die nächste Frage beschäftigt sich mit den Problemen, welche in Bezug auf das Datenmanagement festgestellt werden konnten. Auch bei dieser Frage gab es eine Streuung der Antworten, so wurden alle Antworten mindestens 15-mal ausgewählt. Die Zahl der Antworten übersteigt dabei die Teilnehmer, da eine Mehrfachauswahl möglich war. Über 55% der Teilnehmer gaben an, dass Kommunikationsschwierigkeiten der Abteilungen eines der Hauptprobleme darstellen (Abbildung 29).

Abbildung 29: Frage 10: Probleme<sup>163</sup><sup>163</sup> Quelle: Eigene Darstellung.

**Frage 11:**

Eine eindeutige Tendenz gibt es bei Teilnehmern bei der Frage ob operative Entscheidungen auf Grundlage von Daten entschieden werden, diese Frage wurden von über 90% der Teilnehmer mit JA beantwortet (Abbildung 30).

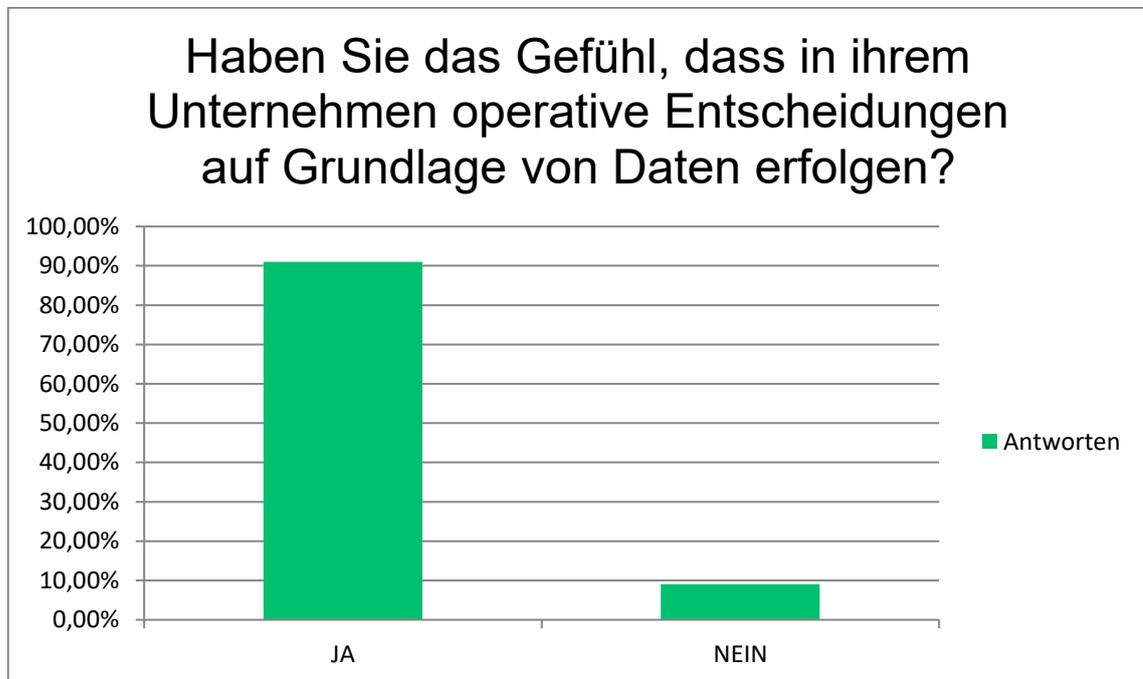


Abbildung 30: Frage 11: Operative Entscheidungen<sup>164</sup>

<sup>164</sup> Quelle: Eigene Darstellung.

**Frage 12:**

Ebenfalls eindeutig, wenn auch mit ein wenig geringerem Zuspruch, wurde die Frage nach den strategischen Entscheidungen beantwortet. Hier gaben knapp unter 84% der Teilnehmer an, dass diese auf Grundlagen von Daten getroffen werden (Abbildung 31).

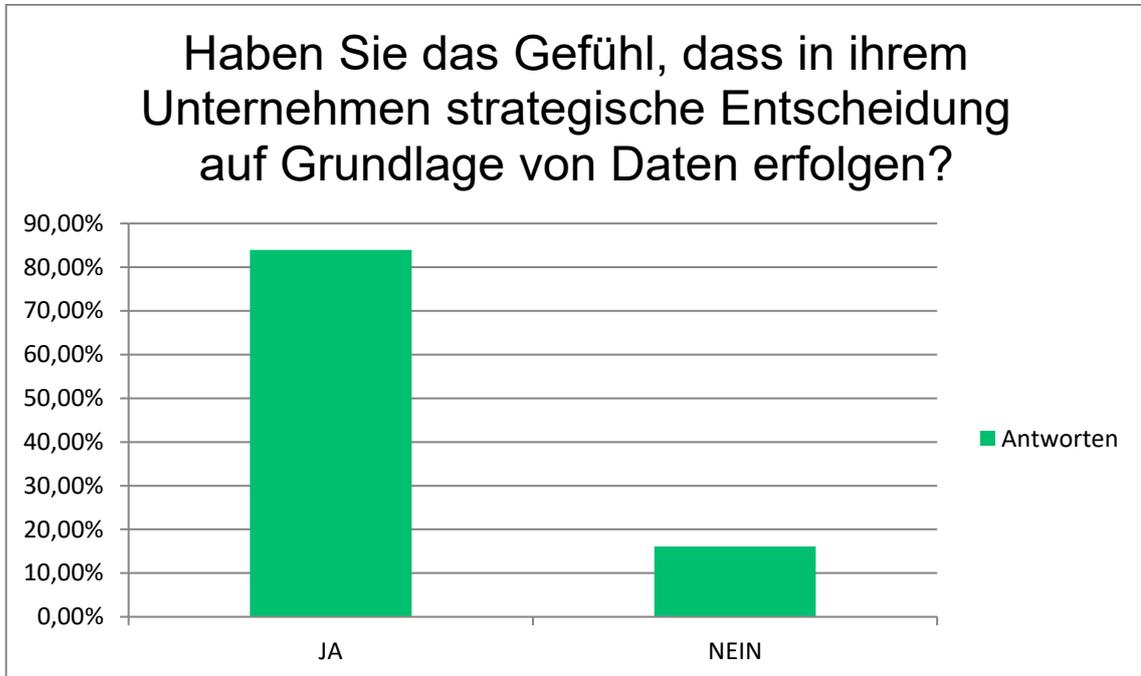


Abbildung 31: Frage 12: Strategische Entscheidungen<sup>165</sup>

<sup>165</sup> Quelle: Eigene Darstellung.

### 6.3 Technische Fragestellungen

#### Frage 13:

72,32% der Teilnehmer geben an, dass Auswerteperspektiven und Key Performance Indicators soweit standardisiert sind (Abbildung 32), dass ein abteilungsübergreifender Vergleich der Daten möglich ist, ohne dass diese noch umgestaltet werden müssen. Dieser Wert scheint im Hinblick auf die Teilnehmerstruktur nachvollziehbar.

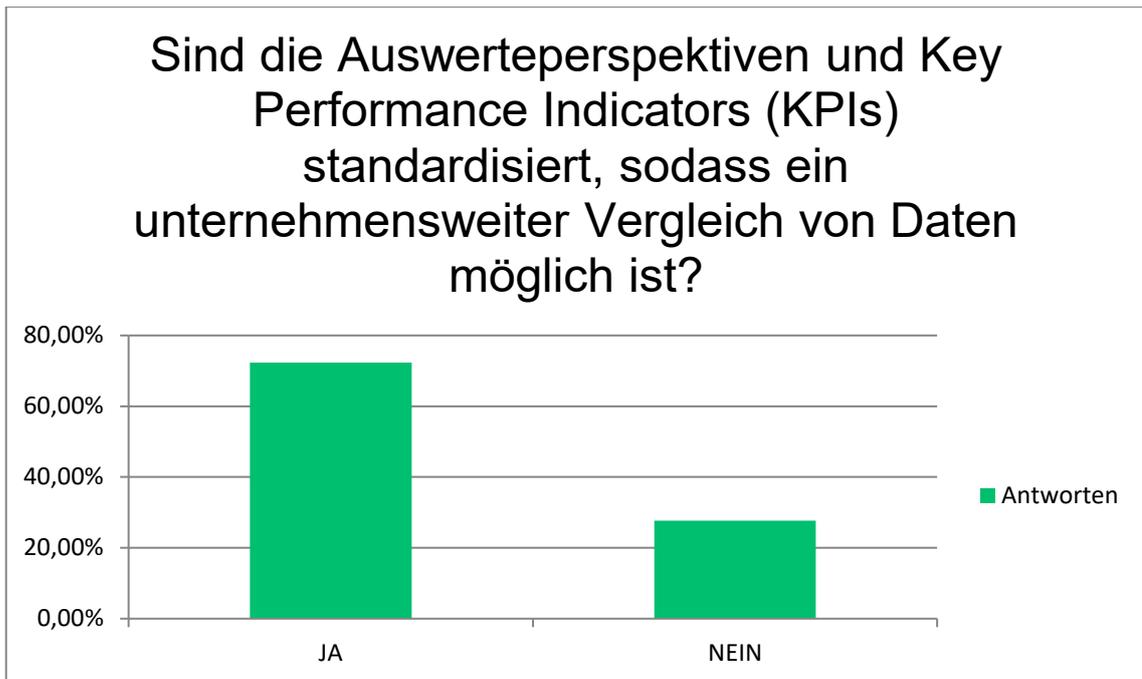
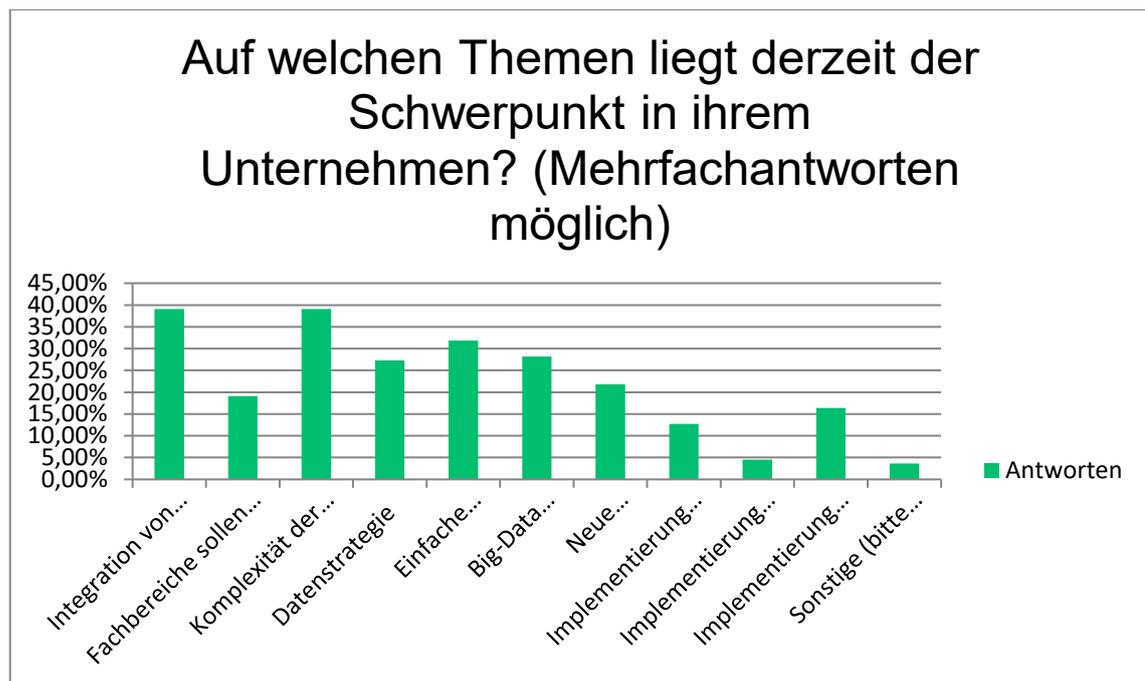


Abbildung 32: Frage 13: Standardisierung<sup>166</sup>

<sup>166</sup> Quelle: Eigene Darstellung.

**Frage 14:**

Bei Frage 14 war es den Teilnehmer möglich mehrere Antworten zu geben, da es durchaus sein kann, dass gerade unterschiedliche Schwerpunkte gleichzeitig gesetzt werden (Abbildung 33). Die Verteilung der Antworten lässt darauf schließen, dass viele unterschiedliche Schwerpunkte in Unternehmen gesetzt werden. Nur das Thema der Implementierung eines Data Lakes wurde von weniger als 5% der Teilnehmer ausgewählt. Am öftesten wurden die Antworten der Integration von heterogenen Datenquellen, der Komplexitätsminimierung der Dateninfrastruktur und der einfachen Integration von neuen Datenquellen ausgewählt.

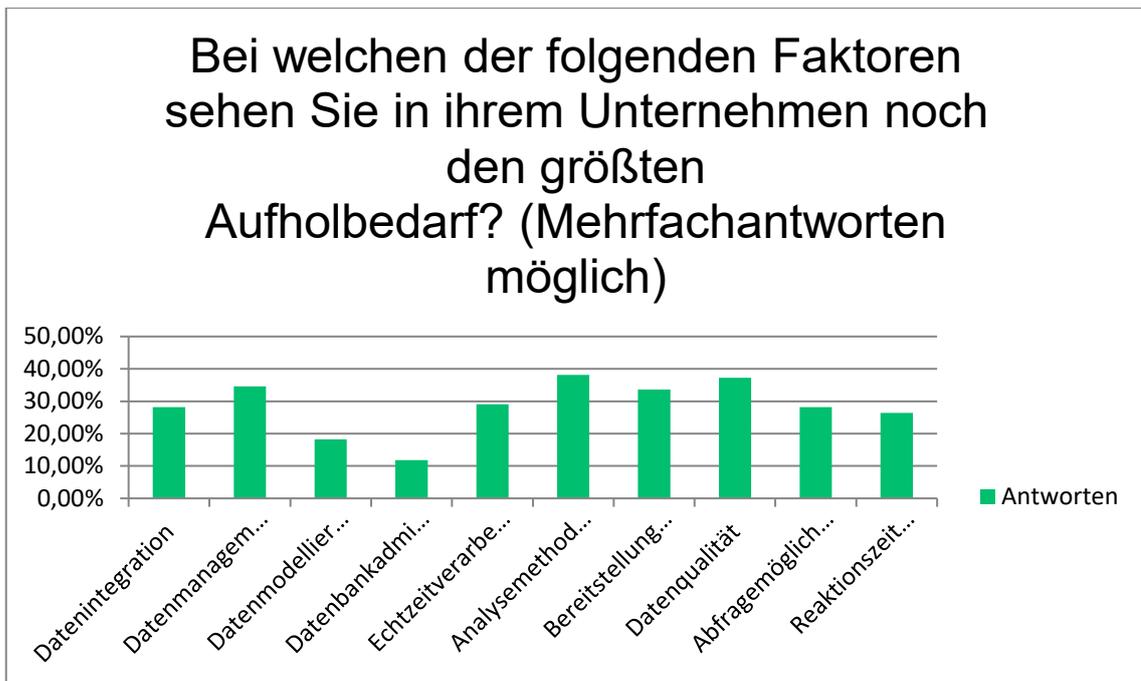


**Abbildung 33: Frage 14: Technischer Schwerpunkt<sup>167</sup>**

<sup>167</sup> Quelle: Eigene Darstellung.

**Frage 15:**

Ebenfalls sehr ausgewogen waren die Antworten bei der Frage nach den Faktoren bei welchen im jeweiligen Unternehmen der größte Aufholbedarf besteht. Hier wurde das Thema der Datenintegration, des Datenmanagements, der Echtzeitverarbeitung, der Analysemethoden, der Bereitstellung von Daten, der Datenqualität, der Abfragemöglichkeiten und der Reaktionszeit auf Änderungen oder neue Anforderungen von ca. 30% oder mehr genannt (Abbildung 34).



**Abbildung 34: Frage 15: Aufholbedarfsfaktoren<sup>168</sup>**

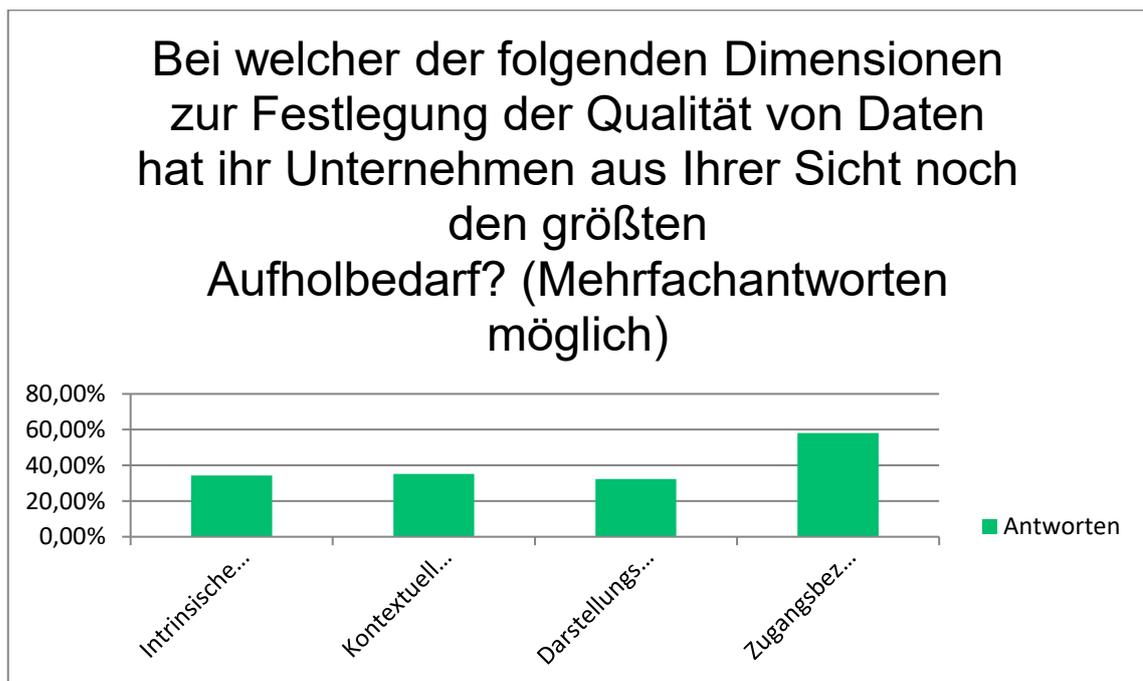
<sup>168</sup> Quelle: Eigene Darstellung.

**Frage 16:**

Die vier zur Antwort stehenden Dimensionen der Datenqualität wurden im Detail beschrieben, um einerseits diese Begriffe zu erklären, aber auch um eine klare Abgrenzung der Begriffe zu gewährleisten nach der die Antwort gegeben werden kann. Die Dimensionen lauten:

- Intrinsische Datenqualität: Genauigkeit, Objektivität und Fehlerfreiheit der Daten
- Kontextuelle Datenqualität: Sind die Daten vollständig, dass diese sinnvoll genutzt werden können
- Darstellungsbezogene Datenqualität: Sind die Daten übersichtlich, also in weiterer Folge auch leicht verständlich
- Zugangsbezogenen Datenqualität: Sind die Daten leicht auffindbar, kann auf diese leicht zugegriffen werden und können diese einfach bearbeitet werden

Dabei wurde die intrinsische, kontextuelle und darstellungsbezogene Datenqualität von ca. 30% ausgewählt. Nur die zugangsbezogene Datenqualität wurde von deutlich mehr Teilnehmer ausgewählt (58,10%) und scheint damit für die Unternehmen die wichtigste Dimension zu sein (Abbildung 35).

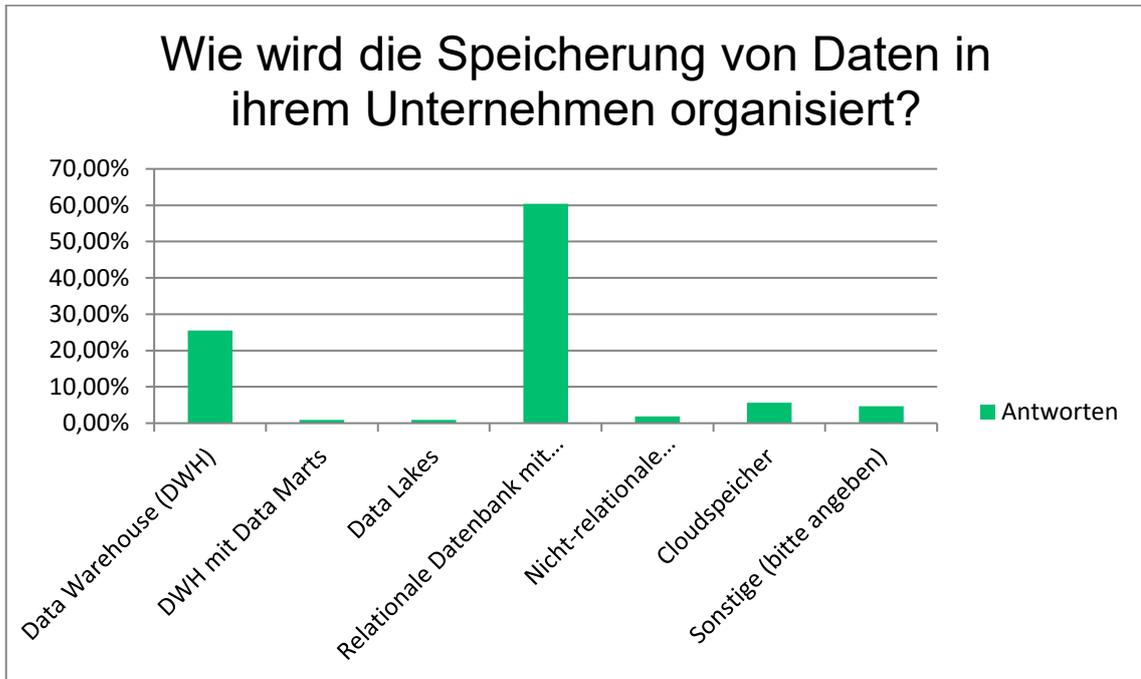


**Abbildung 35: Frage 16: Dimensionen der Datenqualität<sup>169</sup>**

<sup>169</sup> Quelle: Eigene Darstellung.

**Frage 17:**

Die Speicherung der Daten erfolgt bei den Unternehmen mehrheitlich in relationalen Datenbanken (60,38%). Am zweithäufigsten, jedoch mit deutlich weniger Stimmen, wurde angegeben, dass ein Data Warehouse (DWH) verwendet wird. Die neue Technologie der Data Lakes wird hingegen nur von einem der teilnehmenden Unternehmen verwendet (Abbildung 17).

**Abbildung 36: Frage 17: Speicherung<sup>170</sup>**

<sup>170</sup> Quelle: Eigene Darstellung.

## 6.4 Fragen zu Big Data und Datenanalysemethoden

### Frage 18:

In dieser Rubrik des Fragebogens geht es um Erfahrung mit Big Data Projekten und komplexeren Datenanalyseverfahren. Zu Beginn wurde nach der Erfahrung mit solchen Projekten gefragt. Diese Frage war schlicht mit JA oder NEIN zu beantworten und sollte einen ersten groben Überblick zu diesem Thema bieten. Überraschenderweise wurde diese Frage von über 50% der Teilnehmer mit NEIN beantwortet (Abbildung 37).

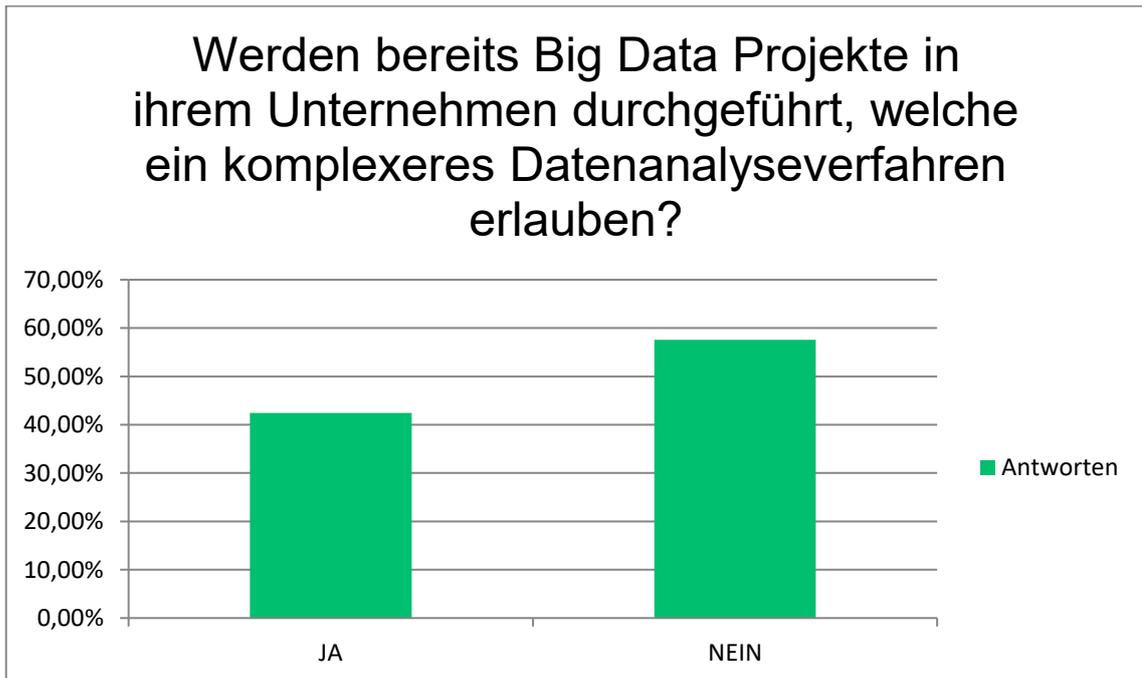
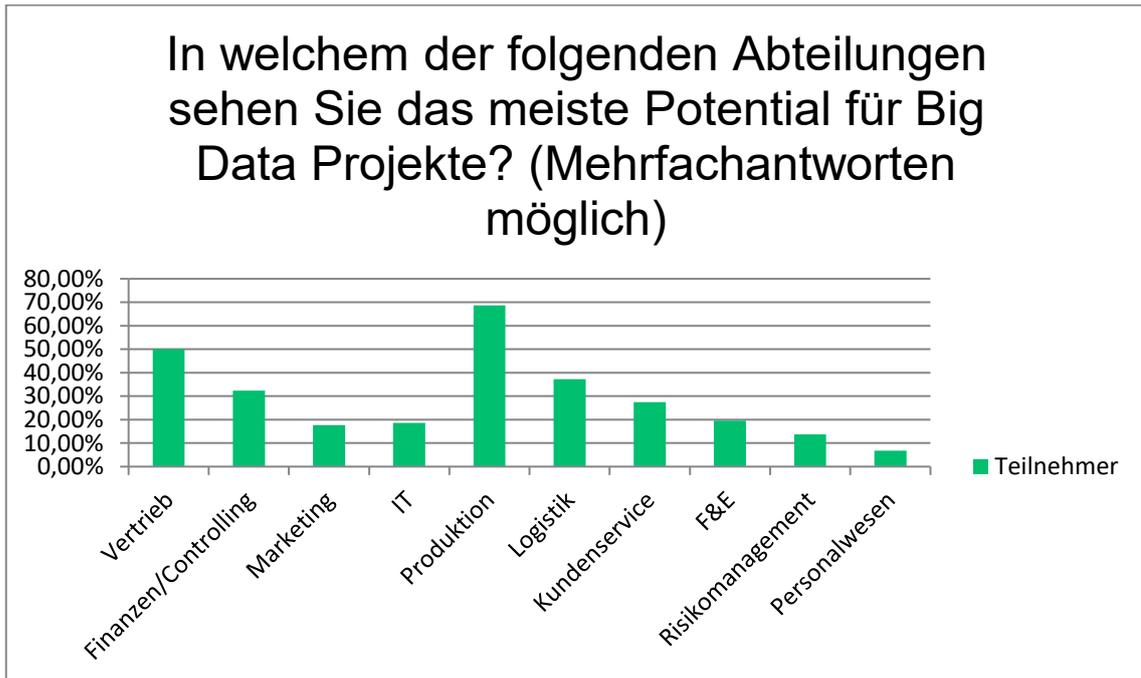


Abbildung 37: Frage 18: Erfahrung mit Big Data<sup>171</sup>

<sup>171</sup> Quelle: Eigene Darstellung.

**Frage 19:**

Aufbauend auf der letzten Frage, wurden die Teilnehmer gefragt, in welcher Fachabteilung sie das größte Potential für Big Data Projekte sehen (Abbildung 38). Diese Frage kann entweder aus persönlicher Erfahrung im Unternehmen beantwortet werden, aber auch von jenen Teilnehmern, welche die letzte Frage mit NEIN beantwortet haben. Es wurde von den Teilnehmer Potential in jeder der vorgeschlagenen Abteilungen gesehen. Am meisten Potential sehen die Teilnehmer jedoch in der Produktion (68,63%) und im Vertrieb (50%).

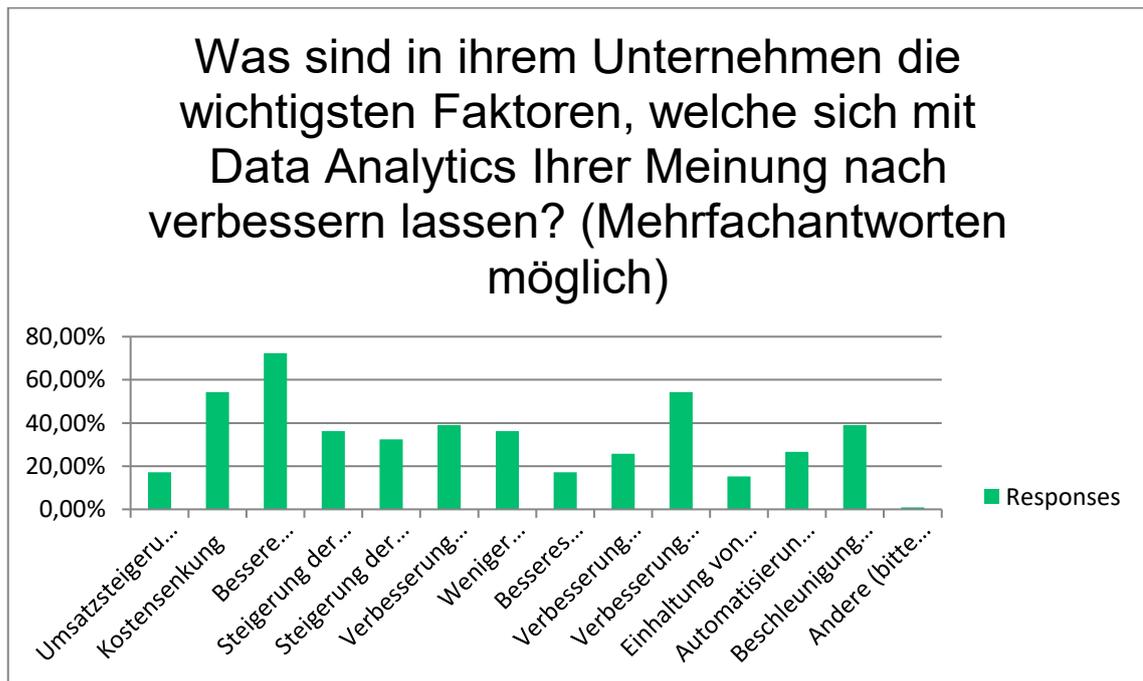


**Abbildung 38: Frage 19: Potential für Big Data Projekte<sup>172</sup>**

<sup>172</sup> Quelle: Eigene Darstellung.

**Frage 20:**

Die nachfolgende Frage geht explizit auf das Thema der Datenanalyseverfahren ein. Es wurden die Faktoren abgefragt wo Analysemethoden den meisten Mehrwert bringen (Abbildung 39). Auch bei dieser Frage wurden alle zur Verfügung gestellten Antworten sehr oft angeklickt, der Mittelwert bezogen auf alle Antworten lag bei 33,33%. Prozentuell die wenigsten Antworten gab es für Umsatzsteigerung, besseres Verständnis der Kundenwünsche und für die Einhaltung von Gesetzen und Regularien. Die TOP 3 Antworten lauten: Bessere Steuerung von operativen Prozessen, Kostensenkung, Verbesserung der Produktionsplanung. Die Topantworten haben jeweils mehr als die Hälfte der Teilnehmer angeklickt, welche diese Frage beantwortet haben.

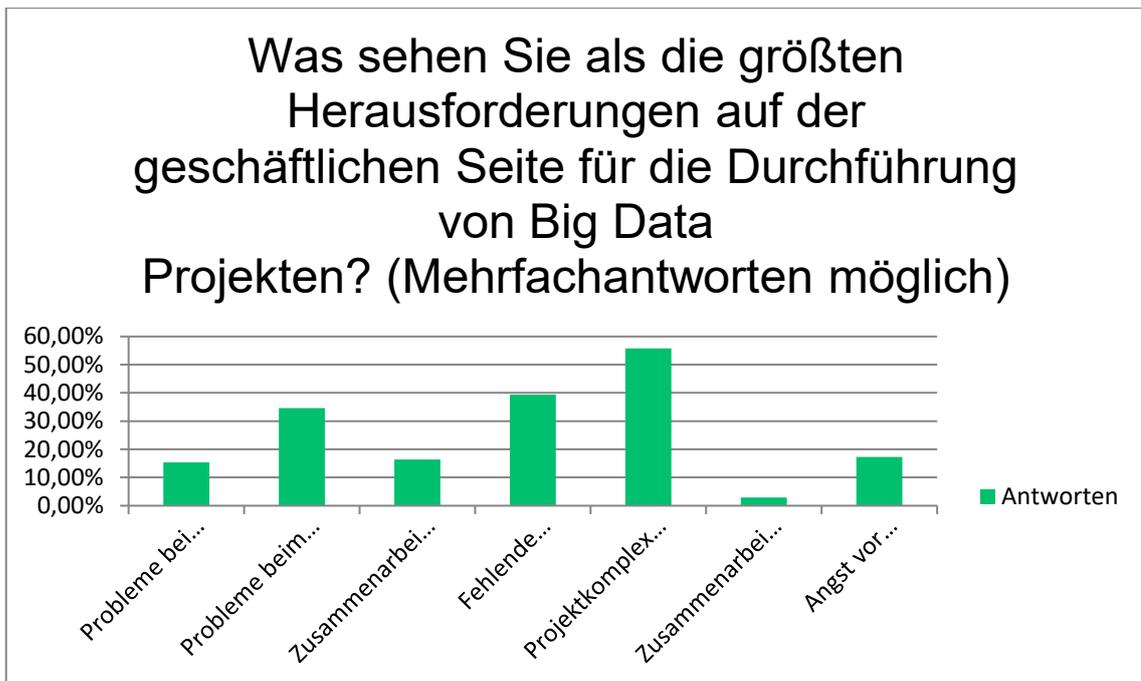


**Abbildung 39: Frage 20: Verbesserungen durch Data Analytics<sup>173</sup>**

<sup>173</sup> Quelle: Eigene Darstellung.

**Frage 21:**

Es ist oftmals sehr schwierig einen Business Case zu finden, welcher ein Big Data Projekt rechtfertigen würde. Bei der Frage sollte überprüft werden, welche Herausforderungen auf der geschäftlichen Seite als die größten angesehen werden (Abbildung 40). Es gab hier drei Antworten, welche sehr oft ausgewählt wurden, es waren dies, die Projektkomplexität, fehlende Schulungen und/oder Kenntnisse der Mitarbeiter und der fehlende Business Case.

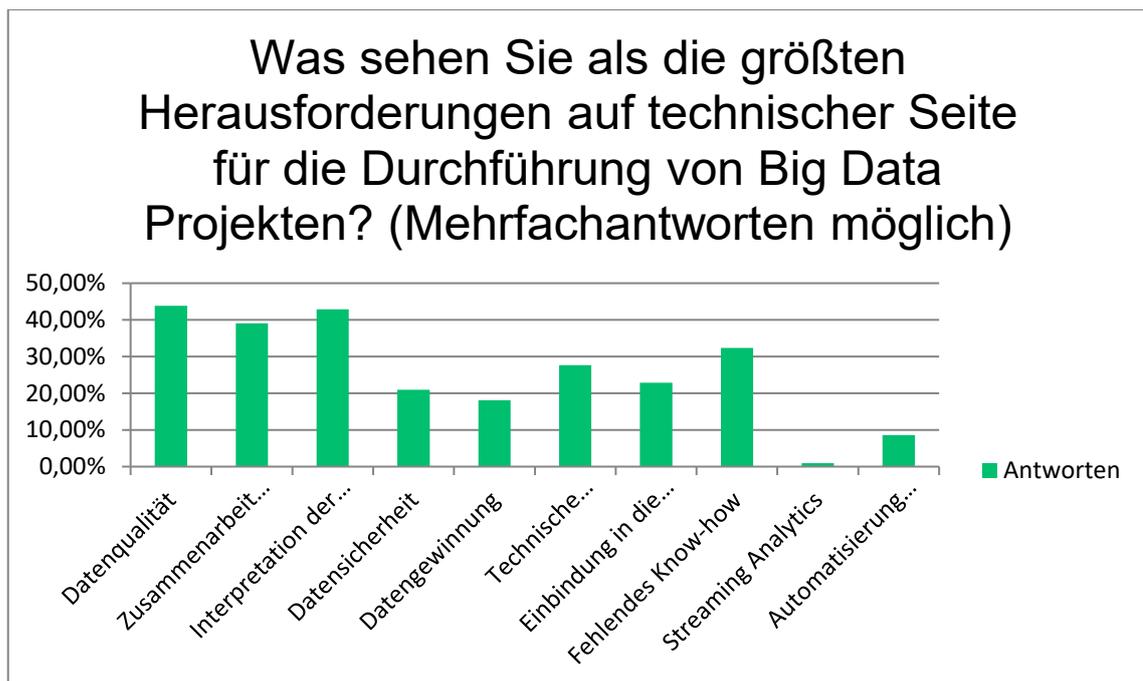


**Abbildung 40: Frage 21: Herausforderungen auf geschäftlicher Seite<sup>174</sup>**

<sup>174</sup> Quelle: Eigene Darstellung.

**Frage 22:**

Ergänzend dazu wurden die technischen Herausforderungen für so ein Projekt abgefragt (Abbildung 41). Es zeigt sich bei dieser Frage wiederholt, dass sich bei den Herausforderungen kein eindeutiges Muster erkennen lässt, weil alle Antworten häufig ausgewählt werden, was jedoch bei einer relativ neuen Technologie oftmals der Fall ist. Als wenig herausfordernd werden nur die Automatisierung von Entscheidungen und Streaming Analytics angesehen. Am öftesten wurde bei dieser Frage die Datenqualität (43,81%), die Interpretation der Daten (42,86%) und das fehlende Know-how (32,38%) als Herausforderung auf technischer Seite angegeben.



**Abbildung 41: Frage 22: Herausforderungen auf technischer Seite<sup>175</sup>**

<sup>175</sup> Quelle: Eigene Darstellung.

**Frage 23 und Frage 24:**

In den nächsten beiden Fragestellungen wurde überprüft welche Art von Analysemethoden derzeit im Unternehmen eingesetzt werden (Abbildung 42) und welche voraussichtlich in fünf Jahren eingesetzt werden (Abbildung 43). Die Analysemethoden welche sich mit Events in der Vergangenheit beschäftigen wurden dabei mit 75,25% (Deskriptive Analyse) bzw. 69,31% (Diagnostische Analyse) mehr als doppelt so häufig angeklickt wie die komplexeren Analyseformen der prädiktiven und präskriptiven Analyse. Bei der Frage nach der Einschätzung wie die Situation in fünf Jahren aussehen wird, zeigt sich genau das umgekehrte Bild, hier sind über 60% der Meinung das prädiktive und präskriptive Analysen vorherrschend sein werden.

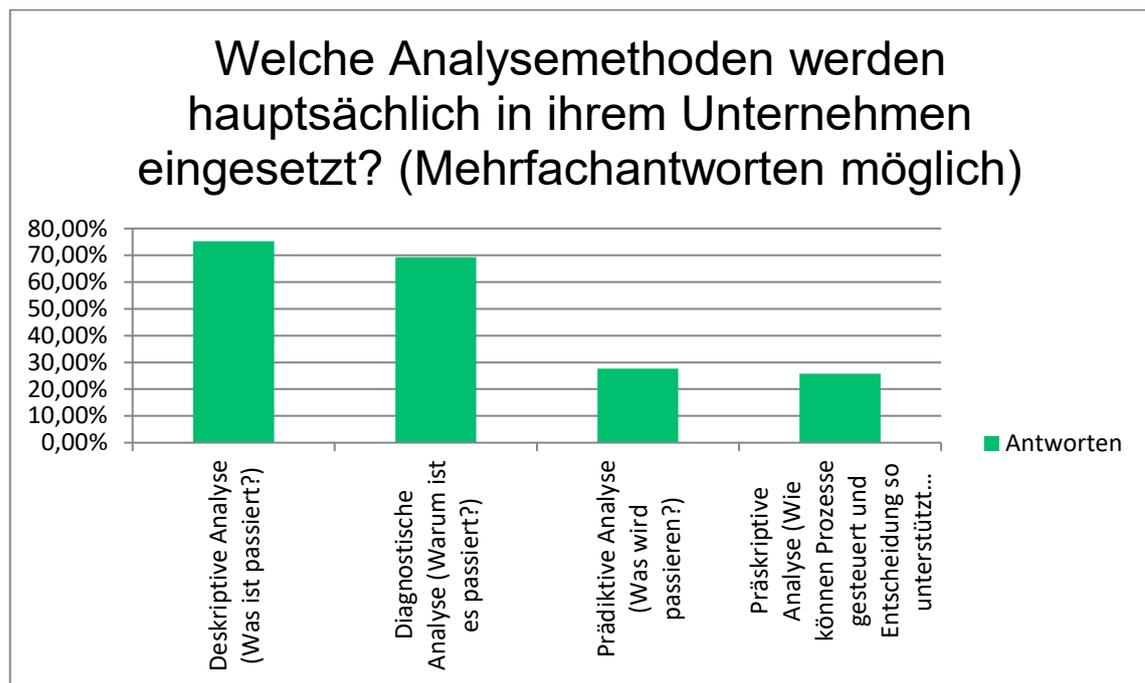


Abbildung 42: Frage 23: Derzeitige Analyseformen<sup>176</sup>

<sup>176</sup> Quelle: Eigene Darstellung.

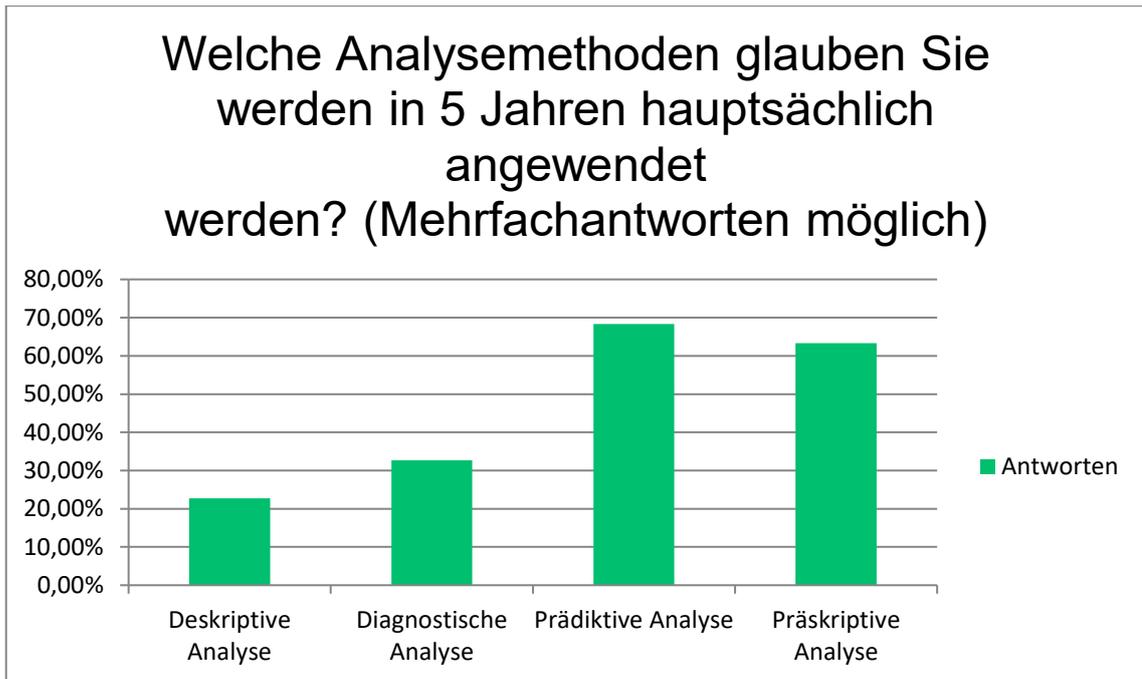
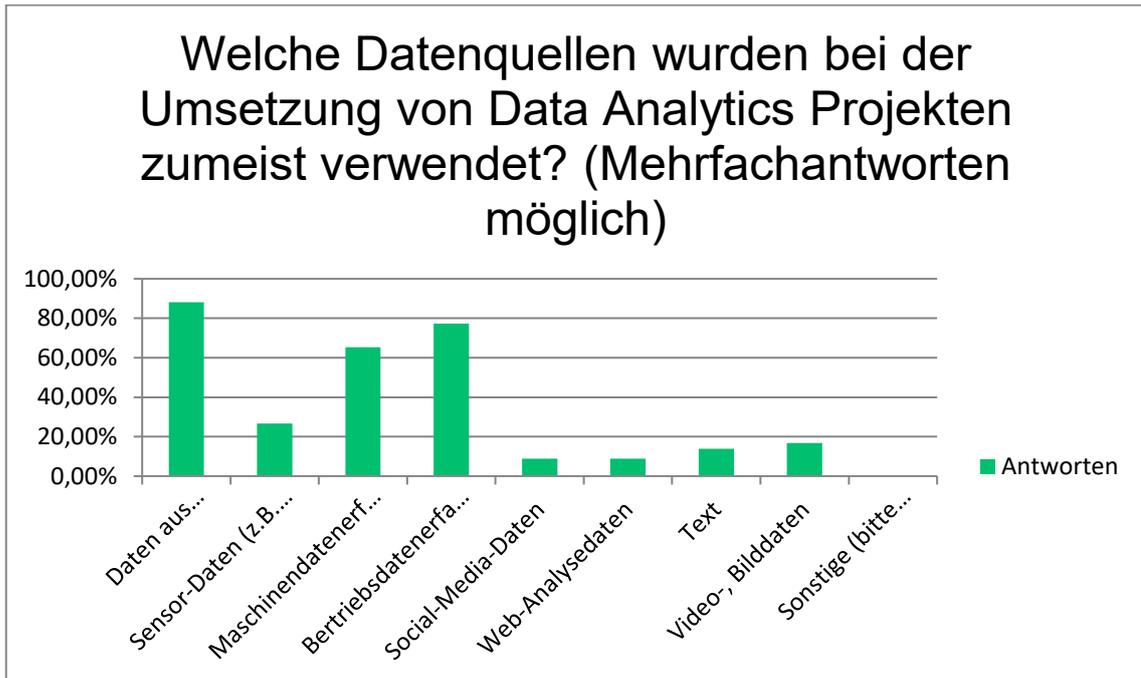


Abbildung 43: Frage 24: Zukünftige Analysemethoden<sup>177</sup>

<sup>177</sup> Quelle: Eigene Darstellung.

**Frage 25:**

Bei der letzten Frage werden die Teilnehmer dazu befragt, welche der folgenden Datenquellen sie für Datenanalyseprojekte verwendet haben. Hier zeigt sich ein klarer Trend zu drei Antworten, welche jeweils von über 60% der Teilnehmer ausgewählt worden sind. Am häufigsten werden Daten aus Transaktionssystemen für Datenanalyseprojekte verwendet, diese werden von 88,12% der Teilnehmer verwendet. Von über 60% werden Betriebs- und Maschinendaten verwendet. Unter 10% verwenden Social-Media und Web-Analysedaten (Abbildung 44).



**Abbildung 44: Frage 25: Datenquellen<sup>178</sup>**

In den folgenden Kapiteln wird nun versucht mit den gewonnenen Daten, neue Erkenntnisse über den Umgang mit Daten in Österreich zu generieren und aufgestellte Hypothesen zu diesem Thema zu überprüfen.

<sup>178</sup> Quelle: Eigene Darstellung.

## 6.5 Diskussion zur Fragebogenerhebung

In diesem Kapitel werden die gewonnenen Ergebnisse aus dem letzten Kapitel genauer analysiert. Es wird versucht zu den Kernthemen, Hypothesen aufzustellen, welche in weiterer Folge auf Grundlage der gewonnenen Daten überprüft werden. Falls bei der Überprüfung der Korrelation die Summe der Antworten zweier Fragen nicht übereinstimmend ist, so ist das dem Umstand geschuldet, dass einer oder mehrere der Teilnehmer sich bei einer Frage der Antwort entzogen haben.

Die nachfolgenden Hypothesen sollen dazu beitragen einen Überblick über den aktuellen Stand des Datenmanagements in österreichischen Industrieunternehmen zu geben und Trends über zukünftige Entwicklungen aufstellen zu können:

### **1. Bestimmte industrielle Sektoren scheinen sowohl von der organisatorischen wie auch von der technischen Seite sehr viele Ressourcen in Datenmanagement investiert zu haben.**

Der Fragebogen wurde an Unternehmen aller Industriesektoren verschickt. Am meisten Rückmeldungen gab es von Unternehmen aus den Sektoren der metalltechnischen Industrie (34 Antworten), der chemischen Industrie (13 Antworten) und der Papierindustrie (11 Antworten). Somit stammen die meisten Teilnehmer aus der metalltechnischen Industrie, insgesamt 24,82% bezogen auf die Gesamteilnehmerzahl. Diese Rücklaufquoten relativieren sich jedoch, da der metalltechnische Sektor mit einem Produktionswert von ca. EUR 37 Milliarden<sup>179</sup> und der chemische Sektor mit einem Produktionswert von ca. EUR 16 Milliarden<sup>180</sup> und der Papiersektor mit ca. EUR 11 Milliarden<sup>181</sup> zu den größten Einzelbranchen in Österreich zählen. Alleine von der Anzahl der Teilnehmer lässt sich somit nicht ableiten ob diese ein gut funktionierendes Datenmanagement implementiert haben. Da auch eine Teilnahme nicht zwangsläufig bedeutet, dass die Fragen positiv beantwortet werden, andererseits ist die Motivation an der Teilnahme bei Umfragen zu vorherrschenden Themen im Normalfall höher. Als Grundlage für die Beantwortung dieser These wurden die Antworten von Frage 6 und 7 analysiert. Hier wurde abgefragt ob schon ein unternehmensweites Datenmanagement implementiert wurde und wer für die Daten in Unternehmen verantwortlich ist. Bei der Auswertung wurde auch darauf geachtet, dass kleinere Unternehmen auch wenn ein unternehmensweites Datenmanagement implementiert ist, aufgrund der Unternehmensgröße wohl keinen Chief Data oder Chief Information Officer angestellt haben. Es kann jedoch bei einer auf das Datenmanagement ausgelegten Organisation davon ausgegangen werden, dass es einen eigens definierten Datenverantwortlichen gibt. Es gibt nur 12 Teilnehmer, die einerseits gesagt haben, dass sie ein unternehmensweites

<sup>179</sup> Quelle: Metalltechnische Industrie, <https://www.metalltechnischeindustrie.at/ueber-uns/> (Zugriff: 22.09.2019).

<sup>180</sup> Quelle: Die Presse, [https://diepresse.com/home/wirtschaft/economist/5635107/Chemische-Industrie\\_Hoffen-dass-die-naechste-Regierung-diesen](https://diepresse.com/home/wirtschaft/economist/5635107/Chemische-Industrie_Hoffen-dass-die-naechste-Regierung-diesen) (Zugriff: 22.09.2019).

<sup>181</sup> Quelle: WKO, [https://www.wko.at/branchen/stmk/industrie/papierindustrie/infos\\_lehre\\_papierindustrie.html](https://www.wko.at/branchen/stmk/industrie/papierindustrie/infos_lehre_papierindustrie.html) (Zugriff: 22.09.2019).

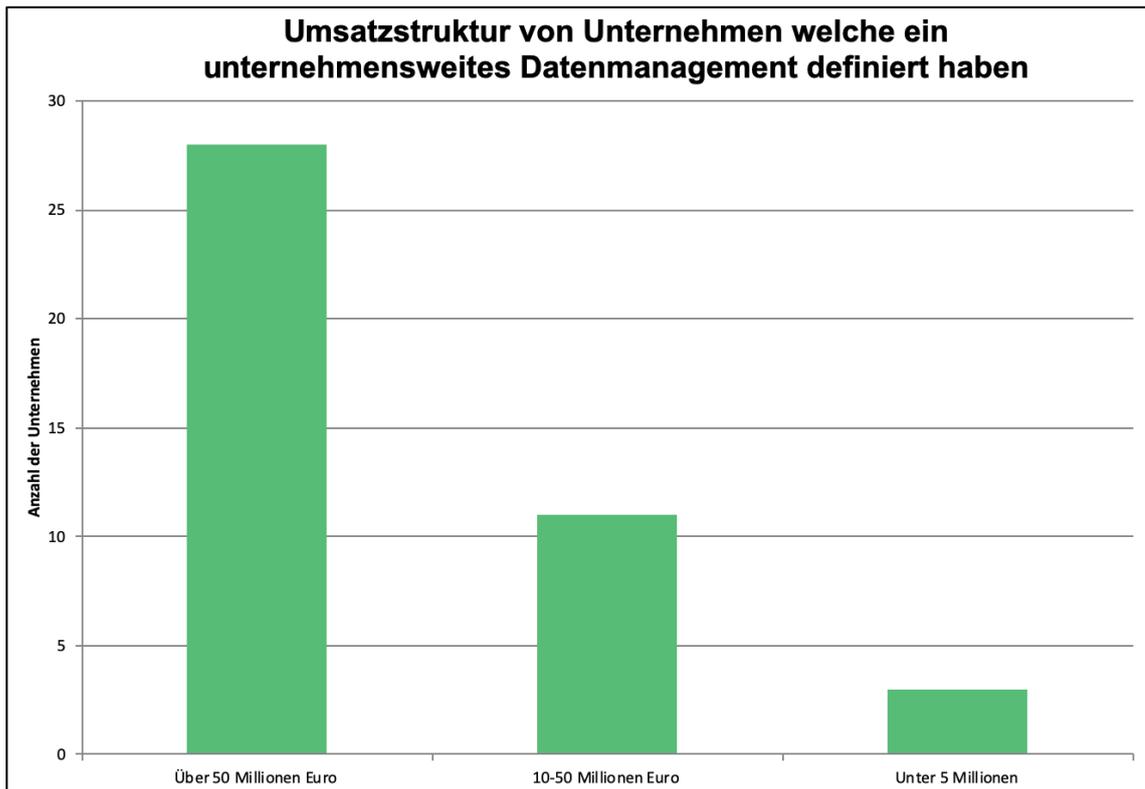
Datenmanagement implementiert haben und andererseits angegeben haben, dass sie eine Person beschäftigen, welche eine zentrale Anlaufstelle für die Zuständigkeit von Daten ist. Nur Unternehmen des metalltechnischen Sektors haben beide Fragen mehrmals positiv beantwortet. Bei der metalltechnischen Industrie haben 5 der insgesamt 34 Teilnehmer aus diesem Sektor die beiden Fragen positiv beantwortet, dies entspricht einem Prozentsatz von ca. 14%, was nicht genügend ist, um diese Hypothese bestätigen zu können. Zumal es die meisten Rückmeldungen aus diesem Sektor gab, somit ist auch ohne eine Korrelation dieser Faktoren, die Wahrscheinlichkeit hoch, dass es hier Unternehmen gibt, welche beide Fragen positiv beantworten.

**2. Es besteht ein Zusammenhang zwischen Unternehmen, welche kein unternehmensweites Datenmanagement implementiert haben und dem Umsatz, den das Unternehmen erwirtschaftet.**

44 Teilnehmer haben angegeben, dass ein unternehmensweites Datenmanagement eingerichtet wurde. Davon haben 28 einen Umsatz von über 50 Millionen Euro und fallen somit in die Kategorie der Großunternehmen in Österreich<sup>182</sup>. Nochmals 11 Teilnehmer haben einen Umsatz von 10-50 Millionen Euro und sind somit in der Kategorie der mittleren Unternehmen angesiedelt. Nur 3 Unternehmen haben einen Umsatz von unter 5 Millionen (Abbildung 45). Es kann also ein klarer Zusammenhang zwischen dem Umsatz des Unternehmens und einem unternehmensweiten Datenmanagement gezeigt werden.

---

<sup>182</sup> Quelle: <https://www.wko.at/service/zahlen-daten-fakten/KMU-definition.html> (Zugriff: 20.09.2019).



**Abbildung 45: Umsatzstruktur von Unternehmen mit implementierten Datenmanagement<sup>183</sup>**

**3. Unternehmen, welche gesonderte Stellen für den Umgang mit Daten geschaffen haben, beschäftigen viele Leute in Österreich und haben Erfahrung mit Big Data Projekten.**

5 der teilnehmenden Unternehmen haben einen Chief Information Officer (CIO), welcher für die Daten verantwortlich ist. Ein Unternehmen hat einen Chief Data Officer und 16 Unternehmen haben einen eigenen Datenverantwortlichen angestellt. Somit haben nur 22 der 141 teilnehmenden Unternehmen angegeben, dass sie eine eigene Person für das Datenmanagement eingestellt haben. 14 dieser Unternehmen haben über (oder knapp unter) 250 Mitarbeiter, gelten also laut der WKO als Großunternehmen<sup>184</sup>. Dies entspricht ca. 63% jener Unternehmen mit einer fix definierten Person, welche im Unternehmen für Daten zuständig ist. Von diesen Großunternehmen haben 9, also ca. 64 % bereits Erfahrung mit Big Data. Wenn der Faktor der Größe des Unternehmens weggelassen wird so entspricht der Anteil an Unternehmen welche bereits Big Data Projekte durchgeführt haben ungefähr 50 Prozent. Der Gesamtanteil der Teilnehmer, bei welchen die Firmen noch keine Big Data Projekte durchgeführt haben, liegt bei 61, jener von Unternehmen, welche bereits durchgeführt haben bei 45 Teilnehmern. Es ist als ein Trend erkennbar, dass Unternehmen welche gesonderte Stellen für die Datenverantwortlichkeit definiert haben Großunternehmen sind und Erfahrung mit Big Data Projekten haben. Allerdings

<sup>183</sup> Quelle: Eigene Darstellung.

<sup>184</sup> Quelle: <https://www.wko.at/service/zahlen-daten-fakten/KMU-definition.html> (Zugriff: 20.09.2019).

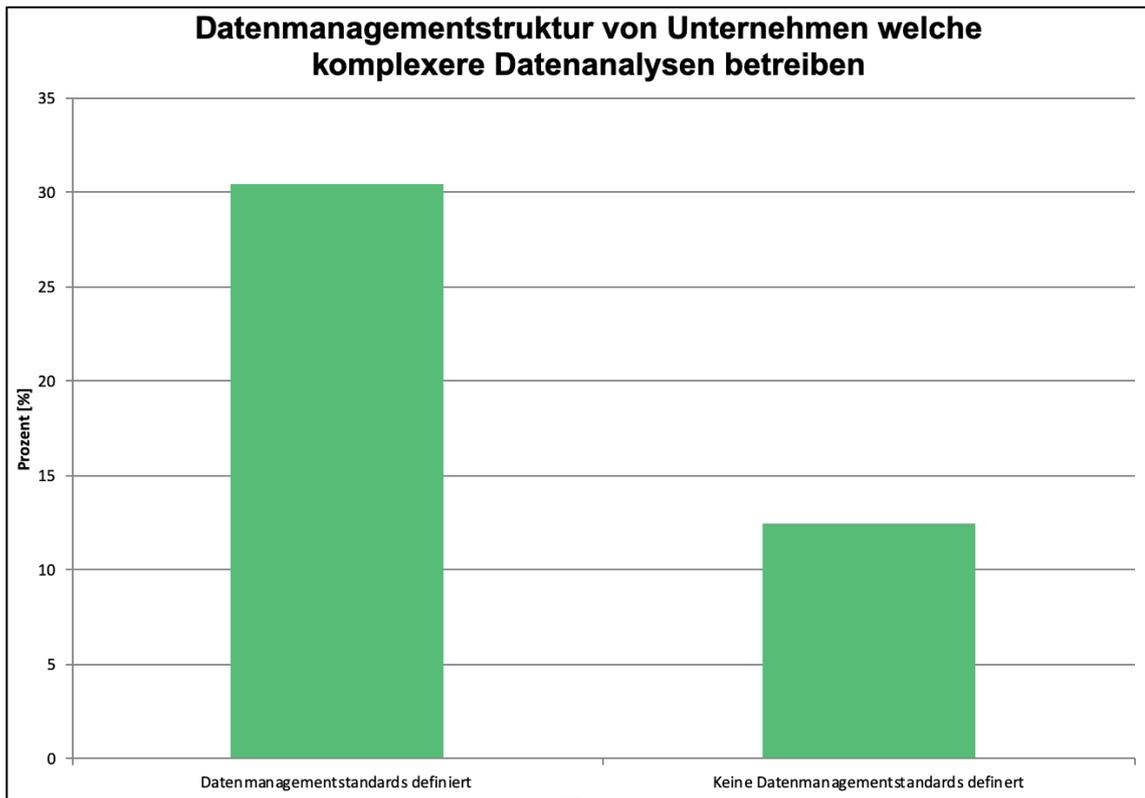
kann man hier nicht von einer generellen Regel sprechen, da dafür die Tendenzen zu klein sind.

**4. Unternehmen bei welchen jede Fachabteilung selbst für die Daten verantwortlich ist, haben mit mehr Problemen in Bezug auf das Datenmanagement zu kämpfen.**

Von den teilnehmenden Unternehmen haben 28 angegeben, dass die Fachabteilung selbst für Daten verantwortlich ist. Insgesamt wurden 66 Probleme in Bezug auf das Datenmanagement angegeben, das entspricht durchschnittlich 2,4 angegebene Probleme pro Unternehmen. Am häufigsten wurde von diesen Unternehmen angegeben, dass Kommunikationsschwierigkeiten mit unterschiedlichen Abteilungen herrschen. 113 Unternehmen haben die Verantwortung für Daten anders definiert, hier wurden insgesamt 204 Probleme mit Daten ausgewählt. Durchschnittlich sind dies 1,8 Probleme pro Unternehmen. Dies entspricht 0,6 Problemen weniger in Bezug auf das Datenmanagement. Auch von diesen Teilnehmern wurde die Kommunikation mit anderen Fachabteilungen als größtes Problem angegeben. Es konnte somit nicht gezeigt werden, dass die Kommunikation mit anderen Fachabteilungen nur ein Problem darstellt, wenn es keine zentrale Stelle für die Datenverantwortung gibt. Es ist jedoch ein Trend erkennbar, dass Organisationen, welche die Datenverantwortung zentral festlegen durchschnittlich eine kleinere Anzahl an Problemen mit dem Datenmanagement haben. Die Seriosität und die dazugehörigen Auswirkungen dieses Problems wurden dabei nicht untersucht.

**5. Unternehmen, welche keinen fixen Standard in Bezug auf das Datenmanagement definiert haben und dies nicht planen oder erst in Zukunft planen, betreiben keine komplexeren Datenanalysen.**

12 Unternehmen haben angegeben, dass keine fixen Standards in Bezug auf Datenmanagement bestehen, aber in Zukunft geplant sind. Bei vier Unternehmen ist gar nicht geplant diese Standards in Zukunft aufzusetzen. Von diesen insgesamt 16 Unternehmen betreiben nur 2 komplexe Datenanalysen wie die prädiktive oder präskriptive Analyse. Von den 141 Gesamtteilnehmern betreiben immerhin 43 komplexe Analysen. Dies sind immerhin 30,5%, im Gegensatz zu nur 12,5% bei Unternehmen, welche derzeit noch keine Datenmanagementstandards definiert haben (Abbildung 46).



**Abbildung 46: Datenmanagementstruktur von Unternehmen welche komplexere Datenanalysen betreiben<sup>185</sup>**

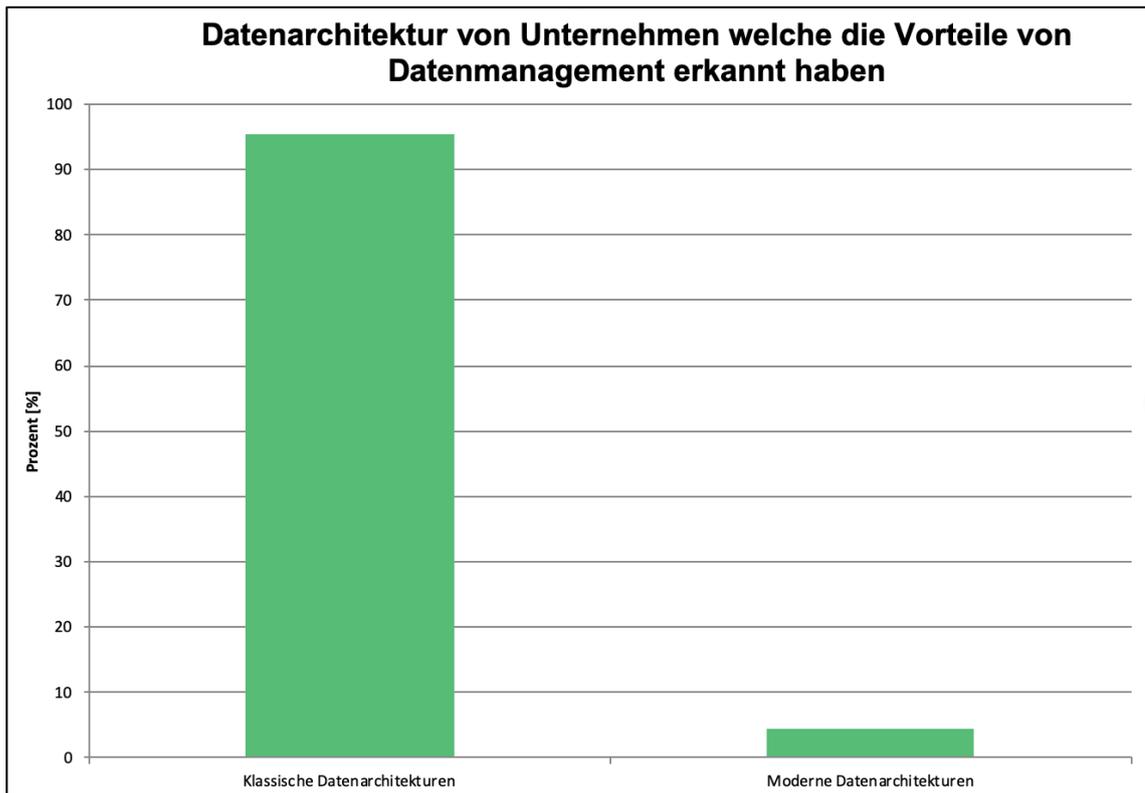
**6. Unternehmen, welche noch kein unternehmensweites Datenmanagement definiert haben, glauben dass sie im direkten Vergleich mit dem Wettbewerb nicht schlechter aufgestellt sind.**

Von den insgesamt 115 Teilnehmern, welche diese Frage beantwortet haben, wurde 16-mal angegeben, dass derzeit keine Standards in Bezug auf Datenmanagement definiert sind. Von diesen Unternehmen glauben jedoch nur 2 damit besser als der direkte Wettbewerb aufgestellt zu sein. 8 Teilnehmer gaben an gleich gut aufgestellt zu sein, wenn es um das Thema Datenmanagement geht. 5 Teilnehmer gaben zu, dass das Unternehmen mit dem jetzigen Handeln schlechter aufgestellt sei als die Konkurrenz. Von den 44 Teilnehmern, welche angegeben haben, dass es zumindest in einzelnen Abteilungen ein definiertes Datenmanagement gibt, glauben 8 davon, damit trotzdem schlechter aufgestellt zu sein als die Konkurrenz. Das sind ca. 18 %, im Gegensatz zu ca. 31 % bei den Teilnehmern ohne implementiertes Datenmanagement. Es ist bei dieser Hypothese ein Trend zu erkennen, welcher bestätigen würde, dass durchaus mehr Teilnehmern bewusst ist, dass ein Unternehmen ohne definiertes Datenmanagement wohl schlechter aufgestellt ist als die Konkurrenz. Der Prozentsatz von 31 % sagt aber auch aus, dass vielen Teilnehmern der Wettbewerbsnachteil nicht bewusst ist. Die aufgestellte Hypothese 2) zeigt jedoch einen klaren Zusammenhang zwischen Umsatz und einem definierten Datenmanagement.

<sup>185</sup> Quelle: Eigene Darstellung.

**7. Unternehmen, welche die Vorteile durch das implementierte Datenmanagement erkannt haben, sind eher bereit in moderne Datenarchitekturen zu investieren.**

Als Faktor ob einem Unternehmen die Vorteile eines definierten Datenmanagements bewusst sind, wurde die Anzahl der Antworten auf die Frage nach den Faktoren, bei denen das definierte Datenmanagement die größten Vorteile gebracht hat herangezogen. Es wurde angenommen, dass ab drei ausgewählten Faktoren das Unternehmen die Vorteile erkannt hat. Nicht berücksichtigt bleibt hierbei natürlich, wenn zwar nur ein Faktor ausgewählt wurde, aber dieser für das Unternehmen erheblich ist. 44 der Befragten haben drei oder mehr Faktoren ausgewählt. Ein Großteil davon verwendet klassische Datenarchitekturen wie Datawarehouses oder relationale Datenbanken (10 Antworten DWH, 26 Antworten relationale Datenbanken, Rest keine Antwort). Nur 2 von diesen Unternehmen verwenden moderne Speichermethoden (Abbildung 47). Von den Gesamtteilnehmern verwenden überhaupt nur 10 modernere Speichermethoden wie Cloudspeicher, DWH mit Data Marts, NoSQL-Datenbanken oder Data Lakes. Dieser Trend lässt sich somit nicht belegen, es zeigt sich jedoch eindeutig, dass in der österreichischen Industrie noch hauptsächlich auf die bewährten Datenarchitekturen gesetzt wird. Ebenfalls könnte es sein, dass die Anzahl der Faktoren bei welchen die Unternehmen Vorteile durch das Datenmanagement sehen kein geeigneter Gradmesser ist um darauf zu schließen, ob ein Unternehmen die Wichtigkeit eines definierten Datenmanagements erkannt hat, da durchaus einige wenige Faktoren für dieses Unternehmen ausschlaggebend sind. Jedoch kann davon ausgegangen werden, dass diese Teilnehmer dann keineswegs das volle Potential und die daraus entstehende Wichtigkeit erkannt haben.



**Abbildung 47: Datenarchitektur von Unternehmen welche die Vorteile von Datenmanagement erkannt haben<sup>186</sup>**

#### **8. Eine Unzureichende Datenqualität ist Hauptschuld, dass keine Big Data Projekte in Unternehmen gestartet werden.**

Eine schlechte Datenqualität wurde von 47 Teilnehmern als größte Herausforderung auf technischer Seite angegeben. Von den Auswahlmöglichkeiten der Herausforderungen auf technischer Seite ist dies somit die meist ausgewählte Antwort. Dahinter folgen die Antworten: Interpretation von Daten (45 Antworten), Zusammenarbeit mehrerer Systemkomponenten (39 Antworten) und fehlendes Know-how (33 Antworten). Es ist somit ein Trend erkennbar, dass die Datenqualität eines der größten technischen Probleme ist, wenn es um die Realisierung von Big Data Projekten geht. Die Stichprobengröße ist jedoch in diesem Fall zu klein und die ebenfalls hohe Anzahl der Teilnehmer welche andere Antworten ausgewählt haben zu groß, um eine klare Aussage treffen zu können, dass dies wirklich das Hauptproblem in der ganzen Industrie ist. Die Top-Antwort auf der technischen Seite ist die Projektkomplexität welche von 58 Teilnehmern als die größte Herausforderung auf geschäftlicher Seite angesehen wurde.

#### **9. Der Schwerpunkt der Unternehmen liegt derzeit in der Implementierung von Big Data Technologien.**

Frage 14 hat die derzeitigen Schwerpunkte im jeweiligen Unternehmen abgefragt, dabei waren die Anzahl der Antworten auf die Auswahloptionen sehr ausgeglichen. Die Top-Antworten in absteigender Reihenfolge waren:

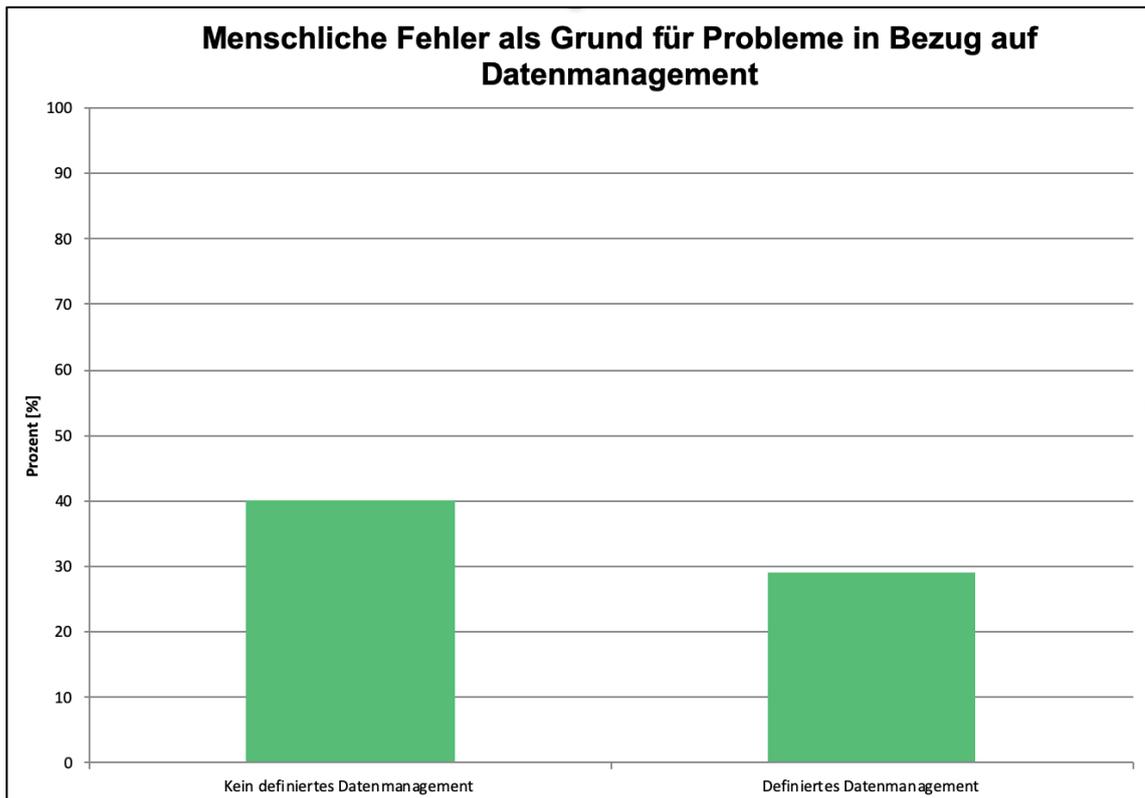
<sup>186</sup> Quelle: Eigene Darstellung.

- Jeweils 39,09 % also 43 Antworten für die Antwortmöglichkeit „Integration von heterogenen Datenquellen“ und für die „Komplexität der Datenarchitektur minimieren“
- 31,82 % also 35 Antworten für die Antwortmöglichkeit „Einfache Integration von neuen Datenquellen“
- 28,18 % also 31 Antworten für die Antwortmöglichkeit „Big Data“

Diese Verteilung zeigt sehr eindeutig, dass Big Data zwar durchaus eines der beherrschenden IT-Themen in Industrieunternehmen sind, es jedoch auch viele weitere Themen gibt, welche als genauso wichtig wahrgenommen werden und den Einsatz von Ressourcen des Unternehmens benötigen.

#### **10. Ein definiertes Datenmanagement schützt vor menschlichen Fehlern.**

13 Unternehmen, welche ein unternehmensweites definiertes Datenmanagement eingeführt haben, haben angegeben, dass menschliches Versagen einer der Gründe für Probleme in Bezug auf Datenmanagement ist (ca. 29 %). Die vorherrschenden Gründe für Probleme sind dabei Kommunikationsschwierigkeiten unterschiedlicher Abteilungen und fehlende Ressourcen (22 bzw. 19 Antworten). Von den 60 Unternehmen, welche diese Frage beantwortet haben und entweder nur in einzelnen Fachabteilungen oder gar kein Datenmanagement definiert haben, haben 24 (ca. 40 %) angegeben, dass menschliches Versagen einer der vorherrschenden Gründe für Probleme mit Datenmanagement ist (Abbildung 48). Die am meisten ausgewählte Antwort war „Kommunikationsschwierigkeiten unterschiedlicher Abteilungen“ (33 Antworten). In beiden Fällen wird menschliches Versagen als eines der Hauptprobleme in Bezug auf das Datenmanagement genannt. Prozentuell ist aber erkennbar, dass ein unternehmensweites Datenmanagement dieses Problem eindämmen kann. Dies verwundert nicht, da ein gut ausgearbeitetes Datenmanagement viele Prozesse und Zuständigkeiten fix definiert und somit weniger Freiräume für Fehler zulässt.

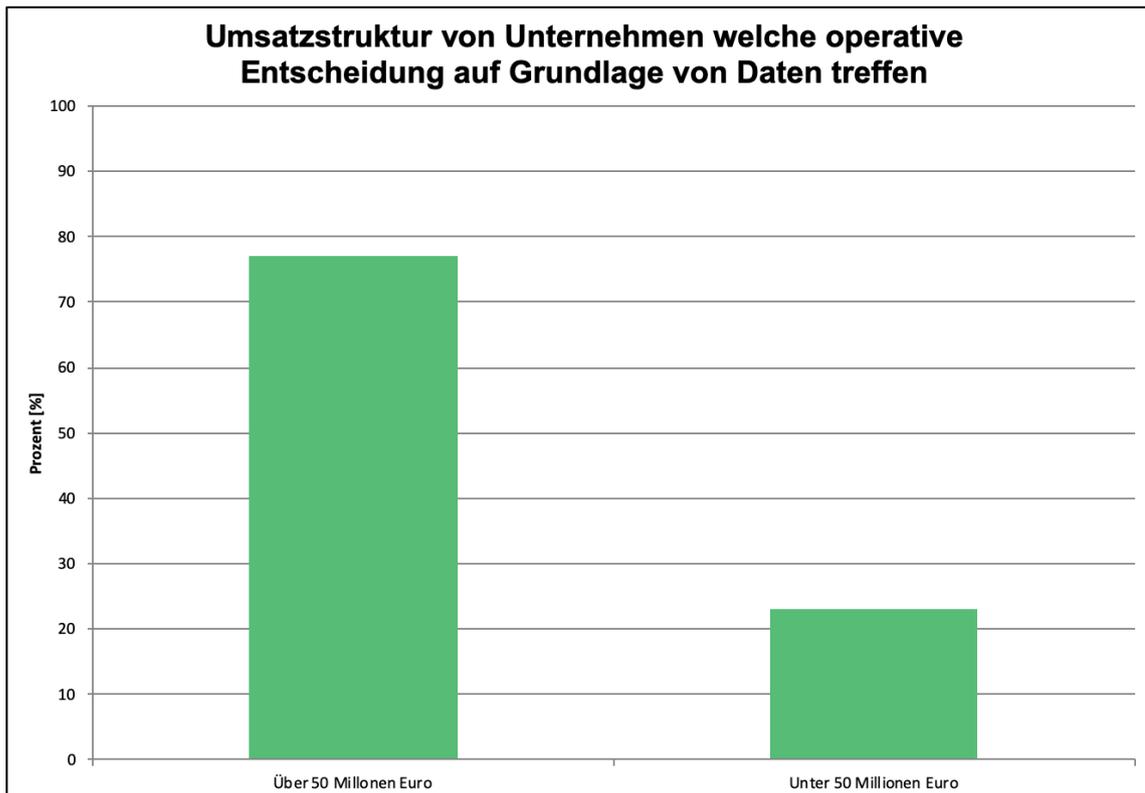


**Abbildung 48: Menschliche Fehler als Grund für Probleme in Bezug auf Datenmanagement<sup>187</sup>**

### **11. Unternehmen welche operativen Entscheidungen nicht auf Grundlagen von Daten treffen sind umsatzmäßig nicht erfolgreich.**

101 Unternehmen haben angegeben, dass operative Entscheidungen hauptsächlich auf Grundlage von Daten getroffen werden, von diesen Unternehmen haben 78 einen Umsatz zwischen 50 Millionen und mehr als 1,5 Mrd. Euro. Dies entspricht einem Prozentsatz von ca. 77 % (Abbildung 49). 11 Unternehmen haben diese Frage verneint, von diesen bewegen sich 7 in diesem Umsatzbereich, was einem Prozentsatz von ca. 63 % entspricht. Ein gewisser Trend kann von diesem Ergebnis abgelesen werden, jedoch kann nicht gesagt werden, ob der Umsatz durch den Gebrauch oder Nicht-Gebrauch von Daten für die operative Entscheidungsfindung verantwortlich ist, oder ob Unternehmen mit höheren Umsätzen professioneller aufgestellt sind und hier die Nutzung von Daten zur Entscheidungsfindung konzernweit vorgeschrieben ist.

<sup>187</sup> Quelle: Eigene Darstellung

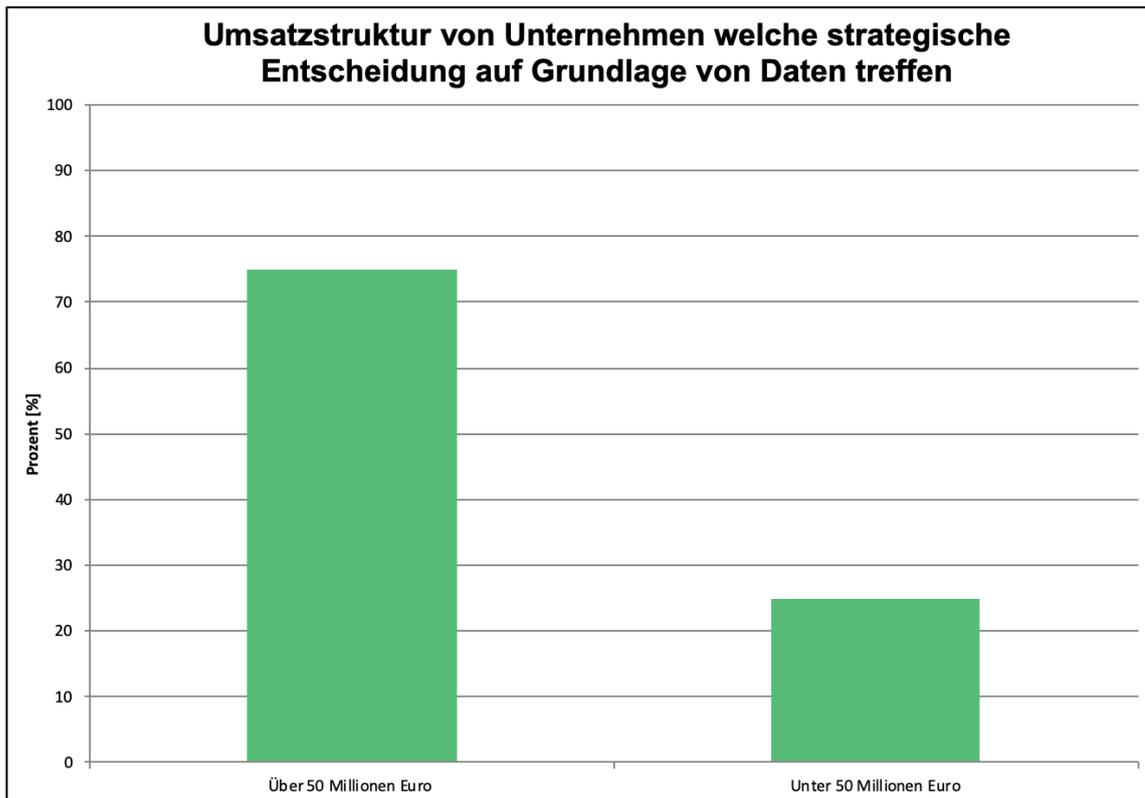


**Abbildung 49: Umsatzstruktur von Unternehmen welche operative Entscheidung auf Grundlage von Daten treffen<sup>188</sup>**

## **12. Unternehmen welche strategischen Entscheidungen nicht auf Grundlagen von Daten treffen sind umsatzmäßig nicht erfolgreich.**

17 Unternehmen haben angegeben, dass strategische Entscheidungen meist nicht auf Grundlage von Daten getroffen werden. Von diesen 17 Unternehmen haben 13 einen Umsatz von über 50 Millionen Euro. Dies entspricht einem Prozentsatz von ca. 76 %. 95 Unternehmen gaben an, dass strategische Entscheidungen auf Grundlage von Daten getroffen werden, davon haben 72 einen Umsatz von über 50 Millionen, was einem Prozentsatz von ca. 75 % entspricht (Abbildung 50). Bezugnehmend auf die oben erstellte Hypothese ist hier kein Trend zu erkennen. Der Prozentsatz der Unternehmen mit einem Umsatz von über 50 Millionen Euro ist in beiden Fällen annähernd gleich. Eine mögliche Erklärung dieses Umstandes ist, dass eine schlechte Kommunikation seitens der Geschäftsführung stattfindet und Entscheidung wenig oder gar nicht argumentiert werden, aber dass diese größtenteils sehr wohl auf Grundlage von Daten getroffen werden.

<sup>188</sup> Quelle Eigene Darstellung

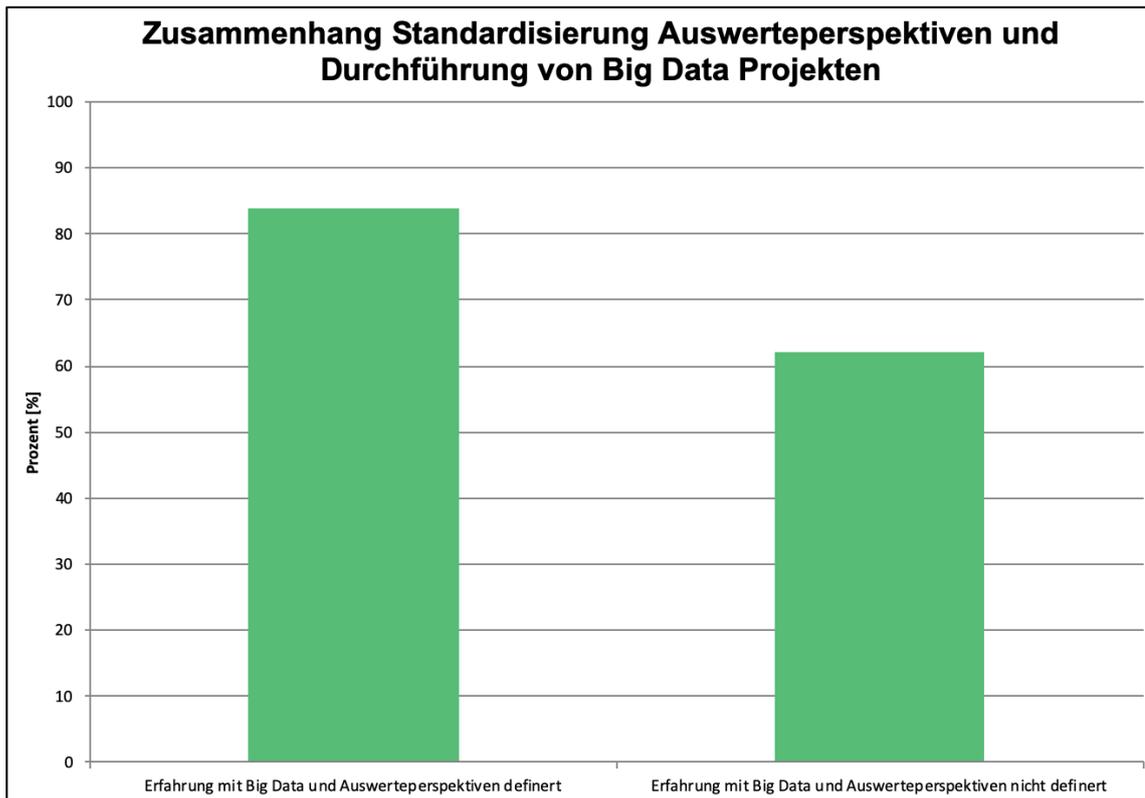


**Abbildung 50: Umsatzstruktur von Unternehmen welche strategische Entscheidung auf Grundlage von Daten treffen<sup>189</sup>**

### **13. Big Data Projekte werden nicht gestartet, weil Auswerteperspektiven und Key Performance Indicators nicht unternehmensweit definiert sind.**

45 Teilnehmer haben angegeben, dass im Unternehmen bereits Big Data Projekte stattgefunden haben, bei 61 Unternehmen ist das nicht der Fall. Von diesen Unternehmen haben 38 die Auswerteperspektiven unternehmensweit definiert (62 % der Teilnehmer mit undefiniertem Auswerteperspektiven). Bei den Unternehmen welche bereits Erfahrung mit Big Data Projekten haben, haben 38 der 45 Teilnehmer angegeben, dass die Auswerteperspektiven unternehmensweit einheitlich sind. Dies entspricht einem Prozentsatz von ca. 84 %, welcher klar über jenen der Unternehmen liegt, welche hier noch Aufholbedarf haben (Abbildung 51). Es zeigt sich das dieser Punkt ein essentieller Bestandteil ist, um erfolgreich Big Data Projekte durchführen zu können, da die Vergleichbarkeit und damit auch die leichtere Verwendbarkeit der zur Verfügung stehenden Daten gewährleistet wird.

<sup>189</sup> Quelle: Eigene Darstellung.



**Abbildung 51: Zusammenhang Standardisierung Auswerteperspektiven und Durchführung von Big Data Projekten<sup>190</sup>**

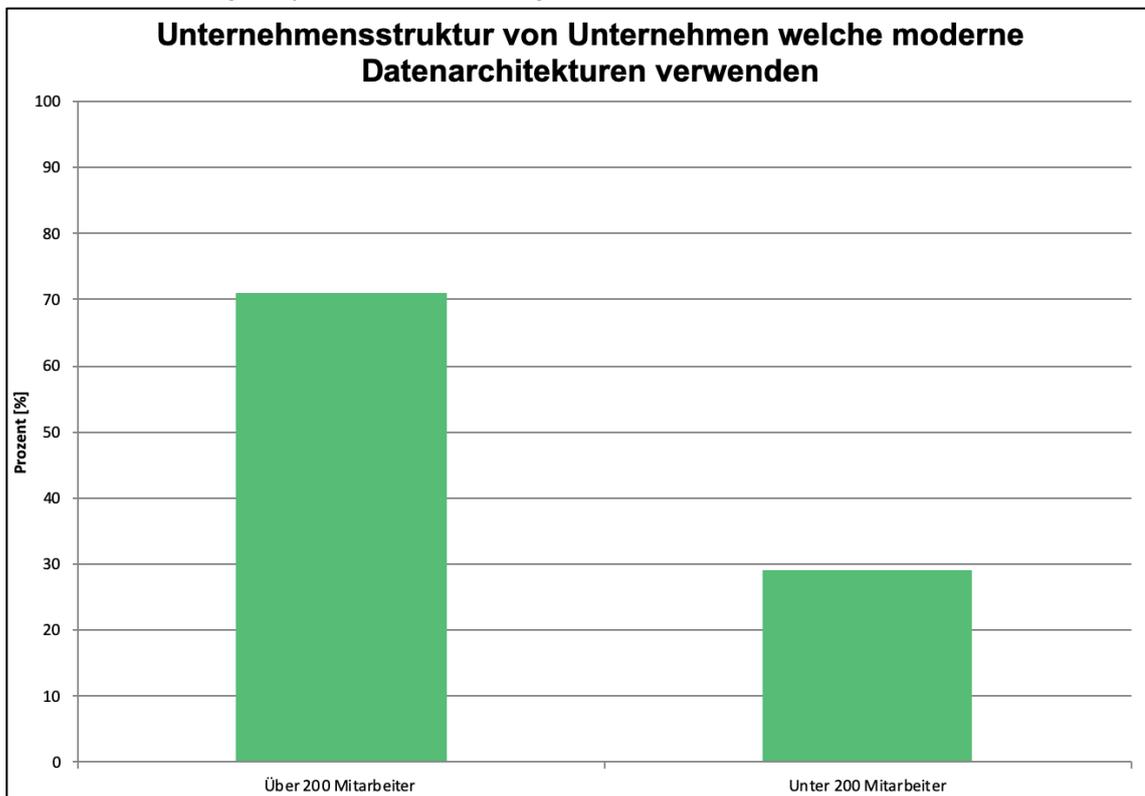
**14. Unternehmen welche moderne Datenarchitekturen wie z.B. DWH, NoSQL, Data Lakes bzw. Cloud Speichern arbeiten, setzen sich eher mit Big Data auseinander.**

Insgesamt haben 10 Teilnehmer angegeben, dass in ihren Unternehmen moderne Datenarchitekturen wie DWH mit Data Marts, NoSQL, Cloudspeicher oder Data Lakes eingesetzt werden. 5 von diesen Unternehmen haben bereits Big Data Projekte durchgeführt (ca. 50%). 93 Teilnehmer verwenden klassische Datenarchitekturen, davon haben 37, also ca. 40% schon Big Data Projekte durchgeführt. Es gibt also nur einen schwachen Zusammenhang zwischen dem Einsatz von modernen Datenarchitekturen und der Durchführung von Big Data Projekten erkennbar. Insgesamt in festzuhalten, dass sowohl moderne Datenarchitekturen wie auch komplexere Datenanalysen derzeit nur von sehr wenigen Unternehmen durchgeführt werden. Diese Analysen lassen sich mit den klassischen Architekturumgebungen nur schwerer realisieren und auch die Anpassungsmöglichkeiten dieser sind begrenzt. Was nicht bedeutet, dass DWH Architekturen vom Markt verdrängt werden, eher wird es Hybridlösungen geben in denen mehrere Architekturen parallel laufen werden, um so die größtmöglichen Vorteile für das Unternehmen lukrieren zu können.

<sup>190</sup> Quelle: Eigene Darstellung.

**15. Es besteht eine Korrelation zwischen gewählter Datenarchitektur und der Anzahl der Mitarbeiter in Österreich.**

Von den 93 Unternehmen welche klassische Speichermethoden wie DWH oder relationale Datenbanken verwenden haben 68 über 200 Mitarbeiter, was einem Prozentsatz von ca. 71 % entspricht. Von den 10 Unternehmen welche modernere Speichermethoden wie DWH mit Data Marts, Cloudspeicher oder Data Lakes verwenden, haben 8 (80 %) über 200 Mitarbeiter (Abbildung 52). Aufgrund der wenigen Industrieunternehmen in Österreich welche modernere Speichermethoden nutzen, lässt sich hierzu keine Aussage treffen. Es gibt eine große Zahl von sehr großen Unternehmen in Österreich welche keine der angegebenen modernen Speicherarten nutzen. Was jedoch im Umkehrschluss nicht bedeutet, dass die Speichermethode nicht trotzdem einem sehr fortschrittlichen Stand der Technik entspricht. Auffallend ist ferner, dass nur ein Unternehmen angegeben hat ein Data Lake zu verwenden, dieses Unternehmen über 10.000 Mitarbeiter und einen Umsatz von mehr als 1,5 Mrd. Euro hat, also zu den Big Playern in Österreich gehört.



**Abbildung 52: Unternehmensstruktur von Unternehmen welche moderne Datenarchitekturen verwenden<sup>191</sup>**

**16. Best-in-class Unternehmen haben bereits Erfahrung mit Big Data Projekten und erkennen die Vorteile in verschiedenen Anwendungsbereichen und Abteilungen.**

Als Gradmesser für ein best-in-class Unternehmen wurde ein Umsatz von über 0,5 Mrd. Euro genommen. Es werden somit die Branchenführer der großen

<sup>191</sup> Quelle: Eigene Darstellung.

Industriesektoren berücksichtigt, sicherlich gibt es kleinere Industriesektoren, bei welchen Unternehmen mit weniger Umsatz, Branchenführer sind. Diese Unternehmen werden in weiterer Folge nicht für die Korrelationsfindung eingerechnet. 23 Teilnehmer der Umfrage haben angegeben, dass ihre Unternehmen einen Umsatz von über 0,5 Mrd. Euro haben. 13 dieser Unternehmen haben Erfahrung mit Big Data Projekten und 4 haben mehr als drei Faktoren ausgewählt, bei welchen sie Vorteile durch das implementierte Datenmanagement sehen. Somit haben ca. 56 % Erfahrung mit Big Data Projekten und ca. 43 % gaben mehr als drei Vorteile durch ein definiertes Datenmanagement an. Von den anderen Teilnehmern haben 26 % bereits Erfahrung mit Big Data und ca. 35 % haben mehr als drei Vorteile durch ein definiertes Datenmanagement angegeben. Es ist somit ein Zusammenhang erkennbar, dass große Unternehmen die Vorteile eines implementierten Datenmanagements erkannt haben und auch eher erkannt haben das Big Data Projekte für die Wettbewerbsfähigkeit unverzichtbar sind und klare Vorteile bringen. Unterstützend haben diese die größten zur Verfügung stehenden Ressourcen, um diese Projekte realisieren zu können.

**17. Big Data Projekte führen langfristig zu hohem Umsatz.**

Von den 45 Unternehmen welche bereits Big Data Projekte durchgeführt haben, haben 15 einen Umsatz von unter oder gleich 100 Millionen Euro. Bei den Unternehmen, welche keine Big Data Projekte durchgeführt haben, sind 35 von 61 mit dem Umsatz gleich oder unter 100 Millionen. Prozentuell gesehen lautet das Ergebnis ca. 33 % zu 57 % (Abbildung 53). Hieraus lässt sich lesen, dass Big Data Projekte in Österreich derzeit eher von größeren Unternehmen durchgeführt werden, jedoch kann man daraus nicht auf eine Umsatzentwicklung der Firmen schließen. In unserer sehr wettbewerbsorientierten Wirtschaftswelt können mit Big Data Vorteile zur Konkurrenz aufgebaut werden, die wohl unbestritten sind. Streitbar ist nur wie sehr sich diese Vorteile auswirken und wie schnell sich diese Vorteile auch umsatzmäßig darstellen lassen. Darüber hinaus lässt sich nur schwer beurteilen ob nicht erst der hohe Umsatz die finanziellen Mittel für Big Data Projekte ermöglicht.

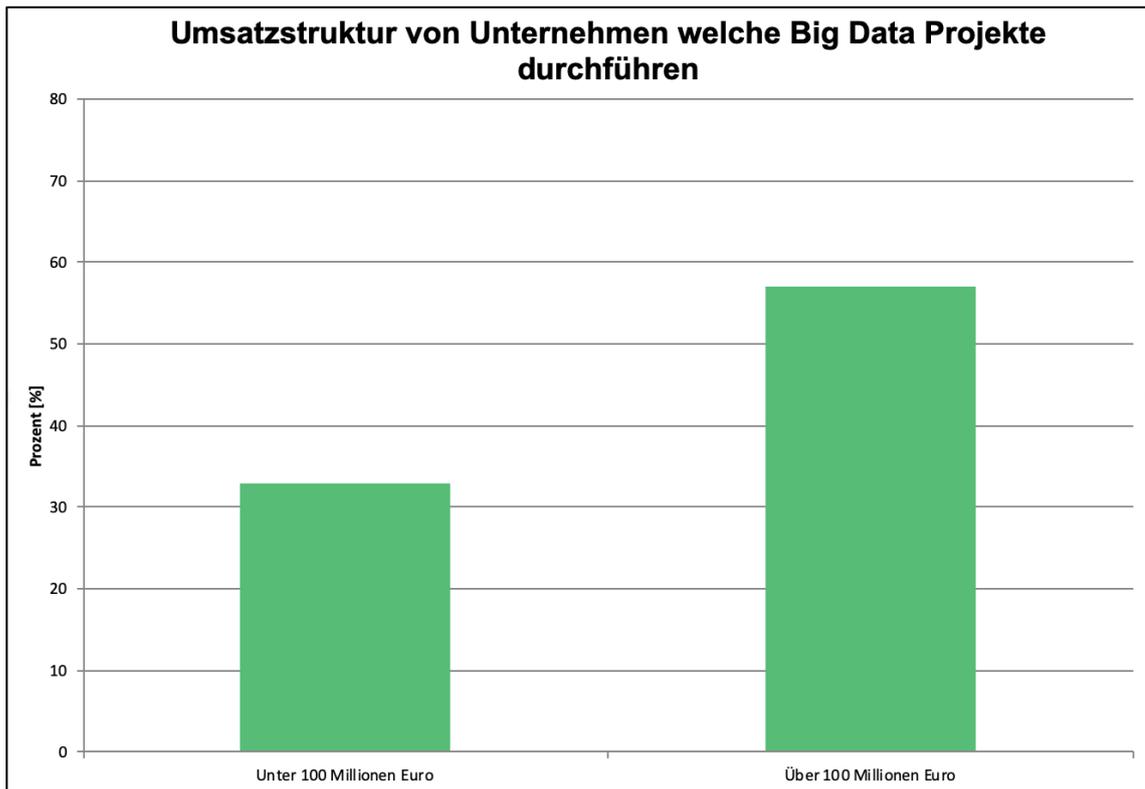


Abbildung 53: Umsatzstruktur von Unternehmen welche Big Data Projekte durchführen<sup>192</sup>

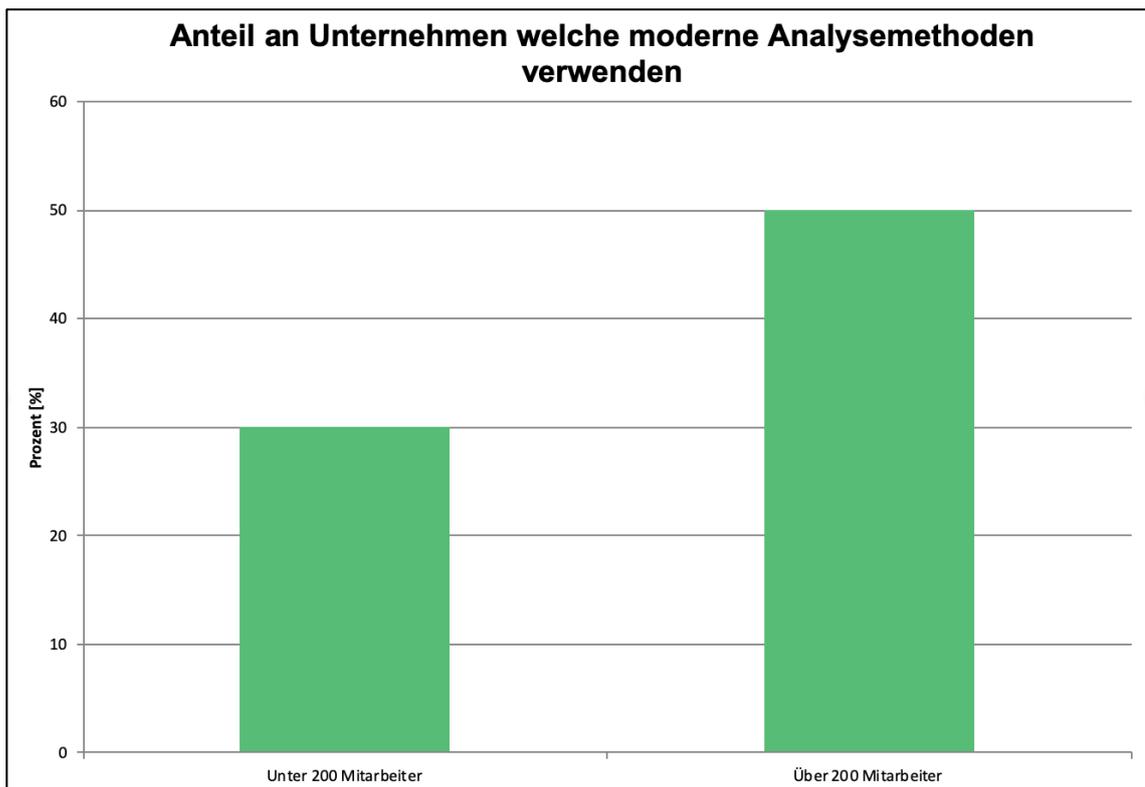
**18. Kleinere Unternehmen verwenden derzeit eher deskriptive und diagnostische Analysemethoden, während größere Unternehmen schon moderne Analysemethoden verwenden.**

Von den Unternehmen mit 1-50 Mitarbeitern, von welchen 7 bei der Umfrage teilgenommen haben, verwenden 2 moderne Analysemethoden und 2 Unternehmen etablierte Methoden wie die deskriptive und diagnostische Analyse (3 Unternehmen haben diese Frage nicht beantwortet). Unternehmen mit einer Mitarbeiteranzahl von 51-200 Mitarbeitern verwenden zu ca. 60 % die bereits am Markt etablierten Analysemethoden (17 von 28 Teilnehmern), nur 6 Teilnehmer verwenden moderne Analysemethoden wie die prädiktive oder präskriptive Analyse. Wenn wir diese zwei Auswahlmöglichkeiten als Grenze zwischen kleineren und größeren Unternehmen setzen, so verwenden 70 % klassische Analysemethoden welche vergangenheitsbezogen sind (Dieser Prozentsatz bezieht sich nur auf jene Unternehmen welche die entsprechende Unternehmensgröße haben und die Frage nach der verwendeten Analysemethode beantwortet haben). Von den Unternehmen mit über 200 Mitarbeitern, die meisten fallen wohl in die Kategorie Großunternehmen mit über 250 Mitarbeitern<sup>193</sup>, verwenden 37 klassische und 35 moderne Analysemethoden. Somit verwenden bei diesen Unternehmen ca. die Hälfte schon moderne Analysemethoden (Abbildung 54). Falls Unternehmen sowohl

<sup>192</sup> Quelle: Eigene Darstellung.

<sup>193</sup> Quelle: <https://www.wko.at/service/zahlen-daten-fakten/KMU-definition.html> (Zugriff: 29.09.2019).

klassische wie auch moderne Analysemethoden ausgewählt haben, wurden diese zu den Unternehmen gezählt welche moderne Analysemethoden verwenden. Es zeigt sich somit schon ein Trend, dass modernere Technologien, welche dementsprechend auch teuer im Einsatz sind, zuerst von großen Unternehmen verwendet werden. Wobei auch zu erkennen ist, dass kleinere Industrieunternehmen in Österreich dieses Thema nicht vernachlässigen und einige durchaus versuchen mit den großen Unternehmen Schritt zu halten. Wie schon in einer der vorigen Hypothesen beschrieben ist dieser Umstand wohl dem geschuldet, dass es spezialisierte Sektoren gibt in denen kleinere Unternehmen Marktführer sind und viel in moderne Technologien investieren, um diesen Marktvorteil zu erhalten.



**Abbildung 54: Anteil an Unternehmen welche moderne Analysemethoden verwenden<sup>194</sup>**

### **19. Die Zukunft der Datenanalyse liegt in der prädiktiven und präskriptiven Analyse.**

Derzeit nutzen nur 28,30 % der Teilnehmer die prädiktive Analyse und 25,47 % die präskriptive Analyse. Die Teilnehmer haben angegeben, dass sie schätzen, dass sich diese Zahlen in den nächsten 5 Jahren auf 68,87 % bzw. 63,21 % ändern. Es ist somit klar bewiesen worden, dass von Seiten der österreichischen Industrieunternehmen eine klare Vorstellung besteht, dass in den nächsten Jahren zukunftsbezogene Analysemethoden vorherrschend sein werden damit Maschinenfehler, Produktionspausen etc. erst gar nicht auftreten und geplante Unternehmensziele sicher erreicht werden.

<sup>194</sup> Quelle: Eigene Darstellung

## **20. Österreichische Industrieunternehmen setzen weiterhin auf klassische Datenquellen und verwenden sehr selten Social-Media-Daten oder Sensordaten.**

„Alte“ Datenquellen sind noch immer vorherrschend, was jedoch auch durch den Einsatz der Analysemethoden in Industrieunternehmen in Produktions- oder Lagerumgebungen erklären lässt. Die meisten Daten kommen aus Transaktionssystemen, also Datenbanken (86,79%), gefolgt von Betriebsdaten (72,64 %) und Maschinendaten (62,26 %). Jedoch ist auch hier klar erkennbar, dass bereits von vielen Unternehmen, Daten aus RFID Chips (24,53 %), Social-Media (9,43 %) oder Video- und Bilddateien (16,04 %) verwendet werden, um hier in den verschiedensten Bereichen einen Marktvorteil zur Konkurrenz zu erzielen. Der Einsatz von RFID Chips ermöglicht so z.B. die Kommunikation von Werkstück mit Industrierobotern um so den Arbeits- und Herstellungsprozess jederzeit kontrollieren und optimieren zu können. Das Wachstum an Daten ist aber nicht nur auf Unternehmen beschränkt, sondern gilt genauso für Privatanwender. Die dabei entstandenen Daten führen dazu, dass Unternehmen immer neue Möglichkeiten in Bereich von Direktmarketing, Cross- und Upselling erkennen und daraus neue Erkenntnisse gewinnen können.

Die Ergebnisse der Fragebogenerhebung und die Thesen, welche teilweise bewiesen worden sind oder wo keine Korrelation gefunden werden konnte, liefern einen guten Überblick über den aktuellen Stand des Datenmanagements in österreichischen Industrieunternehmen. Im nachfolgenden Kapitel sollen der Ergebnisse mit anderen Studien verglichen werden.

## **6.6 Vergleich mit anderen Studien**

Nachfolgend soll der Vergleich mit Vergleichsstudien gezogen werden und so Unterschiede, Vorteile und Nachteile zur österreichischen Industrie gezogen werden.

Da es in diesem Bereich eine begrenzte Anzahl von umfassenden Studien gibt, welche sich mit den gleichen Fragestellungen beschäftigt haben, wurden sowohl ältere wie auch neuere Studien als Vergleich herangezogen. Ein direkter Vergleich der gegenständlichen Fragebogenerhebung mit älteren Studien wie der BARC Studie 2014, zu Datenmanagement in der DACH Region<sup>195</sup> oder der BARC: Big Data Analytics Studie 2014<sup>196</sup> macht in der schnelllebigen IT nur zum Feststellen von Veränderungen im Lauf der Jahre Sinn. Als aktuelle Studien wurden die Bima Studie 2017/18 zu Business Intelligence und Advanced Analytics in der DACH Region<sup>197</sup>, der 2018 Global Data

<sup>195</sup> Vgl. BARC GmbH (2014).

<sup>196</sup> Vgl. BARC GmbH (2014).

<sup>197</sup> Vgl. BARC GmbH (2018).

Management Benchmark Report<sup>198</sup> und der Industrial Analytics Studie aus den Jahren 2016/2017<sup>199</sup> verwendet.

In der durchgeführten Fragebogenerhebung haben 22 der 141 Teilnehmer, bzw. der 117 Teilnehmer, welche diese Frage beantwortet haben, angegeben, dass sie über eine eigene Stelle verfügen, welche die Datenverantwortlichkeit regelt. Dies entspricht 18 % der Teilnehmer, welche diese Frage beantwortet haben. Zum Vergleich laut dem 2018 Global Management Benchmark Report haben in den USA bereits 10 % einen CDO, 20% einen CIO und 14 % einen eigenen Datenverantwortlichen<sup>200</sup>. Wobei auch hier in 51 % der Fälle die IT-Abteilung für Daten im Unternehmen verantwortlich ist<sup>201</sup>. Dies zeigt, dass Daten in Amerika ein höherer Stellenwert zugerechnet wird, vor allem weil vermehrt C-Level Positionen geschaffen werden, welche in der Durchsetzung von neuen Projekten oder Prozessen deutlich höheren Einfluss auf die ganze Organisation haben.

Die Frage ob ein unternehmensweites Datenmanagement implementiert wurde, haben 38,26 % der Teilnehmer aus österreichischen Industrieunternehmen mit der Antwort „Bereits fix im Unternehmensprozess integriert“ beantwortet. 2014 wurde eine ähnliche Frage im Zuge der BARC Studie 2014 bezüglich der festgelegten Datenstrategie in Unternehmen gestellt, hier haben 2014, nur 10 % der Unternehmen aus der DACH-Region angegeben, dass eine zentrale Datenstrategie existiert, die allen Daten umfasst<sup>202</sup>. Es ist also in den fünf Jahren, welche zwischen den Fragebogenerhebungen liegen ein deutlicher Anstieg von Unternehmen zu beobachten welche die Datenstrategie nicht nur in einzelnen Unternehmen verwenden, sondern diese unternehmensweit definiert haben. Laut dem Industrial Analytics Report haben 68 % der Unternehmen, welche aus der ganzen Welt an dieser Umfrage teilgenommen haben, eine unternehmensweit definierte Datenstrategie in Bezug auf die Datenanalyse<sup>203</sup>. Österreichische Industrieunternehmen haben somit seit 2014 ordentlich in Bezug auf das Datenmanagement aufgeholt, sind aber in Anbetracht dieses Ergebnisses unter dem weltweiten Durchschnitt (Abbildung 55).

---

<sup>198</sup> Vgl. Experian (2018).

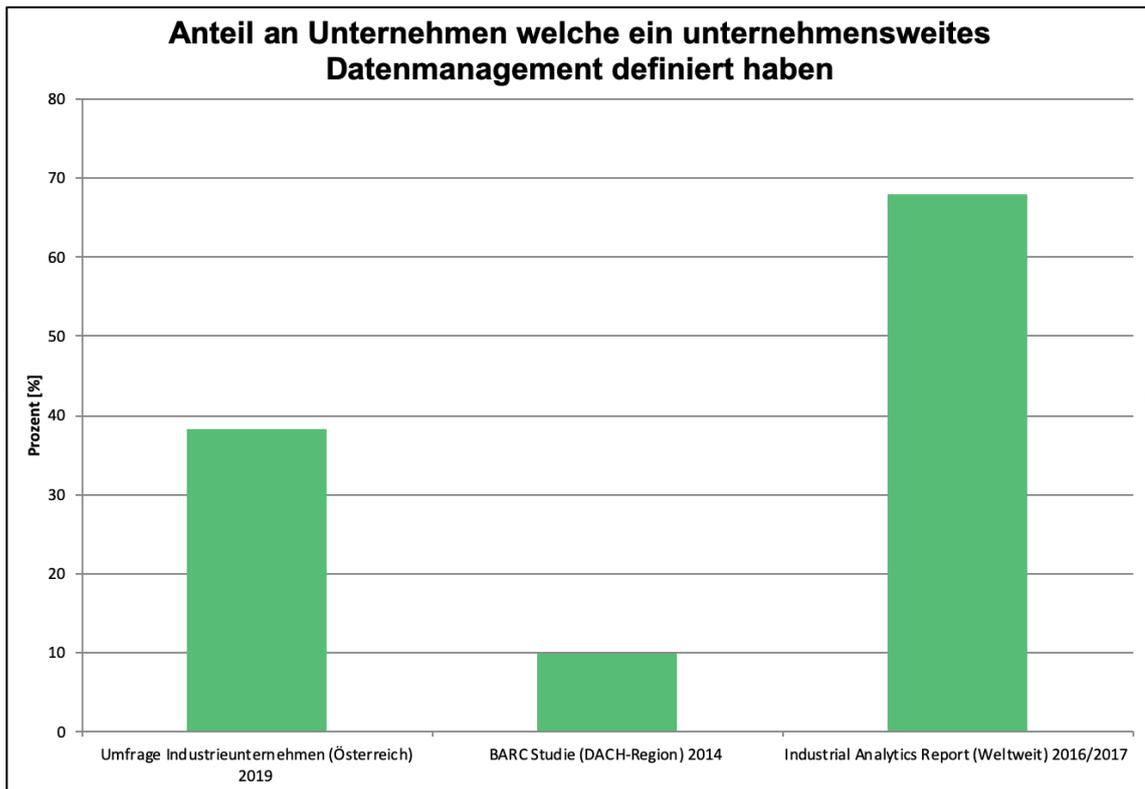
<sup>199</sup> Vgl. IoT Analytics (2016).

<sup>200</sup> Vgl. Experian (2018), S.15.

<sup>201</sup> Vgl. Experian (2018), S.15.

<sup>202</sup> Vgl. BARC GmbH (2014), S. 13.

<sup>203</sup> Vgl. IoT Analytics (2016), S. 17.



**Abbildung 55: Anteil an Unternehmen welche ein unternehmensweites Datenmanagement definiert haben<sup>204</sup>**

Bei Frage 9 wurde nach den Faktoren gefragt, wo das definierte Datenmanagement gefühlt die meisten Vorteile gebracht hat, hier haben über 50% der Teilnehmer die Bereitstellung von Daten und die Datenqualität genannt. 2014 waren in der Barc Studie die meistgenannten Antworten „Datenintegration“ und „Data Warehouse“<sup>205</sup>. Es ist davon auszugehen, dass diese Punkte durch Investitionen in die IT-Infrastruktur in der Zwischenzeit in den meisten Unternehmen deutlich besser geworden sind und somit andere Vorteile als aktueller empfunden werden.

Als größte Probleme in Bezug auf das Datenmanagement wurden in der gegenständlichen Erhebung „Kommunikationsschwierigkeiten unterschiedlicher Abteilungen“, „Fehlende Ressourcen“ und „Menschliches Versagen“ am häufigsten genannt. Auch bei der BARC Studie 2014 war die Topantwort zu den größten Herausforderungen in Bezug auf Datenmanagement „Unzureichende Ressourcen“, dahinter wurde die fehlende Bereitschaft zu Änderung genannt<sup>206</sup>. Daraus lässt sich schließen, dass immer mehr Unternehmen nun die Notwendigkeiten in Bezug auf das Thema Daten erkannt haben, aber gefühlt noch immer zu wenig investiert wird, um dieses Thema aus Sicht der Teilnehmer korrekt umzusetzen.

Aus der BARC Studie 2014 geht ebenfalls hervor, dass nur 41 % damals angaben, dass Informationen einen hohen Stellenwert bei der Entscheidungsfindung haben, aber 61 % haben damals auch angegeben, dass sich dieser Wert voraussichtlich in den nächsten

<sup>204</sup> Quelle: Eigene Darstellung, BARC GmbH (2014), S. 13., IoT Analytics (2016), S. 17.

<sup>205</sup> Vgl. BARC GmbH (2014), S. 14.

<sup>206</sup> Vgl. BARC GmbH (2014), S. 22.

Jahren deutlich erhöhen wird<sup>207</sup>. Dies ist jedenfalls eingetreten, da in der aktuellen Umfrage 90,99%, angegeben haben, dass operative Entscheidungen auf Grund von Daten getroffen werden und 83,93 % haben angegeben, dass strategische Entscheidungen auf Grundlage von Daten getroffen werden. Ebenfalls wurde folgendes Thema bestätigt, 2014 wurde innerhalb der gleichen Studie angegeben, dass die Integration von heterogenen Datenquellen das Hauptproblem in Bezug auf Datenmanagement ist<sup>208</sup>, dieses Thema wurde im durchgeführten Fragebogen als der meistgenannte Schwerpunkt genannt, welcher derzeit in den Unternehmen gesetzt wird. Laut der Bima Studie 2017/2018 haben nur 29 % angegeben, dass Auswerteperspektiven unternehmensweit standardisiert sind<sup>209</sup>, bei der Umfrage welche im Zuge dieser Arbeit durchgeführt wurde lag dieser Wert bei 72,32 %. Dies entspricht einem deutlichen Unterschied, welcher sich entweder durch eine andere Auslegung der Fragestellung erklären lässt, oder darauf zu schließen ist, dass Industrieunternehmen aufgrund der durchgeführten Tätigkeiten, schon längere Zeit KPI standardisieren, weil sich diese auch leichter standardisieren lassen. So wurden z.B. wohl die Produktionsmenge oder Stehzeiten einer Maschine schon lange erfasst und ausgewertet.

Laut der durchgeführten Umfrage verwenden unter 10% der österreichischen Industrieunternehmen moderne Speichermedien oder Datenarchitekturen wie z.B. Data Lakes, Cloudspeicher, NoSQL oder DWH mit Data Marts. Aus der BARC Studie ist ersichtlich, dass dieser Wert bereits 2014 in der DACH Region deutlich höher war. Hier wurden Data Marts mit zentralen DWH von 38% aller Teilnehmer genutzt und Data Lakes bereits von 3% aller Teilnehmer<sup>210</sup> (im Gegensatz zu 0,94 % bei der gegenständlichen Umfrage). Bezüglich Architekturlandschaften sind somit österreichische Industrieunternehmen eindeutig nicht auf dem neuesten Stand, bzw. wird hier sehr konservativ mit der Verwendung neuer Technologien umgegangen (Abbildung 56).

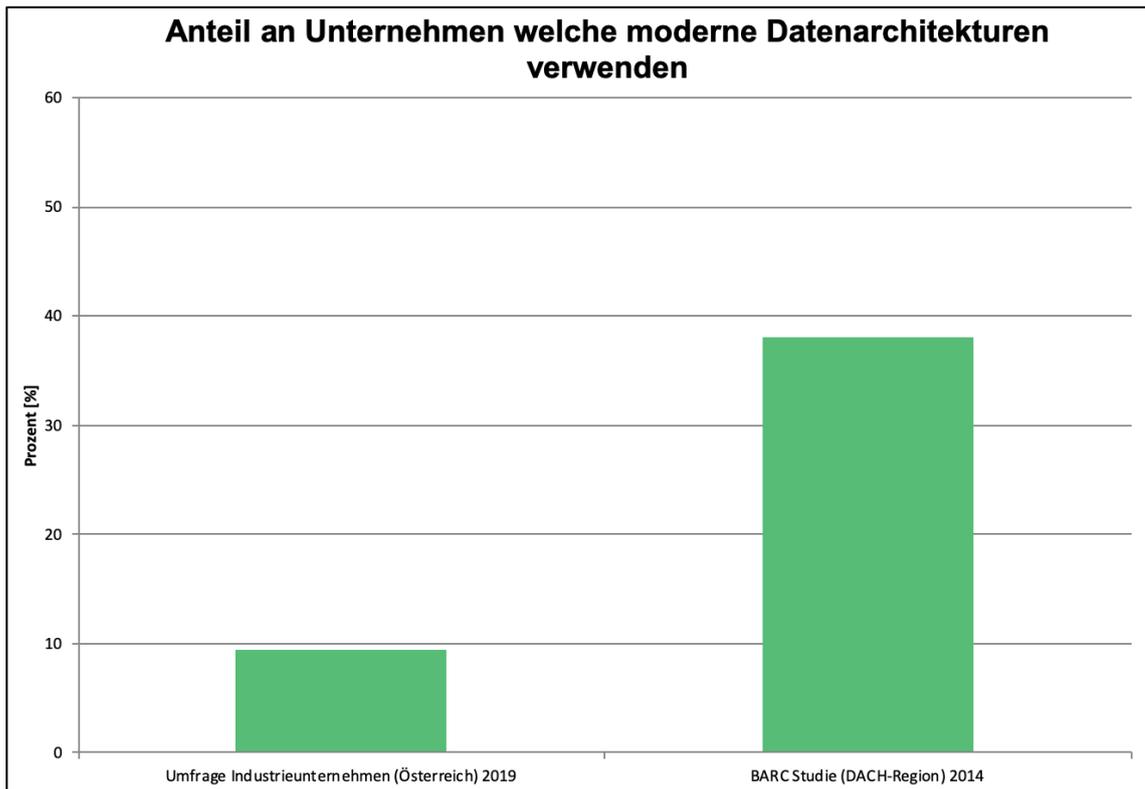
---

<sup>207</sup> Vgl. BARC GmbH (2014), S. 12.

<sup>208</sup> Vgl. BARC GmbH (2014), S. 17.

<sup>209</sup> Vgl. BARC GmbH (2018), S. 44.

<sup>210</sup> Vgl. BARC GmbH (2014), S. 23.



**Abbildung 56: Anteil an Unternehmen welche moderne Datenarchitekturen verwenden<sup>211</sup>**

In der BARC Studie 2014 haben nur 6% der Unternehmen angegeben, dass Big Data ein wichtiges Thema im Unternehmen sei<sup>212</sup>, ebenfalls sagten 22% im Rahmen der BARC: Big Data Analytics Studie im Jahre 2014, dass keine Big-Data-Initiative geplant sei nur 12% sagten, dass bereits Big Data Projekte im Unternehmen umgesetzt werden<sup>213</sup>. Ebenfalls gering ist die Zahl der Unternehmen, welche im Industrial Analytics 2016/2017 Report angegeben haben, dass bereits komplexere Datenanalysen durchgeführt wurden, hier lag die Zahl bei nur 30%<sup>214</sup>. Ausgehend von den Daten der in Österreich durchgeführten Studie, haben Unternehmen aus der österreichischen Industrie hier sogar einen Innovationvorsprung, da 42,65 % bereits solche Projekte umgesetzt haben. Es muss jedoch angemerkt werden, dass in Bezug auf dieses Thema ein Jahr eine lange Zeit darstellt, in der sich auch in der DACH-Region diese Zahlen deutlich erhöht haben können. Während jedoch in Österreich die Anwendungsgebiete von einem Großteil im Vertrieb und in der Produktion gesehen werden, sagen fast 50 % in der Bima Studie 2017/18 aus, dass die meisten Vorteile im Finanzbereich erreicht werden können<sup>215</sup>. Dieser Umstand ist ganz klar darauf zurückzuführen, dass in der gegenständlichen Studie nur Industrieunternehmen teilgenommen haben, bei welchen dieser Bereich kein Kernthema darstellt. Im Industrial Analytics Report 2016/17 welcher Fragen dazu an Industrieunternehmen aus der ganzen Welt gestellt hat, haben 45 % der

<sup>211</sup> Quelle: Eigene Darstellung, BARC GmbH (2014), S. 23.

<sup>212</sup> Vgl. BARC GmbH (2014), S. 10.

<sup>213</sup> Vgl. BARC GmbH (2014), S. 17.

<sup>214</sup> Vgl. IoT Analytics (2016), S. 17.

<sup>215</sup> Vgl. BARC GmbH (2018), S. 45.

Teilnehmer den Bereich der prädiktiven und präskriptiven Wartung von Maschinen als wichtigstes Thema in diesem Bezug gesehen<sup>216</sup>. Die Top-Antworten der aktuell durchgeführten Studie waren die „Bessere Steuerung von operativen Prozessen“ und die „Verbesserung der Produktionsplanung“. Hier zeigt sich, dass prädiktiven und präskriptiven Analysen in Österreich vorerst noch nicht von besonders vielen Unternehmen durchgeführt werden und andere Dinge aktuell deshalb noch wichtiger erscheinen. So haben nur jeweils ca. 25 % der Unternehmen angegeben, dass diese Datenanalyse derzeit schon durchgeführt werde.

---

<sup>216</sup> Vgl. IoT Analytics (2016), S. 17.

## 7 Zusammenfassung und Ausblick

Nur mit einem unternehmensweiten Datenmanagement ist eine Wahrung der Datenqualität, ein strukturierter Lebenszyklus der Daten, eine Zugriffssicherung, eine Maximierung des Wertes von Daten und eine Sicherstellung, dass diese zu rechten Zeit bei der richtigen Person im Unternehmen gelangen möglich. Für ein funktionierendes Datenmanagement braucht es neben den organisatorischen und technischen Strukturen auch die Etablierung von Datenverantwortlichen.

Ein Großteil der Industrieunternehmen in Österreich haben zumindest in einigen Fachabteilungen ein Datenmanagement implementiert. Es ist anzunehmen, dass diese Zahl auch weiterhin stetig wachsen wird, da nur so eine Datenqualität gewährleistet werden kann, die für komplexere Datenanalysen notwendig ist. Aufholbedarf gibt es in Österreich in Bezug auf die eindeutige Regelung der Zuständigkeiten für Daten. Weltweit wird diesen Stellen eine weitaus höhere Bedeutung zugeschrieben, in Österreich liegt die Verantwortung meist noch bei der IT-Abteilung. Eine strategische Ausrichtung und eine auf das Datenmanagement ausgelegte Budgetierung wird somit erschwert. Nur wenn diese Investitionen getätigt werden und die organisatorischen Voraussetzungen geschaffen werden, ist eine sinnvolle Umsetzung von Big Data bzw. komplexeren Datenanalyse überhaupt möglich.

Kommunikation wird noch immer als das Hauptproblem in Bezug auf das Datenmanagement genannt, auch hierfür sind klare Kommunikationswege und Zuständigkeiten notwendig, welche von einer Person getroffen werden sollten, welche im Unternehmen eine notwendige Entscheidungskompetenz hat. Überhaupt wird in Zukunft eine noch stärkeres Bekenntnis zum Thema Daten und Datenmanagement notwendig sein.

Die Durchführung von Big Data Projekten, welche durch die Masse, Vielfalt und die Geschwindigkeit von Daten definiert sind wird in den nächsten Jahren eine der größten Herausforderungen, aber auch Chancen für Unternehmen darstellen. Komplexere Datenanalysen haben zwar häufig einen großen wirtschaftlichen Impact, aber um diese erzielen zu können, müssen zuerst erhebliche Investitionen getätigt werden um die notwendigen Strukturen und die Infrastruktur im Unternehmen zu schaffen. Das Finden eines geeigneten Business Cases der diese Investitionen rechtfertigen würde, ist eines der Hauptargumente, welches auf organisatorischer Seite gegen die Durchführung von Big Data Projekten spricht. Hier muss seitens der Unternehmensführung ein Umdenken stattfinden und ein deutlich längerer Zeithorizont betrachtet werden, wenn es um die Beurteilung solcher Projekte geht. In den seltensten Fällen können damit nämlich kurzfristige Gewinne erzielt werden, aber häufig über einen längeren Zeitraum einen Marktvorteil gegenüber der Konkurrenz aufgebaut werden. Außerdem kann das durch ein Projekt gewonnene Know-how in Folgeprojekten eingesetzt werden, was insgesamt die Qualität der Datenanalysen erhöhen sollte.

Österreichische Industrieunternehmen scheinen aber zumindest in Bezug auf die Standardisierung der Auswerteperspektiven und KPIs sehr gut aufgestellt zu sein, da

diese in den meisten Industrieunternehmen schon standardisiert werden. Um die Daten über mehrere Abteilungen hinweg vergleichen und in weiterer Folge auch analysieren zu können, ist dieser Schritt unabdingbar.

Auch sind sich nahezu alle Unternehmen bewusst, dass die Zukunft der Datenanalyse in komplexeren Datenauswertungen wie der prädiktiven und präskriptive Datenanalyse liegt. Allerdings werden diese Analysemethoden derzeit nur von einer sehr überschaubaren Anzahl von Unternehmen eingesetzt und auch nicht als derzeitiger Schwerpunkt in der IT gesehen. Bei weltweit durchgeführten Umfragen ist dies klar eines der Kernthemen, in welches im Moment am meisten investiert wird. Hier müssen österreichische Industrieunternehmen aufpassen, dass sie nicht den Anschluss verlieren. Auch das volle Potenzial und die Anwendungsgebiete in den verschiedenen Abteilungen scheinen vielen Unternehmen noch nicht vollständig bewusst zu sein. So werden von den meisten Unternehmen nur jene Abteilungen genannt wo derzeit viele industrielle Datenanalysen durchgeführt werden (Vertrieb und Produktion). Wie im Abschnitt 4.1 beschrieben ergeben sich durch Big Data jedoch völlig neue Wege Daten analysieren zu können und bis dato nicht dagewesene Erkenntnisse und Theorien daraus aufstellen zu können. Für diese Art der Analysen werden jedoch in den meisten Fällen auch Investitionen in neue Datenarchitekturen notwendig, da bestehende Data Warehouse Umgebungen für diese Art der Analysen oftmals nicht ausgerichtet sind und sich auch nur schwer anpassen lassen.

Insgesamt kann ein gemischtes Urteil für die österreichische Industrie in Bezug auf das Datenmanagement ausgestellt werden. Gut sind die Unternehmen in Bezug auf die Standardisierung und das Datenmanagement aufgestellt. Aufholbedarf gibt es bei den verwendeten Datenarchitekturen, der Erfahrung mit Big Data Projekten und den gewählten Schwerpunkten, welche größtenteils nicht auf der Erarbeitung von komplexeren Datenanalysen liegen. Klar zu erkennen ist ebenfalls, dass best-in-class Unternehmen hier schon weiter sind wie die Konkurrenz und dadurch ihre Stellung wohl in Zukunft noch weiter festigen werden.

## Literaturverzeichnis

- A.T.Kearney (2013): Big data and the creative destruction of today's business models.  
URL:  
<https://www.atkearney.com/documents/10192/698536/Big+Data+and+the+Creative+Destruction+of+Todays+Business+Models.pdf/f05aed38-6c26-431d-8500-d75a2c384919> (Zugriff: 10.12.2018).
- Aunkofer, B. (2017), Entscheidungsbaumverfahren. URL: <https://data-science-blog.com/blog/2017/02/13/entscheidungsbaumverfahren-artikelserie/> (Zugriff: 06.02.2019).
- Bankhofer, U; Vogel, J. (2008): Datenanalyse und Statistik. Gabler Verlag. ISBN: 978-3-8349-0434-8. S. 273.
- BARC GmbH (Hrsg.) (2014): Big Data Analytics, Würzburg.
- BARC GmbH (Hrsg.) (2014): Datenmanagement im Wandel, Würzburg.
- BARC GmbH (Hrsg.) (2018): Zeit für eine neue Kultur durch Business Intelligence & Advanced Analytics, Würzburg.
- Berger, Doris (2010): Wissenschaftliches Arbeiten in den Wirtschafts- und Sozialwissenschaften. Springer Verlag. ISBN: 978-3-8349-2285-4. S. 156-158.
- Bundesministerium für Bildung und Forschung (Hrsg.) (2017): Industrie 4.0, S. 66.
- Bodendorf, F. (2005): Daten und Wissensmanagement. Springer Berlin Heidelberg New York. ISBN 3-540-28743-4, S.1-4.
- Bourg, A.; Paulus, L.; Reise, M. (2006): Neuronale Netze. Seminararbeit, Institut für Psychologie, Leopold-Franzens-Universität Innsbruck. S. 6-13.
- Brüggemann, H.; Bremer, P. (2015): Grundlagen Qualitätsmanagement. 2. Auflage. Springer Verlag. ISBN: 978-3-658-09221-4, S. 3.
- Cleff, T. (2008): Deskriptive Statistik und moderne Datenanalyse. 2. Auflage. Gabler Verlag. ISBN: 978-3-8349-3221-1, S. 4-147.
- Cloudt, T. (2001): 3k. Die Qualitätsregelkarte. URL:  
[https://www.cloudt.de/pdf\\_archiv/3kregelkar.pdf](https://www.cloudt.de/pdf_archiv/3kregelkar.pdf) (Zugriff: 01.02.2019).
- Diebold, F. X. (2000): „Big Data“ Dynamic Factor Models for Macroeconomic Measurement and Forecasting. S. 115.
- Die Presse (2019): Chemische Industrie: „Hoffen dass die nächste Regierung diesen Kurs fortsetzt“. URL:  
[https://diepresse.com/home/wirtschaft/economist/5635107/Chemische-Industrie\\_Hoffen-dass-die-naechste-Regierung-diesen](https://diepresse.com/home/wirtschaft/economist/5635107/Chemische-Industrie_Hoffen-dass-die-naechste-Regierung-diesen) (Zugriff: 22.09.2019).

- Eckey, H.-F.; Kosfeld, R.; Rengers M. (2002): Multivariate Statistik. Gabler Verlag. ISBN: 978-3-322-84476-7. S. 93-110.
- Ernst & Young (Hrsg.) (2016): Internet of Things, London.
- Experian (Hrsg.) (2018): The 2018 global data management benchmark report, Boston.
- Fayyad, U.; Piatetsky-Sharir, G.; Smyth, P. (1996): From Data Mining to Knowledge Discovery in Databases. In: AI Magazine, Jg. 1996, Nr. 3, S. 37–54.
- FFG (2019): KMU Definition. URL: [https://www.ffg.at/recht-finanzen/rechtliches\\_service\\_KMU](https://www.ffg.at/recht-finanzen/rechtliches_service_KMU) (Zugriff: 30.08.2019).
- Forbes (2019): How much data do we create every day?. URL: <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#5654930d60ba> (Zugriff: 30.10.2019).
- Franzki, C. (2012): Hypothesentest. URL: <http://www.mathematik-wissen.de/hypothesentest.htm> (Zugriff: 01.02.2019).
- Freie Universität Berlin (2019): Wikis der freien Universität Berlin. URL: <https://wikis.fu-berlin.de/pages/viewpage.action?pageId=712409813> (Zugriff: 30.01.2019).
- Freiknecht, J.; Papp, S. (2018): Big Data in der Praxis. Zweite Ausgabe. Hanser Verlag. ISBN: 978-3-446-45601-3, S. 23.
- Gabriel, R.; Gluchowski, P.; Pastwa, A. (2009): Data Warehouse & Data Mining. W3L-Verlag. ISBN: 3-937137-66-7. S. 2-7.
- Gartner (2019): It-Glossary. URL: <https://www.gartner.com/it-glossary/big-data/>, (Zugriff: 11.01.2019).
- Gandomi, A.; Haider, M. (2014): Beyond the hype. Big data concepts, methods, and analytics. In: International Journal of Information Management, Jg. 2015, Nr. 35, S.138-139.
- Gleißner, W. (2014): Wahrscheinlichkeiten, Bayes-Theorem und statistische Analysen. In: Controller Magazin, Jg. 2014, Nr. 2, S. 68-74.
- Gordon, K. (2013): The Principles of Data Management. Facilitating Information Sharing Edition. Second Edition. Bcs The Chartered Institute for IT. ISBN 978-1-78017-185-2, S.57-66.
- Gorelik, A. (2019): The Enterprise Big Data Lake. O'Reilly Verlag. ISBN: 978-1-491-93155-4, S. 2-24.
- Gräf, Lorenz (2010): Online-Befragung. Eine praktische Einführung für Anfänger. Band 3. LIT Verlag. ISBN: 978-3-8258-9771-0. S. 58-65.
- Han, J.; Kamber, M.; Pei, J. (2012): Data Mining. Third Edition. Morgan Kaufman. ISBN: 978-0-12-381479-1. S.3-10.

- Herrmann, J; Fritz, H. (2016): Qualitätsmanagement. 2. Auflage. Hanser Verlag. ISBN: 978-3-446-44043-2. S. 151.
- Hudec, M. (2011): Anwendung von Ensemble Methoden für Klassifikationsaufgaben. URL: <http://www.stat.tugraz.at/Statistiktage2011/PHudec.pdf>. (Zugriff: 18.05.2019).
- Hurwitz, J.; Kirsch, D. (2018): Machine Learning For Dummies, IBM Limited Edition. John Wiley & Sons, Inc. ISBN: 978-1-119-45494. S. 13-19.
- IBM (Hrsg.) (2012): Analytics: Big Data in der Praxis.
- IBM (2018): Diagnostic analytics: Why did it happen. URL: <https://www.ibm.com/blogs/business-analytics/diagnostic-analytics-101-why-did-it-happen/> (Zugriff: 31.01.2019).
- IBM (2013): 2,5 quintillion bytes of data created every day. URL: <https://www.ibm.com/blogs/insights-on-business/consumer-products/2-5-quintillion-bytes-of-data-created-every-day-how-does-cpg-retail-manage-it/> (Zugriff: 16.11.2018).
- IOT Analytics (Hrsg.) (2014): Industrial Analytics 2016/2017, Hamburg.
- Jäggi, S; Portmann, C. (2012): Kommunikation in Marketing und Verkauf: Grundlagen mit zahlreichen Beispielen, Repetitionsfragen mit Antworten und Glossar. 3. Auflage. Compendio Bildungsmedien. ISBN: 978-3715596433. S. 193
- Konen, W. (2009): Einführung Data Mining. URL: <http://www.gm.fh-koeln.de/~konen/WPF-DM-Cup/01-Einleitung-CRISP.PDF> (Zugriff: 16.01.2019).
- Krcmar, H. (2015): Informationsmanagement. 6.Auflage. Springer Verlag. ISBN 978-3-662-45862-4. S. 11-59.
- Krypczyk, V; Bochkor, O., (2018): Der Blick in die Glaskugel: Predictive Analytics in Theorie und Praxis. URL: <https://entwickler.de/online/development/predictive-analytics-praxis-tipps-579847089.html> (Zugriff: 04.02.2019).
- Laudon, K. C.; Laudon, J. P.; Schoder, D. (2016): Wirtschaftsinformatik, Dritte Auflage. Pearson Verlag. ISBN: 978-3-86326-765-0, S. 291-292.
- LeBlanc, D. (2004): Statistics. Concept and Applications for Science. Joint & Barlett Publishers. ISBN: 978-0763717292. S. 48-49.
- Leser, U. (2005): Data Warehousing, Vorlesung, Sommersemester 2005. URL: [https://www.informatik.hu-berlin.de/de/forschung/gebiete/wbi/teaching/archive/sose05/dwh/02\\_architektur\\_prozesse.pdf](https://www.informatik.hu-berlin.de/de/forschung/gebiete/wbi/teaching/archive/sose05/dwh/02_architektur_prozesse.pdf). (Zugriff: 14.01.2019).
- Metalltechnische Industrie (2019): Über uns. URL: <https://www.metalltechnischeindustrie.at/ueber-uns/> (Zugriff: 22.09.2019).

- Mosley, M.; Brackett, M.; et al. (2009): The DAMA Guide to The Data Management of Body of Knowledge (DAMA-DMBOK Guide). Technics Publications LLC. ISBN 978-0-9771400-8-4. S. 4-83.
- Otto, B. (2015): Quality and Value of the Data Resource in Large Enterprises. In: Information Systems Management, Jg. 32, Nr. 3, S. 234–251.
- Otto, B.; Hinderer, H. (2009): Datenqualitätsmanagement im Lieferanten-Controlling: Fallbeispiele, Architektorentwurf und Handlungsempfehlungen. In: Controlling & Management, Jg. 53, Nr. 1, S. 4.
- Otto, B.; Österle H. (2016): Corporate Data Quality. Springer Berlin Heidelberg New York. ISBN 978-3-662-46805-0, S. 29 f.
- Oussous, A.; Benjelloun, F.; Ait Lahcen, A.; Belfkih, S. (2017): Big Data technologies: A survey. In: Journal of King Saud University – Computer and Information Sciences, Jg. 2018, Nr. 30, S. 433.
- Paganoni, M. C. (2019): Framing Big Data: A Linguistic and Discursive Approach. Palgrave Pivot. ISBN 978-3030167875. S. 3.
- Poornima, S.; Pushpalatha, M. (2016): A Journey FROM BIG DATA TOWARDS PRESCRIPTIVE ANALYTICS. In: ARPN Journal of Engineering and Applied Sciences, Jg. 2016, Nr. 19, S. 472.
- Probst, G.; Raub, S.; Romhardt, K. (2010): Wissen managen. 6. Auflage. Gabler Verlag. ISBN: 978-3-8349-1903-8, S.28-40.
- PwC (Hrsg.) (2014): Chancen und Herausforderungen der vierten industriellen Revolution. München.
- Quality & Service Business GmbH (2019): Lexikon. URL: <https://www.quality.de/lexikon/qualitaetsregelkarte-qrk/> (Zugriff: 01.02.2019).
- Question Pro (2019): Rücklaufquote bei Online-Umfragen. URL: <https://www.questionpro.de/responsequote-ruecklaufquote-online-umfrage/> (Zugriff: 21.08.2019).
- RedHat (2019): Datei-, Block- oder Objekt- Storage. URL: <https://www.redhat.com/de/topics/data-storage/file-block-object-storage> (Zugriff: 20.10.2019).
- Rohweder, P.; Kasten, G.; Malzahn, D.; Piro, A.; Schmid, J. (2008): Daten- und Informationsqualität. Springer Verlag. ISBN: 978-3-8348-0321-4. S. 28.
- Romeike, F.; Hager, P. (2016): Predictive Analytics: Looking into the future. FIRM Jahrbuch, Bd. 2016. Frankfurter Institut für Risikomanagement und Regulierung, Frankfurt am Main. S. 168-171.
- Ryte (2016): Regressionsanalyse. URL: <https://de.ryte.com/wiki/Regressionsanalyse> (Zugriff: 31.01.2019).

- Sampaio, P.; Saraiva, P. (2016): Quality in the 21<sup>st</sup> Century, Springer Verlag. ISBN: 978-3-319-21331-6. S. 27.
- Sappelli, M.; Smit, S. K.; de Boer, Maaike H.T.; Bomhof, F. (2017): A Vision of Prescriptive Analytics. In: KESA (Hrsg.): ALLDATA 2017. Venedig. ISBN: 978-61208-552-4, S. 45-49.
- SAS, (2018): Predictive Analytics – Was es ist und was man darüber wissen sollte. URL: [https://www.sas.com/de\\_at/insights/analytics/predictive-analytics.html](https://www.sas.com/de_at/insights/analytics/predictive-analytics.html) (Zugriff: 04.02.2019).
- Schroeck, M.; Shockley, R.; Smart, J.; Romero-Morales, D.; Tufano, P. (2012): Analytics: Big Data in der Praxis. URL: <https://www-935.ibm.com/services/de/gbs/thoughtleadership/GBE03519-DEDE-00.pdf> (Zugriff: 12.01.2019).
- Schulze, P. M.; Porath, D. (2012): Statistik. 7. Auflage. Oldenbourg Wissenschaftsverlag GmbH. ISBN: 978-3-486-71847-8. S. 449-457.
- Schumacher M.; Schulgen-Kristiansen, G. (2008): Methodik klinischer Studien. Springer Verlag. ISBN: 978-3-540-85135-6.
- Spitta, M.; Bick, M. (2008): Informationswirtschaft. Eine Einführung. Zweite Ausgabe. Springer Verlag. ISBN 978-3-540-85115-8, S. 157-160.
- Survey Monkey (2019): Privacy Policy. URL: <https://de.surveymonkey.com/mp/legal/privacy-policy/> (Zugriff: 10.08.2019).
- Tankariya, V.; Parmar B. (2017): AWS Certified Developer – Associate Guide. Packt Publishing. ISBN 978-1787125629, S.232 ff.
- Toutenburg, H.; Heumann, C. (2008): Deskriptive Statistik. Sechste Ausgabe. Springer Verlag. ISBN: 978-3-540-77788-5, S. 40-351.
- Treyer, O. A. G. (2003): Business Statistik. Compendio Bildungsmedien AG. ISBN: 3-7155-9104-8. S. 113.
- Volker, P. A.; Hänisch, T. (2015): Internet der Dinge. Springer. ISBN: 978-3-658-06728-1. S. 9.
- Voß, S.; Gutenschwager, K. (2001): Informationsmanagement. Springer Verlag. ISBN: 978-3-540-67807-6, S.65-72.
- Wang, R. Y.; Strong, D. M. (1996): Beyond Accuracy: What Data Quality Means to Data Consumers. In: Journal of Information Systems Management, Jg. 12, Nr. 4, S. 5-34.
- Wittmann, W. (1959): Unternehmung und unvollkommene Information. Westdeutscher Verlag. ISBN: 978-3-322-98938-3, S. 14.
- WKO (2019): Branchenauswahl. URL: <https://www.wko.at/branchen/Branchenauswahl.html> (Zugriff: 08.03.2019).

WKO (2018): News. URL: <https://news.wko.at/news/oesterreich/wirtschaftskraft-kmu2018.pdf> (Zugriff: 30.08.2019).

WKO (2018): Papierindustrie Steiermark. URL: [https://www.wko.at/branchen/stmk/industrie/papierindustrie/infos\\_lehre\\_papierindustrie.html](https://www.wko.at/branchen/stmk/industrie/papierindustrie/infos_lehre_papierindustrie.html) (Zugriff: 22.09.2019).

Zennaro, M. (2017): Introduction to the Internet of Things. URL: [https://www.itu.int/en/ITU-D/Regional-Presence/AsiaPacific/SiteAssets/Pages/Events/2017/Nov\\_IOT/NBTC-ITU-IoT/Session%201%20IntroIoT-new%20template.pdf](https://www.itu.int/en/ITU-D/Regional-Presence/AsiaPacific/SiteAssets/Pages/Events/2017/Nov_IOT/NBTC-ITU-IoT/Session%201%20IntroIoT-new%20template.pdf). (Zugriff: 20.01.2019).

## Anhang A1 - Fragebogen



### Umfrage - Datenmanagement in Österreich Fragebogen an österreichische Industrieunternehmen

Aufgrund des weltweit stark gestiegenen Datenvolumens bekommen alle Bereiche und Tätigkeiten, die sich mit dem Thema Daten befassen einen immer höheren Stellenwert in Unternehmen. Dieser Fragebogen soll dazu dienen herauszufinden welchen Stand das Datenmanagement derzeit in österreichischen Industrieunternehmen einnimmt. Behandelte Themen hierbei sind: Zuständigkeiten, Umsetzung, Ziele, Hindernisse, Industrial Analytics und zukünftige Projekte.

Durch die Auswertung dieser Daten und den Vergleich mit anderen Studien (International und DACH-Region) soll der aktuelle Stand des Datenmanagements in Österreich dargestellt werden und Abweichungen zu anderen Ländern aufgezeigt werden. Daraus können Trends und Hindernisse für die nächsten Jahre ableiten werden.

Der nachfolgende Fragebogen besteht aus insgesamt 25 Fragen, die Beantwortung dauert rund 6 Minuten.

Ihre Teilnahme an der Umfrage ist freiwillig und Sie können jederzeit den Fragebogen ohne Angabe von Gründen abbrechen. Die Befragung ist anonym und es werden keine persönlichen Informationen erfragt, mit Ausnahme vom Unternehmensbereich in welchem Sie tätig sind.

Mit Ihrer Teilnahme tragen Sie zur Erforschung der oben beschriebenen Fragestellungen bei. Bei der Veröffentlichung in einer Diplomarbeit oder in einer wissenschaftlichen Fachzeitschrift wird aus den Daten nicht hervorgehen, wer an dieser Umfrage teilgenommen hat. Ihre persönlichen Daten unterliegen der Datenschutzgrundverordnung (DSGVO). Die Einwilligung zur Erhebung und Verarbeitung der Daten ist unwiderruflich, da aufgrund der anonymisierten Form der Umfrage keine teilnehmerbezogene Löschung durchgeführt werden kann. Weiters gilt auch die Datenschutzrichtlinie von Survey Monkey, welche unter <https://de.surveymonkey.com/mp/legal/privacy-policy/> eingesehen werden kann.

Bitte wenden Sie sich bei Rückfragen an Martin Wallner ([martin.wallner@stud.unileoben.ac.at](mailto:martin.wallner@stud.unileoben.ac.at)). Gerne können Sie nach der Auswertung, die Ergebnisse dieser Umfrage als PDF-Dokument bekommen. Dazu bitte ich Sie um eine kurze Nachricht an die angegebene E-Mail Adresse.

Indem Sie nun auf den untenstehenden Button „Weiter“ klicken, geben Sie Ihre Zustimmung zur Teilnahme an der Umfrage und erklären sich mit der vorstehend geschilderten Vorgehensweise einverstanden.



## Umfrage - Datenmanagement in Österreich

### Allgemeine Fragen

#### 1. In welchem Sektor ist ihr Unternehmen hauptsächlich tätig?

- |  |  |
|--|--|
| <input type="radio"/> Bauindustrie                         | <input type="radio"/> Holzindustrie  |
| <input type="radio"/> Bergwerke u. eisenerzeug. Industrie  | <input type="radio"/> Nahrungs- u. Genussmittelindustrie                           |
| <input type="radio"/> Chemische Industrie                  | <input type="radio"/> Industrielle Hersteller von Produkten aus Papier oder Karton |
| <input type="radio"/> Elektro- u. Elektronikindustrie      | <input type="radio"/> Textil-, Bekleidungs-, Schuh- und Lederindustrie             |
| <input type="radio"/> Fahrzeugindustrie                    | <input type="radio"/> Mineralölindustrie   |
| <input type="radio"/> Gas- und Wärmeversorgungsunternehmen | <input type="radio"/> NE-Metallindustrie   |
| <input type="radio"/> Glasindustrie                        | <input type="radio"/> Papierindustrie  |
| <input type="radio"/> Metalltechnische Industrie           | <input type="radio"/> Stein- und keramische Industrie                              |
| <input type="radio"/> Anderer (bitte angeben)              |  |

#### 2. Wie viele Mitarbeiter sind in Österreich an allen Standorten in ihrem Unternehmen tätig?

- |                                |                                  |
|--------------------------------|----------------------------------|
| <input type="radio"/> 1-50     | <input type="radio"/> 1001-5000  |
| <input type="radio"/> 51-200   | <input type="radio"/> 5001-10000 |
| <input type="radio"/> 201-1000 | <input type="radio"/> 10000+     |

3. Welcher Jahresumsatz wurde insgesamt an allen Standorten in Österreich erzielt [€]?

- |                                  |   |
|----------------------------------|---|
| <input type="radio"/> Bis 1 Mio. | <input type="radio"/> 50-100 Mio.       |
| <input type="radio"/> Bis 5 Mio. | <input type="radio"/> 100-500 Mio.      |
| <input type="radio"/> 5-10 Mio.  | <input type="radio"/> 0,5-1,5 Mrd.      |
| <input type="radio"/> 10-50 Mio. | <input type="radio"/> Mehr als 1,5 Mrd. |

4. Hat ihr Unternehmen im Vergleich mit dem direkten Wettbewerb letztes Jahr umsatzmäßig eher besser oder schlechter abgeschnitten?

- Eher besser
- Eher schlechter

5. In welchem Unternehmensbereich sind Sie tätig?

- |   |                                     |
|---|-------------------------------------|
| <input type="radio"/> IT                      | <input type="radio"/> F&E           |
| <input type="radio"/> Finanzen                | <input type="radio"/> Produktion    |
| <input type="radio"/> Controlling             | <input type="radio"/> Logistik      |
| <input type="radio"/> Geschäftsführung        | <input type="radio"/> Einkauf       |
| <input type="radio"/> Vertrieb                | <input type="radio"/> Personalwesen |
| <input type="radio"/> Marketing               |                                     |
| <input type="radio"/> Anderer (bitte angeben) |                                     |



## Umfrage - Datenmanagement in Österreich

### Fragen zum Datenmanagement

6. Wer ist für Daten innerhalb des Unternehmens verantwortlich?

- |   |   |
|---|---|
| <input type="radio"/> IT-Abteilung                    | <input type="radio"/> Eigens definierte/r<br>Datenverantwortliche/r |
| <input type="radio"/> CEO (Chief Executive Officer)   | <input type="radio"/> Jede Fachabteilung selbst                     |
| <input type="radio"/> CIO (Chief Information Officer) | <input type="radio"/> Die Verantwortung ist nicht festgelegt        |
| <input type="radio"/> CDO (Chief Data Officer)        |   |
| <input type="radio"/> Andere (bitte angeben)          |   |

Datenmanagement ist eine Unternehmensfunktion welche sich mit der Entwicklung, Implementierung und Wartung von Regeln und Abläufen beschäftigt, welche den Wert von Daten und Informationen kontrolliert, bzw. steigert und diese Daten schlussendlich zu den richtigen Personen liefert.

7. Ist ein unternehmensweites Datenmanagement implementiert?

- |   |   |
|---|---|
| <input type="radio"/> Standards werden derzeit umgesetzt<br>(Pilotphase)      | <input type="radio"/> Keine fixen Standards, aber diese sind in<br>Zukunft geplant        |
| <input type="radio"/> Bereits fix im Unternehmensprozess<br>integriert        | <input type="radio"/> Keine fixen Standards existent und in<br>Zukunft auch nicht geplant |
| <input type="radio"/> Datenmanagement in einzelnen<br>Fachbereichen vorhanden |   |

8. Bezugnehmend auf die letzte Frage, wie ist Ihrer Meinung nach ihr Unternehmen im direkten Vergleich mit dem Wettbewerb im Bereich des Datenmanagement aufgestellt?

- Besser
- Gleich
- Schlechter

9. Bei welchen Faktoren hat das implementierte Datenmanagement den höchsten Mehrwert in ihrem Unternehmen gebracht? (Mehrfachantworten möglich)

- Datenintegration
- Echtzeitverarbeitung
- Analysemethoden
- Bereitstellung von Daten
- Datenqualität
- Abfragen
- Reaktionszeit auf Änderungen oder neue Anforderungen

10. Welche Ursachen konnten Sie im Zusammenhang mit Problemen in Bezug auf das Datenmanagement festmachen? (Mehrfachantworten möglich)

- Menschliches Versagen
- Kommunikationsschwierigkeiten unterschiedlicher Abteilungen
- Fehler im definierten Datenmanagement
- Fehlende Unterstützung des Managements
- Sonstiges (bitte angeben)
- Fehlende Ressourcen
- Unzureichendes Budget
- Technischer Stand unzureichend
- Unzureichende Schulung für den derzeitigen technischen Stand



## Umfrage - Datenmanagement in Österreich

### Fragen zum Datenmanagement

11. Haben Sie das Gefühl, dass in ihrem Unternehmen **operative** Entscheidungen auf Grundlage von Daten erfolgen?

- JA
- NEIN

12. Haben Sie das Gefühl, dass in ihrem Unternehmen **strategische** Entscheidung auf Grundlage von Daten erfolgen?

- JA
- NEIN

13. Sind die Auswerteperspektiven und Key Performance Indicators (KPIs) standardisiert, sodass ein unternehmensweiter Vergleich von Daten möglich ist?

- JA
- NEIN

**14. Auf welchen Themen liegt derzeit der Schwerpunkt in ihrem Unternehmen? (Mehrfachantworten möglich)**

- |   |  |
|---|--|
| <input type="checkbox"/> Integration von heterogenen Datenquellen                 | <input type="checkbox"/> Big-Data Technologien   |
| <input type="checkbox"/> Fachbereiche sollen Daten selbständig integrieren können | <input type="checkbox"/> Neue Datenarchitektur   |
| <input type="checkbox"/> Komplexität der Datenarchitektur minimieren              | <input type="checkbox"/> Implementierung eines Data Warehouses                                 |
| <input type="checkbox"/> Datenstrategie   | <input type="checkbox"/> Implementierung eines Data Lakes                                      |
| <input type="checkbox"/> Einfache Integration von neuen Datenquellen              | <input type="checkbox"/> Implementierung eines Cloud-basierten Data Warehouses oder Data Lakes |
| <input type="checkbox"/> Sonstige (bitte angeben)                                 |  |

**15. Bei welchen der folgenden Faktoren sehen Sie in ihrem Unternehmen noch den größten Aufholbedarf? (Mehrfachantworten möglich)**

- |  |   |
|--|---|
| <input type="checkbox"/> Datenintegration        | <input type="checkbox"/> Analysemethoden                                      |
| <input type="checkbox"/> Datenmanagement         | <input type="checkbox"/> Bereitstellung von Daten                             |
| <input type="checkbox"/> Datenmodellierung       | <input type="checkbox"/> Datenqualität  |
| <input type="checkbox"/> Datenbankadministration | <input type="checkbox"/> Abfragemöglichkeiten                                 |
| <input type="checkbox"/> Echtzeitverarbeitung    | <input type="checkbox"/> Reaktionszeit auf Änderungen oder neue Anforderungen |



## Umfrage - Datenmanagement in Österreich

### Fragen zum Datenmanagement

16. Bei welcher der folgenden Dimensionen zur Festlegung der Qualität von Daten hat ihr Unternehmen aus Ihrer Sicht noch den größten Aufholbedarf? (Mehrfachantworten möglich)

- Intrinsische Datenqualität: Genauigkeit, Objektivität und Fehlerfreiheit der Daten.
- Kontextuelle Datenqualität: Sind die Daten soweit vollständig, dass diese sinnvoll genutzt werden können.
- Darstellungsbezogene Datenqualität: Sind die Daten übersichtlich, also in weiterer Folge auch leicht verständlich.
- Zugangsbezogene Datenqualität: Sind die Daten leicht auffindbar, kann auf diese leicht zugegriffen werden und können diese einfach bearbeitet werden.

17. Wie wird die Speicherung von Daten in ihrem Unternehmen organisiert?

- Data Warehouse (DWH)
- DWH mit Data Marts
- Data Lakes
- Relationale Datenbank mit Datenbankmanagementsystem (z.B. SQL Server)
- Nicht-relationale Datenbanken (z.B. NoSQL)
- Cloudspeicher
- Sonstige (bitte angeben)

18. Werden bereits Big Data Projekte in ihrem Unternehmen durchgeführt, welche ein komplexeres Datenanalyseverfahren erlauben?

- JA
- NEIN

19. In welchem der folgenden Abteilungen sehen Sie das meiste Potential für Big Data Projekte? (Mehrfachantworten möglich)

- Vertrieb
- Finanzen/Controlling
- Marketing
- IT
- Produktion
- Logistik
- Kundenservice
- F&E
- Risikomanagement
- Personalwesen

20. Was sind in ihrem Unternehmen die wichtigsten Faktoren, welche sich mit Data Analytics Ihrer Meinung nach verbessern lassen? (Mehrfachantworten möglich)

- |   |   |
|---|---|
| <input type="checkbox"/> Umsatzsteigerung                           | <input type="checkbox"/> Besseres Verständnis der Kundenwünsche |
| <input type="checkbox"/> Kostensenkung                              | <input type="checkbox"/> Verbesserung der Lieferkette           |
| <input type="checkbox"/> Bessere Steuerung von operativen Prozessen | <input type="checkbox"/> Verbesserung der Produktionsplanung    |
| <input type="checkbox"/> Steigerung der Kundenzufriedenheit         | <input type="checkbox"/> Einhaltung von Gesetzen und Regularien |
| <input type="checkbox"/> Steigerung der Produktqualität             | <input type="checkbox"/> Automatisierung von Entscheidungen     |
| <input type="checkbox"/> Verbesserung der Logistik                  | <input type="checkbox"/> Beschleunigung von Entscheidungen      |
| <input type="checkbox"/> Weniger Stehzeiten der Maschinen           |   |
| <input type="checkbox"/> Andere (bitte angeben)                     |   |



## Umfrage - Datenmanagement in Österreich

### Fragen zum Datenmanagement

21. Was sehen Sie als die größten Herausforderungen auf der **geschäftlichen** Seite für die Durchführung von Big Data Projekten? (Mehrfachantworten möglich)

- |  |   |
|--|---|
| <input type="checkbox"/> Probleme bei der Zuordnung von Aufgaben         | <input type="checkbox"/> Projektkomplexität                     |
| <input type="checkbox"/> Probleme beim Finden des Business Cases         | <input type="checkbox"/> Zusammenarbeit mit Konkurrenten        |
| <input type="checkbox"/> Zusammenarbeit mit externen Partnern            | <input type="checkbox"/> Angst vor „falscher“ Nutzung von Daten |
| <input type="checkbox"/> Fehlende Schulungen und/ oder Kenntnisse der MA |   |

22. Was sehen Sie als die größten Herausforderungen auf **technischer** Seite für die Durchführung von Big Data Projekten? (Mehrfachantworten möglich)

- Datenqualität
- Zusammenarbeit mehrerer Systemkomponenten
- Interpretation der Daten
- Datensicherheit
- Datengewinnung
- Technische Infrastruktur
- Einbindung in die Firmeninfrastruktur
- Fehlendes Know-how
- Streaming Analytics
- Automatisierung von Entscheidungen

23. Welche Analysemethoden werden hauptsächlich in ihrem Unternehmen eingesetzt? (Mehrfachantworten möglich)

- Deskriptive Analyse (Was ist passiert?)
- Diagnostische Analyse (Warum ist es passiert?)
- Prädiktive Analyse (Was wird passieren?)
- Präskriptive Analyse (Wie können Prozesse gesteuert und Entscheidung so unterstützt werden, dass ein bestimmter Zustand erreicht wird?)

24. Welche Analysemethoden glauben Sie werden in 5 Jahren hauptsächlich angewendet werden? (Mehrfachantworten möglich)

- Deskriptive Analyse
- Diagnostische Analyse
- Prädiktive Analyse
- Präskriptive Analyse

25. Welche Datenquellen wurden bei der Umsetzung von Data Analytics Projekten zumeist verwendet? (Mehrfachantworten möglich)

- Daten aus Transaktionssystemen (Datenbanken)
- Sensor-Daten (z.B. RFID- Radio-frequency identification)
- Maschinendatenerfassung (z.B. Prozessdaten oder Prozessmeldungen)
- Betriebsdatenerfassung (z.B. Auftragsstatus oder Produktionsqualität)
- Social-Media-Daten
- Web-Analysedaten
- Text
- Video-, Bilddaten
- Sonstige (bitte angeben)

Herzlichen Dank für Ihre Teilnahme an der Umfrage!

## Anhang A2 - Teilnehmerliste Umfrageerhebung

respondent	collector_id	date_created	date_modified	ip_address
1,0842E+10	234786001	2019-07-05 12:33:39	2019-07-05 12:34:20	80.121.206.26
1,084E+10	234786001	2019-07-04 9:01:23	2019-07-04 9:06:37	213.225.2.176
1,0838E+10	234786001	2019-07-03 16:34:43	2019-07-03 17:12:32	62.218.50.226
1,0838E+10	234786001	2019-07-03 12:21:13	2019-07-03 15:01:29	184.25.254.101
1,0837E+10	234786001	2019-07-03 9:08:48	2019-07-03 9:10:19	91.112.199.186
1,0836E+10	234786001	2019-07-02 17:25:36	2019-07-02 17:26:58	213.150.228.38
1,0836E+10	234786001	2019-07-02 16:56:49	2019-07-02 17:02:17	91.114.46.180
1,0835E+10	234786001	2019-07-02 6:16:52	2019-07-02 6:25:36	213.208.156.24
1,0833E+10	234786001	2019-07-01 16:55:18	2019-07-01 17:24:54	194.107.82.176
1,0833E+10	234786001	2019-07-01 15:58:35	2019-07-01 16:08:55	165.225.72.207
1,0833E+10	234786001	2019-07-01 14:28:55	2019-07-01 14:37:27	185.81.215.228
1,0833E+10	234786001	2019-07-01 14:28:52	2019-07-01 14:37:58	165.225.73.86
1,0833E+10	234786001	2019-07-01 13:17:52	2019-07-01 13:28:18	185.28.209.3
1,0833E+10	234786001	2019-07-01 10:09:00	2019-07-01 10:12:16	62.212.184.2
1,0833E+10	234786001	2019-07-01 9:44:03	2019-07-01 9:55:10	193.228.211.88
1,0832E+10	234786001	2019-06-30 13:50:24	2019-06-30 13:58:55	88.106.97.160
1,0829E+10	234786001	2019-06-28 16:17:07	2019-06-28 16:22:33	80.120.0.130
1,0829E+10	234786001	2019-06-28 15:27:39	2019-06-28 15:36:06	193.187.250.2
1,0829E+10	234786001	2019-06-28 12:52:21	2019-06-28 13:03:14	83.164.140.2
1,0829E+10	234786001	2019-06-28 11:57:09	2019-06-28 12:34:40	212.183.8.132
1,0829E+10	234786001	2019-06-28 10:02:31	2019-06-28 10:23:21	213.33.101.75
1,0828E+10	234786001	2019-06-28 10:00:03	2019-06-28 10:05:17	82.218.178.18
1,0828E+10	234786001	2019-06-28 9:51:38	2019-06-28 9:56:44	80.110.89.157
1,0828E+10	234786001	2019-06-28 8:08:35	2019-06-28 8:41:14	195.216.229.247
1,0828E+10	234786001	2019-06-28 8:28:40	2019-06-28 8:33:22	193.186.185.102
1,0827E+10	234786001	2019-06-27 14:50:39	2019-06-27 15:05:54	217.75.176.70
1,0826E+10	234786001	2019-06-27 14:36:12	2019-06-27 14:37:35	83.164.140.2
1,0826E+10	234786001	2019-06-27 10:20:36	2019-06-27 10:46:54	165.225.72.207
1,0825E+10	234786001	2019-06-27 0:05:25	2019-06-27 0:41:00	81.217.38.253
1,0825E+10	234786001	2019-06-26 18:38:26	2019-06-26 18:50:46	213.225.36.132
1,0824E+10	234786001	2019-06-26 17:29:29	2019-06-26 17:36:35	82.210.248.72
1,0824E+10	234786001	2019-06-26 17:20:24	2019-06-26 17:26:33	82.210.248.72
1,0824E+10	234786001	2019-06-26 14:52:12	2019-06-26 15:05:37	193.186.185.102
1,0824E+10	234786001	2019-06-26 13:43:04	2019-06-26 13:47:16	80.120.11.165
1,0824E+10	234786001	2019-06-26 13:03:25	2019-06-26 13:13:36	80.120.124.7
1,0824E+10	234786001	2019-06-26 11:36:31	2019-06-26 11:47:10	194.48.190.130
1,0823E+10	234786001	2019-06-26 10:03:43	2019-06-26 10:04:29	80.121.155.205
1,0823E+10	234786001	2019-06-26 8:22:52	2019-06-26 9:06:23	92.60.14.211
1,0823E+10	234786001	2019-06-26 8:20:21	2019-06-26 8:29:27	37.186.149.60
1,0823E+10	234786001	2019-06-26 7:36:16	2019-06-26 7:40:32	178.189.115.35
1,0822E+10	234786001	2019-06-25 22:04:29	2019-06-25 22:05:44	93.82.79.144
1,0822E+10	234786001	2019-06-25 17:01:15	2019-06-25 10:00:36	62.218.23.122
1,0821E+10	234786001	2019-06-25 13:56:06	2019-06-25 14:03:13	108.171.129.170
1,0821E+10	234786001	2019-06-25 13:44:19	2019-06-25 13:51:15	213.33.114.196
1,0821E+10	234786001	2019-06-25 13:15:08	2019-06-25 13:19:31	80.121.212.246
1,0821E+10	234786001	2019-06-25 13:08:51	2019-06-25 13:12:34	213.33.114.196
1,0821E+10	234786001	2019-06-25 12:53:36	2019-06-25 13:02:01	165.225.84.145

respondent	collector_id	date_created	date_modified	ip_address
1,0821E+10	234786001	2019-06-25 12:05:06	2019-06-25 12:14:13	46.183.103.8
1,0821E+10	234786001	2019-06-25 11:20:23	2019-06-25 11:41:46	193.170.242.24
1,0821E+10	234786001	2019-06-25 11:10:39	2019-06-25 11:23:26	143.161.248.25
1,0821E+10	234786001	2019-06-25 10:23:54	2019-06-25 10:31:42	80.120.0.46
1,0821E+10	234786001	2019-06-25 9:16:54	2019-06-25 9:33:01	195.230.170.230
1,0821E+10	234786001	2019-06-25 8:49:33	2019-06-25 8:59:02	165.225.73.16
1,0821E+10	234786001	2019-06-25 8:13:29	2019-06-25 8:20:12	87.243.151.147
1,0821E+10	234786001	2019-06-25 7:10:03	2019-06-25 7:16:45	212.67.242.44
1,082E+10	234786001	2019-06-25 5:58:09	2019-06-25 5:59:34	212.186.29.177
1,0819E+10	234786001	2019-06-24 20:09:49	2019-06-24 20:12:47	185.46.214.64
1,0819E+10	234786001	2019-06-24 18:38:50	2019-06-24 18:54:26	193.186.105.11
1,0819E+10	234786001	2019-06-24 17:42:04	2019-06-24 17:57:30	185.38.169.226
1,0819E+10	234786001	2019-06-24 17:12:12	2019-06-24 17:18:51	82.218.226.39
1,0819E+10	234786001	2019-06-24 17:01:30	2019-06-24 17:13:11	185.46.212.84
1,0819E+10	234786001	2019-06-24 16:28:55	2019-06-24 16:29:31	81.189.155.51
1,0819E+10	234786001	2019-06-24 16:06:34	2019-06-24 16:10:14	217.75.176.70
1,0819E+10	234786001	2019-06-24 15:55:59	2019-06-24 16:03:35	195.202.142.218
1,0819E+10	234786001	2019-06-24 15:42:59	2019-06-24 15:50:49	93.83.113.228
1,0819E+10	234786001	2019-06-24 15:20:08	2019-06-24 15:32:07	91.112.189.52
1,0819E+10	234786001	2019-06-24 8:37:37	2019-06-26 10:47:12	80.123.233.199
1,0819E+10	234786001	2019-06-24 15:16:16	2019-06-24 15:19:49	194.39.218.10
1,0818E+10	234786001	2019-06-24 14:24:22	2019-06-24 15:42:01	194.107.94.54
1,0818E+10	234786001	2019-06-24 14:22:31	2019-06-24 14:37:07	109.68.107.2
1,0818E+10	234786001	2019-06-24 14:20:58	2019-06-24 14:29:13	195.69.193.51
1,0818E+10	234786001	2019-06-24 14:18:40	2019-06-24 14:27:25	194.107.94.54
1,0818E+10	234786001	2019-06-24 14:00:27	2019-06-24 14:09:13	80.120.254.76
1,0818E+10	234786001	2019-06-24 13:46:46	2019-06-24 13:54:44	213.33.67.59
1,0818E+10	234786001	2019-06-24 13:44:54	2019-06-24 13:53:18	88.128.81.61
1,0818E+10	234786001	2019-06-24 13:29:34	2019-06-24 13:54:03	81.217.155.103
1,0818E+10	234786001	2019-06-24 12:56:40	2019-06-24 13:06:27	160.46.252.19
1,0818E+10	234786001	2019-06-24 12:55:53	2019-06-24 12:57:46	85.126.93.34
1,0818E+10	234786001	2019-06-24 12:50:17	2019-06-24 12:57:46	82.218.173.52
1,0818E+10	234786001	2019-06-24 12:45:42	2019-06-24 12:55:27	157.167.59.180
1,0818E+10	234786001	2019-06-24 12:47:14	2019-06-24 12:55:30	85.233.98.34
1,0818E+10	234786001	2019-06-24 12:31:16	2019-06-24 12:40:38	108.171.129.167
1,0818E+10	234786001	2019-06-24 12:12:40	2019-06-24 12:19:17	157.167.59.180
1,0818E+10	234786001	2019-06-24 12:00:59	2019-06-24 12:08:52	31.193.162.66
1,0818E+10	234786001	2019-06-24 11:56:11	2019-06-24 12:12:47	217.115.75.200
1,0818E+10	234786001	2019-06-24 11:48:02	2019-06-24 12:02:20	93.82.249.20
1,0818E+10	234786001	2019-06-24 11:33:52	2019-06-24 12:26:07	213.153.63.219
1,0818E+10	234786001	2019-06-24 11:24:35	2019-06-24 11:31:36	165.225.72.129
1,0818E+10	234786001	2019-06-24 11:22:25	2019-06-24 11:30:32	80.120.85.90
1,0818E+10	234786001	2019-06-24 11:22:04	2019-06-24 11:32:31	80.120.141.215
1,0818E+10	234786001	2019-06-24 11:04:17	2019-06-24 11:09:54	80.121.146.50
1,0818E+10	234786001	2019-06-24 10:40:14	2019-06-24 10:41:26	185.46.214.101
1,0818E+10	234786001	2019-06-24 10:07:11	2019-06-24 10:19:15	213.33.89.18
1,0818E+10	234786001	2019-06-24 9:59:57	2019-06-24 10:10:54	195.69.193.54

respondent	collector_id	date_created	date_modified	ip_address	
1,0818E+10	234786001	2019-06-24 10:00:01	2019-06-24 10:08:15	81.189.253.161	
1,0818E+10	234786001	2019-06-24 9:57:54	2019-06-24 10:10:31	83.164.128.182	
1,0818E+10	234786001	2019-06-24 9:54:05	2019-06-24 9:57:51	82.218.177.149	
1,0818E+10	234786001	2019-06-24 9:51:34	2019-06-24 10:03:15	81.217.114.172	
1,0818E+10	234786001	2019-06-24 9:44:35	2019-06-24 9:53:44	80.120.128.251	
1,0818E+10	234786001	2019-06-24 9:43:10	2019-06-24 9:55:45	195.202.142.218	
1,0818E+10	234786001	2019-06-24 9:37:10	2019-06-24 9:43:00	83.164.137.230	
1,0818E+10	234786001	2019-06-24 9:23:27	2019-06-24 9:29:41	89.16.200.250	
1,0818E+10	234786001	2019-06-24 9:23:17	2019-06-24 9:25:41	195.202.151.40	
1,0818E+10	234786001	2019-06-24 9:21:22	2019-06-24 9:22:36	185.46.213.78	
1,0818E+10	234786001	2019-06-24 9:20:11	2019-06-24 9:28:41	165.225.72.165	
1,0818E+10	234786001	2019-06-24 9:16:08	2019-06-24 9:17:29	31.193.164.146	
1,0818E+10	234786001	2019-06-24 9:00:35	2019-06-24 9:08:19	178.189.78.226	
1,0818E+10	234786001	2019-06-24 8:58:15	2019-06-24 9:05:44	213.174.241.8	
1,0818E+10	234786001	2019-06-24 8:55:31	2019-06-24 9:01:44	81.189.118.2	
1,0818E+10	234786001	2019-06-24 8:46:52	2019-06-24 9:02:30	193.228.211.89	
1,0818E+10	234786001	2019-06-24 8:44:45	2019-06-24 8:49:26	81.189.155.51	
1,0818E+10	234786001	2019-06-24 8:29:51	2019-06-24 8:44:14	193.178.171.60	
1,0818E+10	234786001	2019-06-24 8:23:26	2019-06-24 8:24:53	91.114.29.58	
1,0818E+10	234786001	2019-06-24 8:21:11	2019-06-24 8:27:48	165.225.72.159	
1,0818E+10	234786001	2019-06-24 8:19:23	2019-06-24 8:27:09	62.218.119.3	
1,0818E+10	234786001	2019-06-24 8:16:10	2019-06-24 8:41:48	80.120.164.194	
1,0818E+10	234786001	2019-06-24 8:01:11	2019-06-24 8:09:37	81.223.169.210	
1,0818E+10	234786001	2019-06-24 7:56:17	2019-06-24 7:58:46	188.21.3.133	
1,0818E+10	234786001	2019-06-24 7:55:42	2019-06-24 8:06:45	62.218.145.196	
1,0818E+10	234786001	2019-06-24 7:53:55	2019-06-24 8:03:32	212.52.198.74	
1,0818E+10	234786001	2019-06-24 7:52:38	2019-06-24 7:58:12	170.103.97.53	
1,0818E+10	234786001	2019-06-24 7:47:39	2019-06-24 7:57:48	80.120.160.1	
1,0818E+10	234786001	2019-06-24 7:43:07	2019-06-24 7:50:18	85.158.138.19	
1,0818E+10	234786001	2019-06-24 7:40:38	2019-06-25 13:24:52	193.228.112.247	
1,0818E+10	234786001	2019-06-24 7:39:14	2019-06-24 7:45:50	193.228.0.14	
1,0818E+10	234786001	2019-06-24 7:36:02	2019-06-24 7:47:46	80.121.201.146	
1,0818E+10	234786001	2019-06-24 7:36:05	2019-06-24 7:39:34	194.107.82.176	
1,0818E+10	234786001	2019-06-24 7:34:25	2019-06-24 7:36:38	80.121.204.253	
1,0818E+10	234786001	2019-06-24 7:31:00	2019-06-24 7:38:19	85.158.138.19	
1,0818E+10	234786001	2019-06-24 7:30:00	2019-06-24 7:31:49	80.120.11.84	
1,0818E+10	234786001	2019-06-24 7:09:30	2019-06-24 7:32:02	193.239.226.41	
1,0818E+10	234786001	2019-06-24 7:11:49	2019-06-24 7:22:16	85.158.138.19	
1,0818E+10	234786001	2019-06-24 7:07:47	2019-06-24 7:16:20	80.120.127.20	
1,0818E+10	234786001	2019-06-24 6:56:41	2019-06-24 6:57:51	212.108.4.219	
1,0818E+10	234786001	2019-06-24 6:54:04	2019-06-24 7:01:23	185.46.214.79	
1,0818E+10	234786001	2019-06-24 6:43:17	2019-06-24 6:48:46	195.70.107.97	
1,0817E+10	234786001	2019-06-23 22:53:41	2019-06-23 23:04:24	178.115.130.65	
1,0817E+10	234786001	2019-06-23 22:42:57	2019-06-23 22:48:21	188.22.158.207	
1,0817E+10	234786001	2019-06-23 21:18:12	2019-06-23 21:26:21	193.81.40.9	
1,0817E+10	234786001	2019-06-23 21:15:57	2019-06-23 21:25:45	192.164.130.141	
1,0817E+10	234786001	2019-06-23 18:13:11	2019-06-23 18:19:22	185.46.214.84	

---

respondent	collector_id	date_created	date_modified	ip_address
1,0816E+10	234786001	2019-06-22 12:44:59	2019-06-22 12:46:17	165.225.72.214
1,0816E+10	234786001	2019-06-22 10:37:28	2019-06-22 10:46:18	194.107.94.54
1,0815E+10	234786001	2019-06-22 9:17:46	2019-06-22 9:34:35	84.112.176.58
1,0814E+10	234786001	2019-06-21 20:58:39	2019-06-21 21:04:50	193.81.33.6
1,0814E+10	234786001	2019-06-21 20:08:14	2019-06-21 20:13:34	89.144.219.158
1,0814E+10	234786001	2019-06-21 18:25:55	2019-06-21 18:31:04	194.118.198.150
1,0814E+10	234786001	2019-06-21 17:52:18	2019-06-21 18:04:32	91.133.72.45