

**Hochschule Neu-Ulm**

University of Applied Sciences

Fachbereich IMA

**Bachelorarbeit in**

**Information Management Automotive**

Zur Erlangung des akademischen Grades Bachelor of Science

**Welche Datenqualität wird hinsichtlich des Aufwandes und  
Nutzens im Rahmen eines zur Unternehmenssteuerung  
taugliches BI-System benötigt?**

**Name: Marcel Pfaff**

**Matrikelnummer: 261241**

Erstprüfer: Prof. Dipl. Ing. Manfred Plechaty

Zweitprüfer: Christoph Martini

Abgabedatum: Neu-Ulm, 19.09.2022

## Eidesstattliche Erklärung

Ich versichere, dass ich diese Bachelorarbeit selbständig verfasst und nur die angegebenen Quellen und Hilfsmittel verwendet habe. Ich stimme zu, dass meine Arbeit im Verdachtsfall mit einer Plagiatssoftware überprüft werden darf.

Ochsenhausen, den: 19.09.2022

A handwritten signature in black ink, consisting of stylized, overlapping letters and a long horizontal stroke extending to the right.

## Hinweise zur Formulierung

Der besseren Lesbarkeit halber wurde in dem vorliegenden Bericht auf die Verwendung der Sprachformen männlich, weiblich, diverse verzichtet und die männliche Form verwendet. Sämtliche Personenbezeichnungen gelten gleichermaßen für alle Geschlechter.

## Zusammenfassung

Unternehmen verfolgen alle das Ziel, aus dem BI-System den größtmöglichen Nutzen zu ziehen, um somit das Unternehmen in verschiedene Richtungen lenken zu können. Was jedoch, wenn diese Daten fehlerhaft sind? Daher ist es notwendig zu verstehen, welche Fehlerquellen vor oder im BI-System entstehen können, um dieses beheben zu können.

Das Ziel der vorliegenden Arbeit ist es zu beantworten, welche Datenqualität wird hinsichtlich des Aufwandes und Nutzens, im Rahmen eines zur Unternehmenssteuerung taugliches BI-System benötigt.

Um die Forschungsfrage zu beantworten, wurden Experten befragt, welche aktuelle Probleme beschrieben. Es wurden im Rahmen dieser Arbeit die theoretischen Grundlagen der Datenqualität mit ihrem Kosten und Nutzens hinsichtlich eines BI-Systems, auch anhand statistischer Werte, erläutert. Weiter wird ein Einblick in die Quellsysteme eines BI-Systems gegeben und wie sich dort die Datenqualität verbessern lässt.

Aufbauend auf den Experteninterviews in Verbindung mit den theoretischen Grundlagen, entsteht als Ergebnis dieser Arbeit ein praxisnaher Leitfaden, der Unternehmen bei der Verbesserung der Datenqualität eine Unterstützung darbieten kann.

## **Abstract**

Companies all have the goal of getting the most out of the BI system so that they can steer the company in different directions. But what happens if this data is faulty? Therefore, it is necessary to understand which sources of error can arise before or in the BI system in order to be able to correct them.

The aim of this thesis is to answer the question of what data quality is required in terms of effort and benefit within the framework of a BI system that is suitable for corporate management.

To answer the research question, experts were interviewed who described current problems. The theoretical foundations of data quality with its costs and benefits with regard to a BI system, also on the basis of statistical values, were explained in the context of this work. Furthermore, an insight is given into the source systems of a BI system and how data quality can be improved there.

Based on the expert interviews in conjunction with the theoretical principles, the result of this work is a practical guideline that can support companies in improving data quality.

## Abbildungsverzeichnis

Abbildung 1 – (Pütter, 2007).....	11
Abbildung 2 – eigene Abbildung.....	13
Abbildung 3 - (statista, 2022).....	16
Abbildung 4 – eigene Abbildung, vgl. (Knut Hildebrand, 2022).....	22
Abbildung 5 – eigene Abbildung, vgl. (Enderlein, 2015).....	26
Abbildung 6 – eigene Abbildung, vgl.(Eppler M, 2004) .....	30
Abbildung 7 – eigene Abbildung, vgl.(Klein, 2009) .....	33
Abbildung 8 - (Enderlein, 2015).....	40
Abbildung 9 – eigene Abbildung, vgl. (Enderlein, 2015).....	42
Abbildung 10 - (cmes, 2022).....	43
Abbildung 11 - (collaboration, 2022b).....	46
Abbildung 12 – eigene Abbildung – vgl.(Detlef Apel, 2015) .....	50
Abbildung 13 - (kupil, 2022).....	55,56

## Abkürzungsverzeichnis

DQ	Datenqualität
Bzw.	Beziehungsweise
Bzgl	Bezüglich
DQM	Datenqualitätsmanagement
BI	Business Intelligence
BI-Systeme	Business Intelligence Systeme
Vgl.	Vergleiche
DWH	Data Warehouse
IoT	Internet of Things
OLAP	Online analytical processing
RoI	Return on Investment
SPS	Speicherprogrammierbare Steuerung
ICS	Industrielle Kontrollsysteme
DBMS	Datenbankmanagementsystem
ETL	Extract Transform Load

# Inhaltsverzeichnis

<b>Einleitung</b>	<b>11</b>
<b>Aufbau der Daten-Hierarchie bzw. Informationsfluss</b>	<b>13</b>
Low Level	14
Middle Level	14
High Level	14
Level of production management	15
Level of ressource management	15
Level of strategic management	15
<b>Warum ist Datenqualität ein Faktor für ein funktionierendes BI-System</b>	<b>16</b>
<b>Definition Datenqualität</b>	<b>17</b>
<b>Datenqualität im BI</b>	<b>18</b>
<b>Warum brauchen wir gute Datenqualität im BI</b>	<b>19</b>
<b>Messbarkeit von Datenqualität</b>	<b>20</b>
Datenqualitätskriterien	21
Messung der vorhandenen Datenqualität	22
Vorteile der Kennzahlen	23
Nachteile der Kennzahlen	23
Erklärung der Messpunkte	24
<b>Welche Datenqualität wird benötigt</b>	<b>26</b>
Prozesse	27
Datenarchitektur / Schnittstellen	27
Definition	28
Datenverwendung	28
Datenverfall	28
Zusammenfassung	28

<b>Symptome schlechter Datenqualität</b>	<b>29</b>
<b>Kosten schlechter Datenqualität</b>	<b>29</b>
Inwieweit kosten Fehler Geld	30
Direkte Kosten	31
Indirekte Kosten	31
Präventivkosten	32
Entdeckungskosten	32
Bereinigungskosten	32
<b>Business Case Betrachtung</b>	<b>32</b>
<b>Business Case</b>	<b>33</b>
Kostenunterscheidungen im Business Case	34
Initiale Kosten	34
Laufende Kosten	34
<b>Big Data als neue Evolutionsstufe für BI</b>	<b>35</b>
Definition der Begrifflichkeit	35
Definition der Begrifflichkeit	36
Data Science	38
Data Mining	38
Bedeutung der Datenqualität im BI in Bezug auf Big Data	39
Beispiel Predictive Maintenance oder Predictive Analytics	39
<b>Business Intelligence Systeme</b>	<b>40</b>
Experteninterview über die Grenzen der Quellsysteme und Stärken eines Professionellen BI	41
Beispiel des Experten	41
<b>Data Warehouse</b>	<b>42</b>
<b>Datenqualitätsverbesserung als Prozess</b>	<b>42</b>
<b>Verbesserung der DQ in Quellsystemen</b>	<b>43</b>
Beispiele Möglicher Quellsysteme	43
ERP	44

C-MES _____	44
CRM _____	45
PDM _____	45
SCADA _____	46
Übersicht der Systeme auf „niedrigerer“ Ebene _____	46
<b><i>Datenqualitätsverbesserung in QS</i></b> _____	<b>47</b>
Identifizieren/Priorisieren _____	48
Zielwerte definieren _____	48
Gründe untersuchen _____	48
Prozesse optimieren _____	49
Zielwerte gegenüberstellen _____	49
<b><i>Data Profiling</i></b> _____	<b>50</b>
Beispiel für Data Profiling _____	51
<b><i>Durchführung eines DQ-Monitorings</i></b> _____	<b>51</b>
Initiales Data Profiling _____	52
Laufendes Data Profiling _____	52
<b><i>Experteninterview</i></b> _____	<b>53</b>
Vorgehensweise Organisation des Experteninterviews _____	53
Vorbereitung und Vorgehensweise für das Experteninterview _____	54
Hintergrund des Experten _____	54
Erfahrung des Experten _____	55
Resümee des Experteninterviews _____	57
<b><i>Entwicklung eines Leitfadens für Unternehmen</i></b> _____	<b>59</b>
Entwicklung des Leitfadens _____	59
Grundlegendes _____	59
Ziel _____	59
Schritt 1: Identifizierung _____	60
Schritt 2: Zielwerte definieren _____	60

<b>Schritt 3: Ursachen untersuchen</b>	<b>60</b>
<b>Schritt 4: Prozesse optimieren</b>	<b>60</b>
<b>Schritt 5: Zielwerte gegenüberstellen</b>	<b>61</b>
<b>Motivation für gute Datenqualität</b>	<b>61</b>
<b>Praxisnahes Beispiel für eine Datenqualitätsverbesserung</b>	<b>61</b>
<b><i>Limitierungen und zukünftiger Forschungsbedarf</i></b>	<b>62</b>
<b><i>Kritische Betrachtung der Bachelorarbeit</i></b>	<b>63</b>
<b><i>Gesamtfazit der Arbeit</i></b>	<b>64</b>
<b><i>Anhang</i></b>	<b>66</b>
<b>Transkription</b>	<b>66</b>
<b><i>Literaturverzeichnis</i></b>	<b>71</b>

## Einleitung

**Die besten und teuersten Business Intelligence-Initiativen nützen nichts, wenn die Qualität der zugrunde liegenden Daten nicht stimmt. Und die ist so schlecht, dass den Banken bspw. Kosten in Höhe von 40 bis 50 Prozent ihres IT-Budgets entstehen. Das behauptet der Berater Agens in einer Studie.<sup>1</sup>**

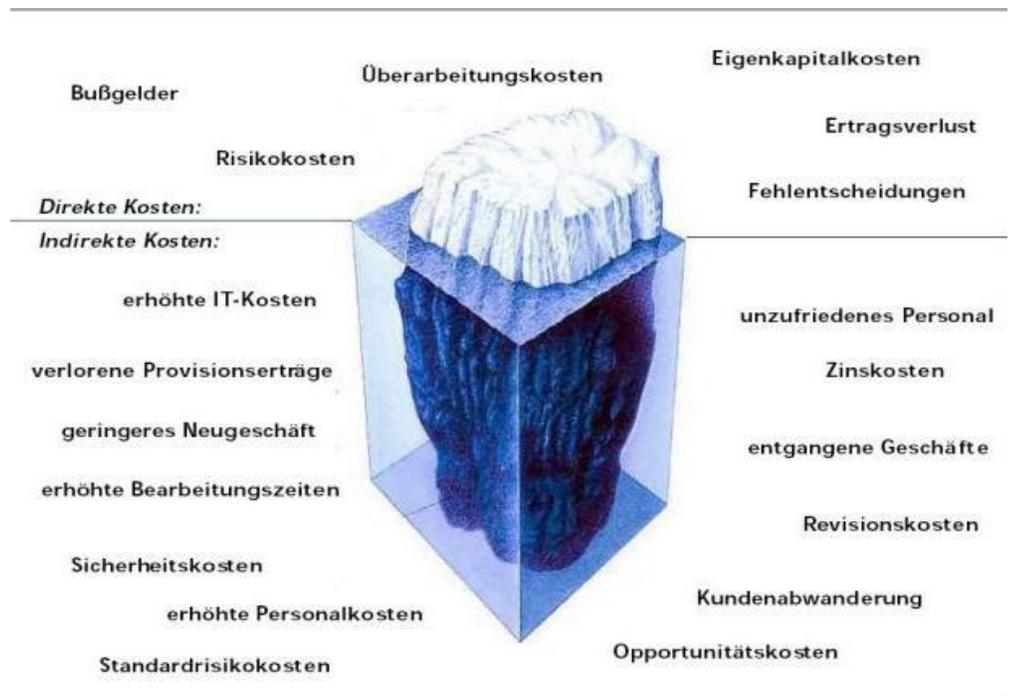


Abbildung 1 – (Pütter, 2007)

Ein Bewusstsein für das Daten-Management, sei bei den Banken vorhanden, so würden nicht umsonst die Kassen bei Business Intelligence Software Anbietern klingeln, so Christiane Pütter. Jedoch ist ein wesentlicher Knackpunkt das fehlende Verständnis dafür, dass die Ergebnisse von BI-Lösungen immer nur so gut sein können, wie die verarbeiteten Daten. Aufgrund anschwellender Einführungen der IT-Systeme in Unternehmen, sowie die Einbeziehung unternehmensexterner Daten (z.B. Big Data), gewinnen die „richtigen Daten“ als Basis immer mehr an Bedeutung. So soll die vorliegende Arbeit nicht nur Datenqualitätsmanagement betrachten, sondern den Fokus auf den Bereich Business Intelligence lenken, wo der Schmerz fehlerhafter Daten am größten ist.<sup>2</sup>

Bisher liegt kein allgemein gültiger Ansatz für eine erfolgreiche Datenqualitätsverbesserung, noch einen Ansatz für eine Kosten-Nutzen-Betrachtung vor.

<sup>1</sup> (Pütter, 2007)

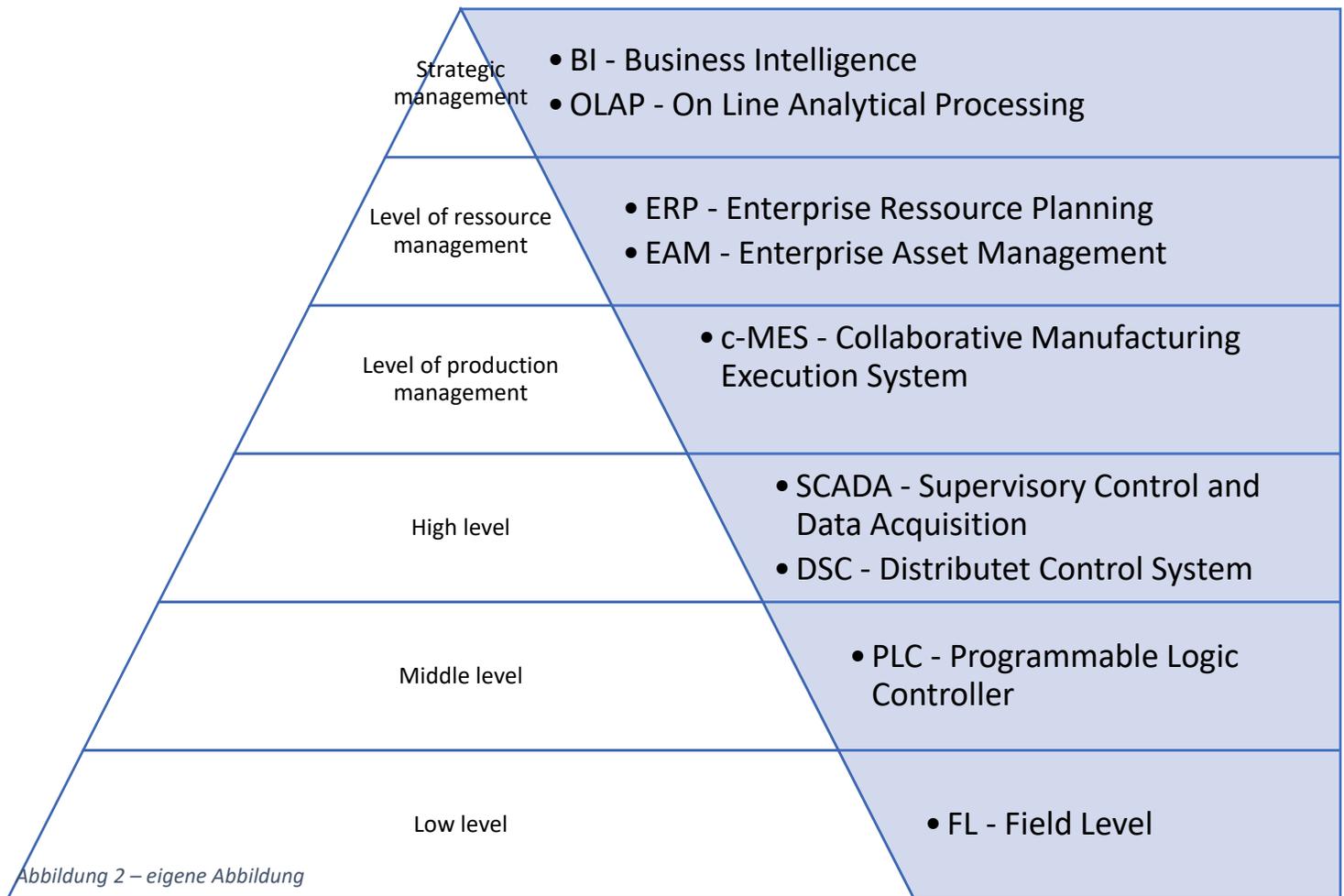
<sup>2</sup> vgl. (Detlef Apel, 2015)a

In dieser Arbeit sollen signifikante Ursachen, Symptome und Vorgehensweisen bei der Beantwortung der Frage, welche Daten in einem BI System vorhanden sein müssen, damit hier ein lukratives Kosten-Nutzen-Verhältnis entsteht.

**Forschungsfrage:** Welche Datenqualität wird hinsichtlich des Aufwandes und Nutzens im Rahmen eines zur Unternehmenssteuerung taugliches BI-System benötigt

Die zugrunde liegenden Daten sind für ein taugliches BI-System maßgeblich und ausschlaggebend um ein gewinnbringendes (sog. lukratives) Kosten Nutzen Verhältnis zu erzeugen. Somit wird, zur Beantwortung der Forschungsfrage genau das analysiert, wie in einem Dokument Grundlagen und Lösungen zusammengefasst werden kann. Die Antwort auf, welche Datenqualität benötigt wird, geht aus der nachfolgenden Arbeit heraus und wird am Ende zusammengefasst.

## Aufbau der Daten-Hierarchie bzw. Informationsfluss



### APMS – Automated Production Management System

- Low level
- Middle level
- High level

APMS bezieht sich hierbei auf die Produktionsebene der Maschinen im Feld.

### APCS – Automated Process Control System

- Level of production management
- Level of resource management
- Level of Strategic management

APCS befasst sich mit den Prozessen, welche die Daten aus den APMS-Systemen beziehen und diese verarbeitet.

In der Abbildung 2 wird ein Überblick über die Schichten in einer Datenhierarchie gegeben. Nachfolgend werden die einzelnen Systeme genauer beschrieben.

## Low Level

Auf dieser Ebene sammeln Computer Daten von Sensoren und Aktoren (Input/Output), bereiten sie auf, verarbeiten sie in Verarbeitungscomputern, implementieren Kontrollmechanismen und übermitteln spezifische Daten an Host-Computer. Darüber hinaus werden die Daten verarbeitet und grafisch dargestellt.<sup>3</sup>

## Middle Level

Eine speicherprogrammierbare Steuerung (SPS) ist ein Computer für die kontinuierliche Automatisierung, Überwachung und Steuerung von industriellen Kontrollsystemen (ICS). SPS arbeitet in Echtzeit, verarbeitet zyklisch Automatisierungsdaten und überwacht den Status nach Kundenvorgaben.<sup>4</sup>

## High Level

Im Allgemeinen, wird im High Level Bereich Prozesse geleitet, um zu Bedienen und zu Beobachten oder auch Messwerte archivieren zu können.

SCADA-Prozesse erstrecken sich über Gebiete, die tausende von Quadratkilometern groß sein können, und haben Dimensionen, die Hunderte, gelegentlich tausende von Kilometern lang sein können. Mittlerweile ist die Technologie ausgereift. SCADA umfasst unter anderem Softwarepakete, die in ein System aus Hardware und Software integriert werden können, um die Sicherheit und Effizienz des Betriebs dieser großen Prozesse zu verbessern.

---

<sup>3</sup> vgl. (ITWissen, 2022)a

<sup>4</sup> vgl. (ITWissen, 2022)b

### Level of production management

Eine Plattform für das Produktionsmanagement in der Fertigung auf der Grundlage von C-MES (**Collaborative Manufacturing Execution System**) ist für alle Aspekte des Produktionsprozesses in der Werkstatt zuständig. Sie hat die Rolle eines Bindeglieds zwischen der Unternehmensplanung und der Produktionssteuerung, zum Übergang von Verwaltung zu Produktion. So ist es wichtig, ein C-MES zu schaffen, dass die Fertigungsausführung, die Informationsintegration und die Zusammenarbeit effektiv realisieren kann, um den Informationsprozess in diskreten Fertigungsunternehmen zu fördern, eine reaktionsschnelle Fertigung zu bewerkstelligen und die Managementebene zu verbessern.<sup>5</sup>

### Level of ressource management

Ressourcen sind unerlässlich, um Ihr Ziel zu erreichen, sei es, dass Sie eine Aufgabe oder ein Projekt abschließen, oder dass sie analysieren, was dafür notwendig ist. Zu den Ressourcen gehören finanzielle Mittel, Personal, Räumlichkeiten, Ausrüstung, Technologie und Zeit. Ziel des Ressourcenmanagements ist es, die beste Kombination von Ressourcen zu nutzen, um die Anforderungen zu erfüllen und gleichzeitig zu erkennen, dass dieselben Ressourcen wahrscheinlich auch an anderer Stelle im Unternehmen benötigt werden.<sup>6</sup>

### Level of strategic management

Strategisches Management beschäftigt sich mit der fortlaufenden Planung, Überwachung, Analyse und Bewertung aller Notwendigkeiten, die eine Organisation benötigt, um ihre Ziele zu erreichen. Veränderungen im geschäftlichen Umfeld machen es erforderlich, dass Organisationen ihre Erfolgsstrategien ständig überprüfen. Der Prozess des strategischen Managements hilft Organisationen, eine Bestandsaufnahme ihrer gegenwärtigen Situation zu machen, Strategien zu entwickeln, sie einzusetzen und die Wirksamkeit der implementierten Managementstrategien zu analysieren. Eine SWOT-Analyse (Strengths, Weaknesses, Opportunities, Threats) ist eine Form des strategischen Managements, in der die Stärken, Schwächen, Möglichkeiten und Gefahren gesammelt werden.<sup>7</sup>

---

<sup>5</sup> vgl.(Huifen Wang, 2014)

<sup>6</sup> vgl.(planview, 2022)

<sup>7</sup> vgl.(Tucci, 2022)

## Warum ist Datenqualität ein Faktor für ein funktionierendes BI-System?

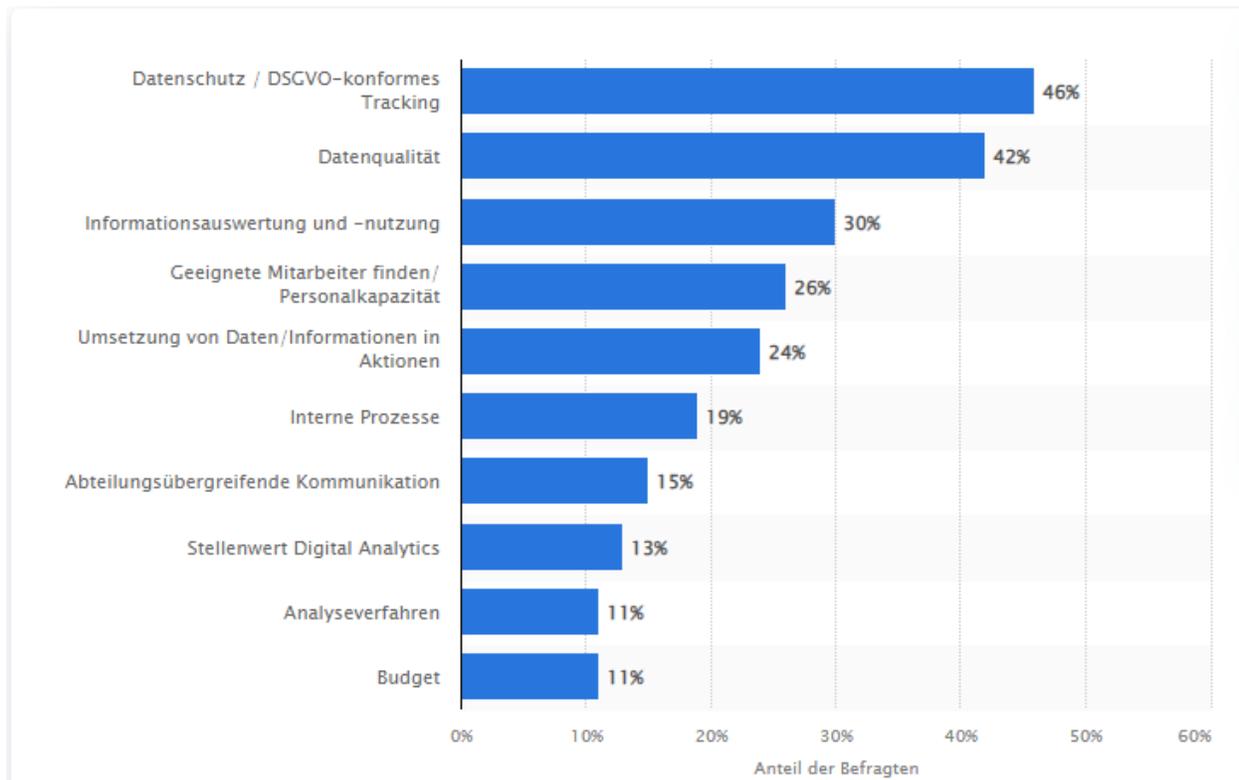


Abbildung 3 - (statista, 2022)

Wie in Abbildung 3 von Statista ersichtlich, ist Datenqualität nach Datenschutz eine der größten Digital Analytics Herausforderungen in Unternehmen im Jahre 2022 und kommt somit direkt nach Datenschutz Thematiken. Die Umfrage wurde vom 10. bis 21. Januar 2022 erhoben und zeigt das Ergebnis zu den größten Herausforderungen hinsichtlich der Digital-Analytics aus Unternehmenssicht.

## Definition Datenqualität <sup>8</sup>

Datenqualität (DQ) wird im heutigen Sprachgebrauch häufig mit Informationsqualität als Synonym benutzt. Jedoch bestehen ein paar maßgebliche Unterschiede. Denn Datenqualität beschreibt alle Ausprägungen von Qualitätsmerkmalen eines Datensatzes oder Bestandes. Daten sind eines der wertvollsten Güter innerhalb eines Unternehmens. Besonders wenn Entscheidungen anhand fälschlicher Daten getroffen werden, welche einen negativen Verlauf und zugleich erhöhte Kosten für das Unternehmen herbeiführen.<sup>9</sup>

Wenn Leute über Information bzw. Datenqualität reden, wird oft nur die Genauigkeit einer Variablen gemeint. Jedoch ist Informationsqualität viel mehr als das, denn auch andere maßgebliche Faktoren wie Vollständigkeit, Konsistenz oder Einheit werden benötigt, um die Qualität von Informationen klassifizieren zu können.

Datenqualität (DQ) ist ein wichtiges Thema für moderne Organisationen, vor allem für die Entscheidungsfindung auf der Grundlage von Informationen, die Lösungen wie CRM, Business Analytics und Big Data nutzen. Um qualitativ hochwertige Daten zu erhalten, ist es zudem notwendig, Methoden, Prozesse und spezifische Techniken zu implementieren. Dies ermöglicht Informationen wie ein Produkt zu behandeln.

DQM ist das aktive Management der Qualität von Daten, damit sie für den Geschäftszweck nützlich sind. Schlechte Datenqualität kann zu schlechten Geschäftsentscheidungen führen. DQ kann in Genauigkeit, Vollständigkeit, Gültigkeit, Einzigartigkeit, Zeitlosigkeit und Konsistenz gemessen werden.

---

<sup>8</sup> vgl.(Knut Hildebrand, 2022)

<sup>9</sup> vgl.(talend, 2022)

## Datenqualität im BI

Eine nicht ausreichende Datenqualität hat in einem Unternehmens- oder auch im organisatorischen Umfeld maßgeblich Einfluss auf spezifische Bereiche. So ist eine Datenqualitätsverbesserung zu jedem Zeitpunkt eine Überlegung wert. Jedoch ist diese Verbesserung ein eher komplexes Unterfangen und sollte auf jeden Fall durch eine **Kosten-Nutzen-Analyse** verifiziert werden. Eine solche Betrachtung wird im Laufe dieser Arbeit noch weiter thematisiert.

Im Umfeld eines BI-Projektes, eignen sich besonders die Kennzahlen, um die Datenqualität messbar zu machen und überwachen zu können. Jedoch birgt eine automatisierte Berechnung der Kennzahlen die Gefahr, dass nur begrenzt eingeschätzt werden kann, wie der tatsächliche Zustand der Datenqualität ist.

## Warum brauchen wir gute Datenqualität im BI

Einen initialen Anstoß für eine Verbesserung der Datenqualität ist oftmals getrieben durch die Geschäftsfelder Wettbewerb, Kosten, Risiken, Compliance und Steuerung. Ein Grad der Aufwendung ist immer in Abhängigkeit der beabsichtigten Anwendung zu setzen.

So ist im Bereich Compliance eine hohe Datenqualität erforderlich, damit eine Revision und regulierende Vorgaben erfüllt werden können.

Eine Reduzierung des operativen Risikos in den Unternehmen kann durch eine frühzeitige Behebung von Datenqualitätsproblemen vorgenommen werden. So ist die Motivation einer Datenqualitätsverbesserung im Bereich Kosten und Risiken hoch, denn es können Durchlaufzeiten und somit Kosten reduziert werden. Besonderes Augenmerk liegt hierbei auf der Datenmigrationen zwischen den Anwendungen, denn Projekte werden immer agiler und es bleibt somit wenig Zeit für fehlerhafte Daten.

Auch im heutigen Wettbewerb ist eine konstante Verbesserung der Qualität der Daten unabkömmlich. Beispielsweise erhöht der Customer Service die Chance, lauernde Unzufriedenheiten der Kunden in eine langfristigen Beziehungen umzuwandeln, um somit verhindern zu können, dass zuverlässige und treue Kunden zur Konkurrenz abwandern.<sup>10</sup>

Bei der Steuerung des Unternehmens müssen sich Mitarbeiter auf die Qualität der Informationen verlassen können, denn diese wirkt sich direkt auf die Qualität der Entscheidungen im Unternehmen aus. Daten von guter Qualität in diesem Bereich, können einen signifikanten Einfluss auf die langfristige Ausrichtung eines Unternehmens haben.

---

<sup>10</sup> vgl.(DataSelf, 2022)

## Messbarkeit von Datenqualität

Damit Qualität oder in diesem Fall Datenqualität messbar sein kann, bedarf es verschiedener Kriterien, diese werden dann den Daten zugeordnet. Man spricht hierbei von Qualitätskriterien.

Qualitätskriterien beschreiben Kriterien, mit denen die Qualität definiert sein muss, sodass diese exakt formuliert und objektiv überprüfbar sind.<sup>11</sup>

Aufgrund praktischer Erfahrungen werden Qualitätskriterien definiert. Der Erfüllungsgrad, welcher durch den Datennutzer ermittelt wird, setzt messbare Qualitätskriterien zugrunde. Ziel ist es, einen möglichst hohen Erfüllungsgrad zu erreichen, dies wird erreicht, indem Sollwerte definiert werden. An diesen Sollwerten wird schlussendlich die Datenqualität gemessen.<sup>12</sup>

---

<sup>11</sup> vgl.(ProjectBase, 2022)

<sup>12</sup> vgl.(Detlef Apel, 2015)b

## Datenqualitätskriterien<sup>13</sup>

Alle Daten sollten nach folgenden Kriterien überprüft werden um sie anschließend ergänzen oder abändern zu können.

**Konsistenz** beschreibt, dass es keine logischen Fehler oder Widersprüche innerhalb eines Datensatzes geben darf.

- User-Feedback
- Schnittstellenreview
- Vergleichswerte zum Abgleich zwischen Systemen
  - o Summe über Umsätze
  - o Kundenanzahl in unterschiedlichen Kategorien
- Zusammenhänge zwischen Systemen
  - o Geburtsjahrabgleich ist konsistent mit Alter in einem anderen System

**Zuverlässigkeit** der Daten ist gewährleistet, wenn ihr Ursprung klar nachvollziehbar ist.

- Anzahl dokumentierter Transformationen
- Anzahl durchgeführter Abgleiche von Daten

**Vollständigkeit** ist die generelle Verfügbarkeit von Inhalten, welche ein Anwender benötigt.

- Vollständigkeit der implementierten Konzepte
- Anzahl fehlender Datenwerte
- Anzahl nicht geladener Schnittstellen

**Genauigkeit** beschäftigt sich mit dem Faktor, wie akkurat ein Wert ist, je nach Anwendungsfall mit den entsprechenden Nachkommastellen beispielsweise.

In einer **Redundanzfreiheit** befinden sich keine Duplikate innerhalb eines Datensatzes.

Durch **Einheitlichkeit** werden Daten fortlaufen gleich abgebildet.

**Eindeutigkeit** ist die genaue Interpretierbarkeit, welche durch Metadaten umgesetzt werden kann.

---

<sup>13</sup> vgl.(Detlef Apel, 2015)c

## Messung der vorhandenen Datenqualität<sup>14</sup>

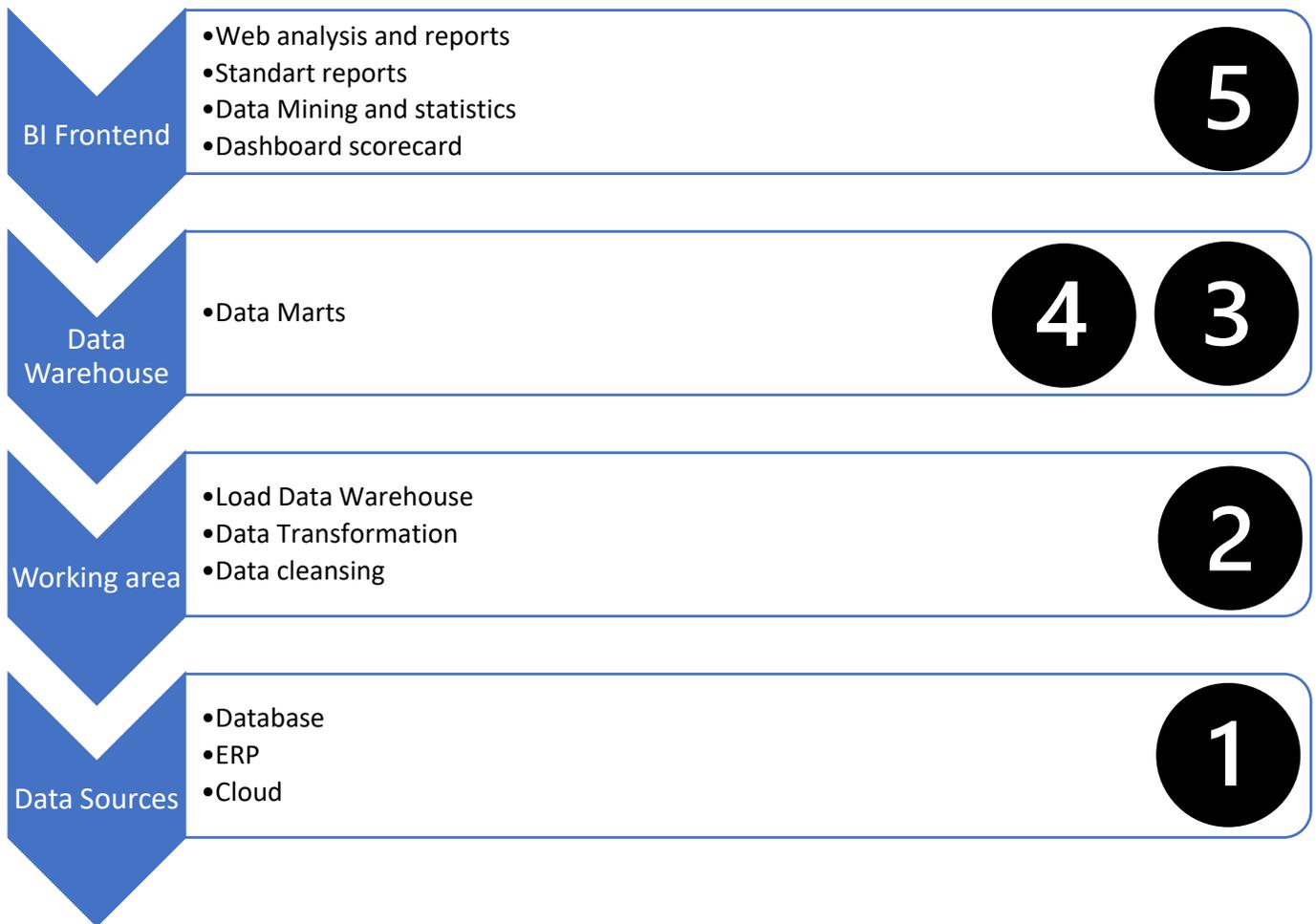


Abbildung 4 – eigene Abbildung, vgl. (Knut Hildebrand, 2022)

### Messpunkte innerhalb der Referenzarchitektur

- 1. Data Sources**
- 2. Working Area**
- 3. DWH**
- 4. Data Marts**
- 5. BI-Frontend**

<sup>14</sup> vgl. (Knut Hildebrand, 2022)b

Um an betitelten Messpunkten Daten abzugreifen, bedarf es inhaltlich einem:

- Wertevergleich zur Vorperiode
- Verteilung von Werten über Wertebereiche
- Validierung von Minimum und Maximum
- Planwerte gegenüber Sollwerten

Eine Bewertung der Qualität von Daten erfolgt unter Zuhilfenahme der Datenqualitätskriterien. Eine Nutzung von Kennzahlen ist so im Bereich von Business-Intelligence Anwendungen eine Möglichkeit Datenqualität messbar zu machen. Eine Aufstellung von Kennzahlen kann über statistische Nützbarmachungen einzelner Attribute erfolgen. Aber auch über umfassende Abbildungen von Geschäftsregeln, welche Zusammenhänge zwischen Entitäten und Domänen auslegen. Jedoch ist, damit eine erfolgreiche Überwachung der Daten erfolgen kann, das Verständnis über die Leistungsmöglichkeiten solcher Kennzahlen von Bedeutung. Auch in der Qualitätsplanung finden Kennzahlen ihren Einsatz.<sup>15</sup>

### Vorteile der Kennzahlen

Ein wesentlicher Nutzen in der Anwendung von Kennzahlen zur Messung der Datenqualität liegt in der Quantifizierung klar dargelegter Problemstellungen. Kennzahlen können zur initialen Qualifizierung von Daten verwendet werden. In durchlaufenden Plan-Ist-Abweichungsanalysen können Problembereiche aufgezeigt werden. Innerhalb von Datenbeladungsprozessen helfen Kennzahlen dabei, fehlerhafte Datenverarbeitungen aufzuzeigen.

### Nachteile der Kennzahlen

Ein Verständnis, sowie ein Bewusstsein der Grenzen von Kennzahlen ist oft bei den Anwendern nur begrenzt vorhanden. Zudem werden nicht immer alle Datenqualitätskriterien gemessen. Kennzahlen können keine einheitliche Sichtweise auf historische Vergleiche der Kennzahlen darstellen, aufgrund von Einflussfaktoren, wie beispielsweise eine Neueinführung eines Produktes

---

<sup>15</sup> vgl.(Knut Hildebrand, 2022)c

## Erklärung der Messpunkte

Nachfolgend eine Erklärung der einzelnen Messpunkte, welche in Abbildung 4 zu sehen sind.

### **1. Data Sources**

Eine Erhöhung der Datenqualitätsprüfungen in Quellsystemen erhöht die Qualität der Daten in den nachgehenden Systemen. Jedoch hängt eine Implementierung von Kennzahlen im Wesentlichen von der Machbarkeit in diesem Bereich ab. Eine Prüfung für die Erfassung der Daten ist in der Umgebung von Quellsystemen besonders sinnvoll. Auch die Überprüfung der oftmals bereits vorhandenen Statistiken im Umfeld eines Datenbankmanagementsystems (DBMS), dass die Anzahl der Datensätze visualisiert, ist eine Möglichkeit Datenqualität in Quellsystemen zu messen.<sup>16</sup>

### **2. Working Area**

Am 2. Messpunkt werden die Kennzahlen mittels Schnittstellen in die Working Area übernommen. So werden Statistiken der ETL-Prozesse mitgeschrieben und als Kennzahlen abgespeichert. Eine weitere Möglichkeit ist ein automatisierter Vergleich aktueller Fehlerlisten mit den Fehlerlisten vorheriger Beladungen und die damit zusammenhängende Ausgabe der jeweiligen Differenzen für eine Überprüfung der Daten. Dieser Messpunkt ermöglicht eine Berechnung trivialer Kennzahlen, wie die Anzahl der Null-Werte je Attribut. Darunter fallen Motive wie bspw. die Prüfung, ob in Transaktionsdaten nur Kunden angeliefert werden, welche bereits in den Stammdaten vorhanden sind.<sup>17</sup>

### **3. Data Warehouse**

Dieser Messpunkt ist besonders geeignet für die Aufstellung von Kennzahlen, welche aus mehreren Datenbereichen oder unterschiedlicher Quellsysteme entstammen. So können Konsistenzprüfungen zwischen Quellsystemen vorgenommen werden, welche Daten an das unternehmensweite Data Warehouse liefern. Ergebnisse der Kontrolle und Messung, werden in dafür vorgesehene Bereiche im Data Warehouse (kurz DWH) abgelegt. Fachlich spezifische Kennzahlen werden in den jeweiligen Data Marts berechnet.<sup>18</sup>

---

<sup>16</sup> vgl.(Detlef Apel, 2015)d

<sup>17</sup> vgl.(Detlef Apel, 2015)e

<sup>18</sup> vgl.(Hahne, 2014)a

#### **4. Data Marts**

Ein Data Warehouse ist eine einfache Form eines Datenspeichers, der sich auf ein einziges Thema, oder einen einzigen Geschäftsbereich konzentriert. Mit Data Maps können Teams schneller auf Daten und Informationen zugreifen, da sie keine Zeit mit der Suche nach komplexeren Data Warehouses oder der manuellen Erfassung von Daten aus verschiedenen Quellen verbringen müssen. Die Kennzahlen in diesem Gebiet enthalten Informationen, die im Kreis der Beladung von Data Marts anfallen. So ist die Berechnung dieser Daten oft ein Ergebnis aus Aggregationen, oder aus dem Fachbereich spezifischen Logiken.<sup>19</sup>

#### **5. BI-Frontend**

Eine einfache Form der Berechnungen der Kennzahlen ist eine solche, in Berichten. Hier kann ein einfacher Austausch zwischen operativen Systemen durchführbar gemacht werden, die dann definierte Summengrößen automatisiert an die leitenden Personen der Quellsysteme melden. Die Häufigkeit der durchgeführten Vergleiche ist mit den fachlich und technisch verantwortlichen Personen abzugleichen.

---

<sup>19</sup> vgl.(Oracle, 2022a)

## Welche Datenqualität wird benötigt

Der Erfolg von Projekten vor allem im BI-Bereich wird oft an der Datenqualität festgemacht. Denn ein erfolgreiches Unternehmen erfordert hochwertige Daten. So können „schlechte Daten“ viele Formen haben und somit auch einen negativen Einfluss auf ein Unternehmen. Einen Überblick hierzu geben, bezogen auf ein BI-System die folgenden wesentlichen Ursachen schlechter Datenqualität.<sup>20</sup>

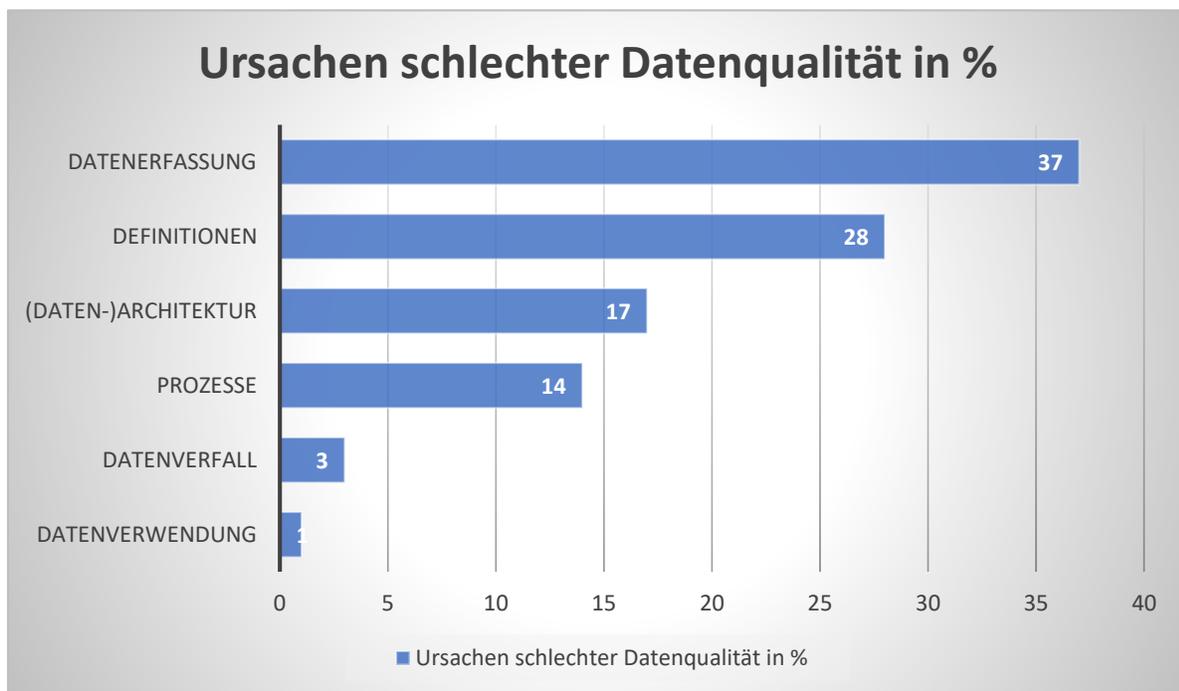


Abbildung 5 – eigene Abbildung, vgl. (Enderlein, 2015)

Um eine korrekte Datenerfassung in meist hochkomplexen Anwendungen sicherstellen bedarf es ein Bewusstsein für Fehlerquellen, anfangend bereits bei Benutzerfehlern, wie etwa Tippfehler, bis hin zu fehlerhaften Daten.

Angefangen bei Benutzerfehlern, besteht ein fehlender Beschäftigungsanreiz für die korrekte Erfassung der Daten, sowie ein Bewusstsein der Auswirkungen schlechter Datenqualität für eine fortlaufende und aufbauende Informationsverarbeitung.

Eingabemasken bilden aufgrund einer oft unzureichenden oder unverständlichen Oberfläche eine weitere Quelle für Probleme bei der Datenerfassung. Als Vorlage hierzu dient eine intelligente Eingabemaske, welche direkt sicherstellt, dass beispielsweise ein Geburtsdatum

<sup>20</sup> vgl.(Ellsworth, 2019)

kein „Frei-Text-Feld“ sein sollte. Aber auch Adressangaben können mit einer intelligenten Eingabemaske durch eine postalische Verifizierung direkt validiert werden.

Externe Daten wie beispielsweise zugekaufte Adressdaten, können im Umfeld einer BI-Anwendung eine weitere Fehlerquelle der Datenerfassung sein, welche es zusätzlich zu überprüfen gilt.

So ist zusammenfassend bereits bei einer Erfassung der Daten ein Potential zur Verbesserung von Daten möglich. Ein angepasstes Bewusstsein der Unternehmen bezüglich der korrekten Datenerfassung kann helfen bereits im Ursprung Fehlerquellen auszubessern.

### Prozesse

Mit der steigenden Komplexität von Prozessen steigt auch die Gefahr mangelhafter Prozesse bezüglich der Datenverarbeitung. Dies reicht von der ersten Erfassung der Daten, bis hin zur Erstellung von manueller Ergänzung, oder Erfassung innerhalb einer Datenerfassungskette.

Eine weitere Gefahr stellt Outsourcing dar, indem Verantwortungen abgegeben werden, dabei steigt gleichzeitig die Gefahr einer nicht korrekten Datenerfassung. So ist eine saubere Erfassung der Stammdaten unabkömmlich, speziell in Bereichen von Business Intelligence und Big Data.<sup>21</sup>

### Datenarchitektur / Schnittstellen

In viele Unternehmen hat sich über die Jahre die Anwendungslandschaft drastisch vergrößert. Es werden neue Funktionen hinzugewonnen, welche in ihrem Ursprung nicht geplant waren. Dies bringt eine heterogene Anwendungslandschaft mit redundanten und inkonsistenten Daten mit sich. Betroffen hierbei sind vor allem die Schnittstellen, welche sich durch verschiedene Anwendungen ergeben. Ein im Ursprung entstehendes Problem vervielfacht sich im Zuge der Datenverarbeitung bis in die BI-Anwendungen. Schnittstellen zwischen verschiedenen Anwendungen führen unweigerlich zu Problemen, so muss ein besonderes Augenmerk immer auf den Datenflüssen zwischen den Anwendungen liegen.<sup>22</sup>

---

<sup>21</sup> vgl.(Detlef Apel, 2015)f

<sup>22</sup> vgl.(Detlef Apel, 2015)g

## Definition

In einer Konzeption vieler Anwendungen, stecken oft nur einzelne Bereiche eines Unternehmens dahinter. So werden Bedürfnisse des Gesamtunternehmens nicht berücksichtigt. Eine Beschreibung der Daten beruht auf Sichtweisen aus den Fachbereichen. Eine generelle Definition von Feldformaten, ist durch unterschiedliche Sichtweisen für identische Felder, in Business Intelligence Anwendungen unvereinbar. <sup>23</sup>

Viele durch Anwender wahrgenommene Datenqualitätsprobleme lassen sich auf mangelhafte Kommunikation innerhalb eines Unternehmens zurückführen. Jedoch ist ein abgestimmtes Vokabular eine Minimalvoraussetzung, um weitere Fortschritte in diesem Gebiet zu erreichen. Durch interne Kommunikation sollen Informationen in erster Linie weitergegeben werden, welche für alle, oder einen Teil der Mitarbeiter Relevant sind. <sup>24</sup>

## Datenverwendung

Eine inkorrekte Nutzung von Anwendungen ist ein maßgeblicher Punkt. Denn sollte sich ein Fehler in einem Datensatz ergeben, so sollten Datenkorrekturen immer in den Quellenanwendungen vorgenommen werden. Zugleich ist das Fehlen von Know-how ein weiterer Breaking Point. Mangelhaftes Wissen für beispielsweise eine korrekte Erstellung von Berichten in diversen Anwendungen ist eine entscheidende Voraussetzung, welche gewährleistet sein sollte.

## Datenverfall

Im Laufe der Zeit treten speziell im BI-Bereich Verfalls-Probleme auf. Ein Beispiel hierfür sind die Telefondaten, welche sich im Laufe der Zeit ändern können. Datenverfall führt so unweigerlich zu einer Ungültigkeit der Daten. So müssen Prozesse eingeführt werden, welche heikle Daten erkennen und aktuell halten.

## Zusammenfassung

Ein fehlendes Bewusstsein für die Auswirkungen schlechter Datenqualität steht als Problem bei Entscheidungsträgern an erster Stelle. Abhilfe kann hier eine Kosten-Nutzen-Rechnung schaffen welche die „Hard-Points“ beleuchtet und bewertet. Mehr hierzu im Laufe dieser Arbeit.

---

<sup>23</sup> vgl.(Detlef Apel, 2015)h

<sup>24</sup> vgl.(Hannig, 2022)

## Symptome schlechter Datenqualität<sup>25</sup>

„Fast 40 Prozent aller Unternehmen sind mit der Qualität ihrer erfassten und bearbeiteten Geschäftspartnerdaten unzufrieden [...] Allerdings ergreift jeder Fünfte keine Maßnahme zur Verbesserung.“(Förg, 2019) So sind häufige Symptome für schlechte Datenqualität:

- Kunden oder Geschäftspartner tauchen mehrfach im System auf
- Hohe Rückläuferquoten
- Mitarbeiter beschwerten sich über manuellen Rechercheaufwand
- Cross-Selling-Möglichkeiten werden verschenkt
- Wenig Response Daten aus dem Bereich Marketing
- Gesetzliche Vorgaben können eventuell nicht eingehalten werden
- Mangelnde Planungssicherheit

In den USA betragen die Kosten aufgrund von schlechter Datenqualität nach einer Studie von TDWI (The Data Warehousing Institute) ca. 600 Milliarden US-Dollar pro Jahr<sup>26</sup>

## Kosten schlechter Datenqualität

Wie bereits beschrieben, verursacht schlechte Datenqualität oft Kosten in Milliarden Höhe (siehe Symptome schlechter Datenqualität). Auch das Abbrechen von BI oder Customer-Relationship-Management-Projekten verursacht hohe Kosten.<sup>27</sup> So wird dies durch eine weitere Gartner Studie belegt, dass der geschätzte durchschnittliche durch Datenqualitätsprobleme verursachte Verlust jährlich 8,2 Millionen US-Dollar verursacht. Von 150 Befragten Unternehmen gaben 4% an, jährlich Verluste von ca. 100 Millionen US-Dollar zu haben.<sup>28</sup>

Ein weiterer Punkt beschreibt die Kosten strategischer Fehlentscheidungen, durch das Management aufgrund fehlerhafter Datenmaterialien.

Daher stellt sich die Frage wie eine Verbesserung der Datenqualität lukrativ wird. Diese Fragen müssen sich die Unternehmen unter aktuell steigendem Kostendruck stellen.

---

<sup>25</sup> (Matthias Förg, 2019)

<sup>26</sup> vgl.(Eckerson, 2002)

<sup>27</sup> (Keizer G, Mai 2004)

<sup>28</sup> vgl.(Friedman, 2010)

## Inwieweit kosten Fehler Geld

Die Dresdner Bank hat Wertpapierdepots von Kunden zweimal in Euro umgerechnet. Dadurch erschienen in den Kontoauszügen auf einmal nur noch ein Viertel des ursprünglichen Guthabens.<sup>29</sup>

Ein Audit brachte in einem europäischen Unternehmen auf, dass ca. 4% der Bestellungen keine zugehörige Rechnung besaßen. So bedeutete dies für eine Unternehmen, mit einem Umsatz von 2 Milliarden Dollar einen Verlust von 80 Millionen US-Dollar.

Die obigen Beispiele sind nur wenige von vielen, daher zeigt sich anschaulich, dass der Nutzen aus verbesserter Datenqualität sich durchaus bestimmen lässt.<sup>30</sup> So ist der oft vorangeschobene Punkt des Managements, dass sich Datenqualitätsmanagement nicht quantifizieren lässt, hinsichtlich der generierten Kosten und Nutzenpotentiale hinfällig. So ist ein Business Case zu entwerfen, nicht komplexer als jede andere Investition.

Regulatorische Anforderungen bilden einen maßgeblichen Punkt in der Datenqualität, welche mit der Zeit immer mehr an Bedeutung gewinnen. Dies erfordert ein proaktives Management der Datenqualität.

Beschränkungen aus der Sicht des Unternehmensbudgets machen eine Datenqualitätsverbesserung in allen Bereichen jedoch sehr schwer. Daher muss eine Priorisierung vorgenommen werden, welche sich nach den wichtigsten Projekten innerhalb

des Unternehmens, mit ihrer jeweiligen Auswirkung richtet. Die Kosten für die Sicherstellung von Datenqualität lassen sich, wie in

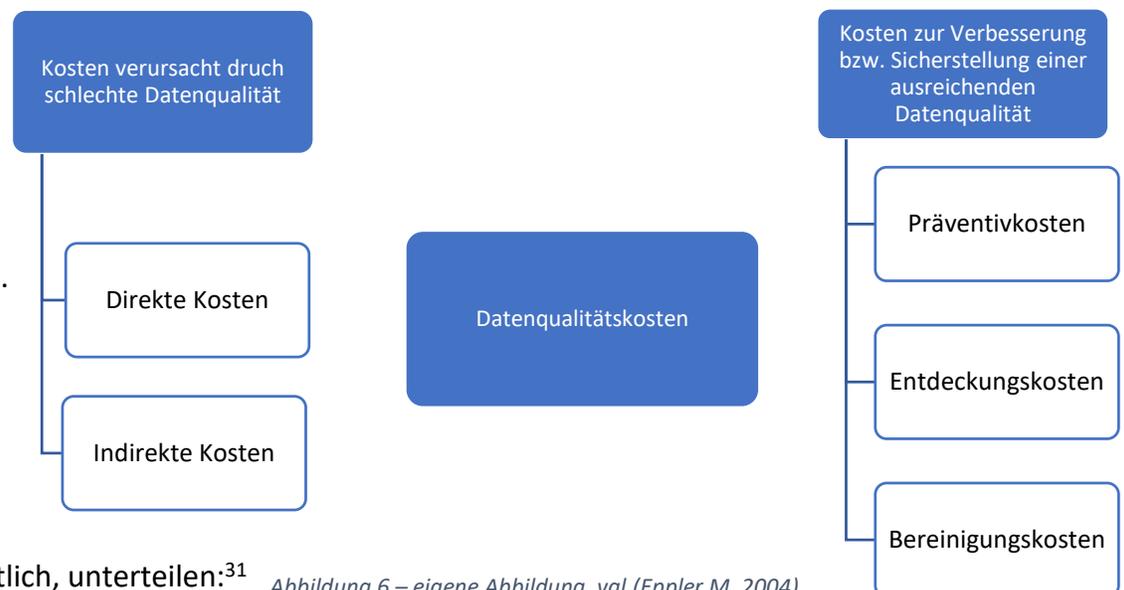


Abbildung 6 ersichtlich, unterteilen:<sup>31</sup> *Abbildung 6 – eigene Abbildung, vgl.(Eppler M, 2004)*

<sup>29</sup> vgl.(Bank, 1999)

<sup>30</sup> vgl.(English, 1999b)

<sup>31</sup> vgl.(Eppler M, 2004)

## Direkte Kosten

Direkte Kosten entstehen sofort, welche unmittelbar durch schlechte Datenqualität hervorgerufen werden. Ein Beispiel hierfür ist eine falsche Währungsumrechnung im Quellsystem, bei dem eine Reihe von Zahlen falsch an das Data Warehouse weitergegeben wurden. Um dies zu korrigieren, müssen die Daten neu geladen werden.

## Indirekte Kosten

Indirekte Kosten entstehen durch falsche Entscheidungen oder Aktionen, wie in folgenden Bereichen ersichtlich:

- Umsatzeinbußen
  - Falsche Preisauszeichnungen
- Verschwendung von Budget
  - Versendung von Werbebriefen an nichtexistierende Adressen
- Fehlentscheidungen
  - Entscheidungen basierten auf falschen Zahlen, nun wurde das Potential einer Zielgruppe unterschätzt
- Imageverlust
  - Wiederholte Unzufriedenheit führt zu Kundenabwanderung durch schlechte Datenqualität
- Gerichtliche Auseinandersetzung
  - Falsche Daten führen zu Rechtsstreitigkeiten zwischen Kunden/Staat und Unternehmen.
- Produktionsmangel
  - Falsch eingestellte Maschinen führen zu teureren Nacharbeitsprozessen oder Rückrufaktionen
- Vertrauensverluste
  - Vertrauensverluste in die eigenen Daten innerhalb des Unternehmens führt zu endlosen Diskussionen über richtige Kennzahlen

## Präventivkosten

Die Eingabe falscher Daten kann durch die Sensibilisierung der Mitarbeiter, bezüglich der Datenqualität vorab minimiert werden. Gezielte Schulungen der Mitarbeiter können hierbei große Fortschritte erwirken. Dies führt jedoch unweigerlich zu erhöhten Kosten, welche aber sehr gut investiert sind.

## Entdeckungskosten

Durch Plausibilitätsberichte lassen sich Unstimmigkeiten in den Daten aufdecken und erkennen. Ist beispielsweise das monatliche Bestellvolumen bekannt, können utopisch hohe Werte direkt als Fehlbuchung identifiziert werden. Eine regelmäßiges Data-Profiling hilft hierbei Datenqualitätsmängel frühzeitig zu erkennen. Im Qualitätsmanagement ist von einer **Zehnerregel** die Rede. Diese besagt, dass Kosten für einen unentdeckten Fehler, sich erhöhen von Stufe zu Stufe der Wertschöpfung um den Faktor 10.<sup>32</sup>

## Bereinigungskosten

Wenn falsche Daten im Data Warehouse aus verschiedensten Quellsystemen landen, sind diese zu bereinigen. Hierfür müssen sogenannte regelmäßige Update Routinen implementiert werden. Eine Korrektur erfolgt hierbei in der Praxis über eine Kombination aus manuellen und automatisierbaren Tätigkeiten.

**Eine Prävention hat im Schnitt um ca. 2/3 geringere Gesamtkosten zur Folge.** <sup>33</sup>

## Business Case Betrachtung

Ein Business Case im Bereich der Datenqualität, bietet keinen signifikanten Unterschied im Vergleich zu jeglichen anderen Projekten und Vorhaben. Bei der Entscheidung, welche das Management zu treffen hat, muss eine klare Aufbereitung der Thematik vorhanden sein, sodass eine klare Antwort auf die WARUM Frage folgen kann. Eine Aufbereitung könnte wie in Abbildung 7 beispielhaft aussehen.

---

<sup>32</sup> vgl.(Florian, 2021)

<sup>33</sup> Vgl.(Redman, 2008)

## Business Case

Project Name	Project Manager
Client	Duration



## Kostenunterscheidungen im Business Case

### Initiale Kosten

Initiale Kosten sind Kosten, welche nur einmalig für Datenqualitätsbereinigungen oder für Implementierungsaufwand anfallen. So sollen Hardware oder Softwarekosten nur ein Beispiel hierfür sein.

### Laufende Kosten

Betriebskosten sind alle Kosten, die beim Betrieb eines Unternehmens anfallen. Beispiele hierfür sind: Personalkosten, Räumlichkeiten, Werbung, Logistik, Fahrzeuge, Bürokosten, Versicherungsprämien oder Finanzierungskosten. Zu den laufenden Kosten gehört beispielsweise ein laufendes Monitoring der Datenqualität.

## Big Data als neue Evolutionsstufe für BI

### Definition der Begrifflichkeit

Big Data ist kein klar definierter Begriff, so sind auch wissenschaftliche Definitionen nur in geringer Zahl erschienen.<sup>34</sup>

Big Data ermöglicht in Zeiten einer digitalen Revolution, Analysen von großen Datenmengen, um so Muster, Trends und Interaktionen zwischen Kunden und Unternehmen besser zu verstehen.<sup>35</sup>

Heutige Big Data Datenquellen lassen sich in folgende Bereiche Unterteilen:

- Internet
  - o Soziale Netzwerke
  - o Blogs
- IoT (Internet of Things)
  - o RFID
  - o Autonomes Fahren
  - o Smart Cities
  - o Smart Home
- Logfiles
  - o Server Logs
  - o Maschinen Logs
- Unstrukturierte Daten
  - o Dokumente
  - o Mails

Haupt Charakteristika von Big Data, werden oft als die vier V's bezeichnet. Volume, Velocity, Variety und Veracity.<sup>36</sup>

---

<sup>34</sup> vgl.(Klein D, 2013)

<sup>35</sup> vgl.(Andreas Krämer, 2022)

<sup>36</sup> vgl.(Gartner, 2011)

## Definition der Begrifflichkeit

So beschreibt in diesem Bezug **Volume** die stark zunehmende Datenmenge, im Vergleich zu vorhergegangenen existierenden Daten aus ERP-Systemen oder CRM-Datenbanken. So steigt auch das weltweite Datenvolumen, so Statista, von 2017 um das Zehnfache bis ins Jahr 2025 auf 163 Zettabyte an.<sup>37</sup> So sind beispielsweise Sensordaten in Fahrzeugen der heutigen Generation, das Internet mit sozialen Netzwerken aber auch E-Mail und Text größere neuartige Datenquellen. Bei großen Datenmengen ist ein klassischer Ansatz der Datenverarbeitung nicht mehr möglich. Hier ist eine rasche Verarbeitung mit der Extraktion sowie Prüfung von Information und Wissen erforderlich. Hierbei bieten Anbieter für ETL, Data Warehouseing oder Business Intelligence, Konnektoren zu einem Hadoop Filesystem an, um gezielt analyserelevanten Daten zu extrahieren.

**Velocity** zielt diesbezüglich auf die Geschwindigkeit der Produktion ab. So sollte diese möglichst zeiteffizient analysiert werden, um so Entscheidungsfindungen bestmöglich unterstützen zu können. Häufig wird die Automatisierung der Entscheidung angestrebt, denn zwei Drittel der C-Level Manager glauben, dass es gerade bei der Automatisierung, geschäftlicher Entscheidungen große Potentiale gibt.<sup>38</sup> Auch die Geschwindigkeit, mit der sich die Daten ändern, wird mit Velocity beschrieben.<sup>39</sup>

**Variety** von Daten steigt ebenfalls an, da mittlerweile Daten aus internen und externen Quellen stammen und unstrukturiert auftreten. 90% der gespeicherten Daten sind unstrukturiert. Unter strukturierten Daten versteht man Datensätze, die ein festes Schema haben. Aus diesem Grund lassen sich diese Daten normalerweise ohne großen Aufwand speichern, verwalten und analysieren (z.B. Speicherung zeilenbasiert in Tabellen). Semistrukturierte Daten enthalten partiell Informationen zu ihrer zugrundeliegenden Struktur. Sie lassen sich, z.B. bei der Verwendung des Austauschformates XML (Extensible-Markup-Language) verändern und erweitern. Im Gegensatz dazu, sind unstrukturierte Daten, meistens nur durch spezielle Verfahren handhabbar. Es kann zu großen Unterschieden bei der Verarbeitbarkeit der Daten kommen, wenn diese z.B. in unterschiedlichen Datenformaten vorhanden sind.<sup>40</sup> Eine Sentiment Analyse kann für einige Fragestellungen

---

<sup>37</sup> vgl.(Andreas Krämer, 2022)

<sup>38</sup> vgl.(Economist, 2012)

<sup>39</sup> vgl.(König C, 2018)

<sup>40</sup> vgl.(Kollmann, 2020)

bei der Erschließung von Information aus textuellen, unstrukturierten Daten unterstützen. So gibt es 2 verschiedene Ansätze eine Sentiment Analyse durchzuführen, einerseits Machine Learning oder wörterbuchbasiert. Herausforderungen bestehen dann, wenn Dialekte, Sarkasmus oder Rechtschreibfehler den Inhalt beeinflussen. In heutigen Unternehmen spielt Echtzeitverarbeitung eine entscheidende Rolle und kann somit einen maßgeblichen Wettbewerbsvorteil sichern. Machine-Learning hilft hierbei Daten schneller zu klassifizieren und zu analysieren.<sup>41</sup>

**Veracity** beschreibt die Richtigkeit der Daten und Datenqualität. Daten kommen oft aus verschiedensten Quellen, häufiger auch aus minderwertigeren und müssen danach aufwendig nachbereitet werden.<sup>42</sup>

Im Allgemeinen schöpfen Big Data Vorhaben ihren Erfolg aus dem Einsatz von Business Analytics, wie etwa Sentiment Analysen über Text Mining oder Data-Mining. Zielgruppen sind Marketing, Vertrieb, Service, Qualitätsmanagement, Supply Chain und Maintenance. So bestehen die meisten Use-Cases in Vorhersage von Wartungsintervallen, Erkennung von Risiken, Erkennung von Trends, Erkennung von Problemen in Lieferketten und die ganzheitliche Sicht auf Kunden. Ein erfolgreicher Einsatz von Big Data erfordert eine zentrale Organisation für Datenqualität und Governance im Unternehmen, welche die Verantwortung für diese wichtigen Dinge tragen muss.

---

<sup>41</sup> vgl.(Finnlay, 2014)

<sup>42</sup> vgl.(Vasarhelyi M, 2015)

## Data Science

Der Begriff Data Science wird typischerweise im Kontext von Big Data verwendet. Data Science, ist die Abkürzung von "Data-intensive Science", dt. Datenintensive Wissenschaft, und bezeichnet die Durchführung von Datenanalysen als empirische Wissenschaft, die direkt aus den Daten selbst lernt.<sup>43</sup> Eine zentrale Aufgabe der Data Science ist es, Muster in Daten zu erkennen und daraus Wissen zu generieren. Dazu werden insbesondere Methoden und Algorithmen des maschinellen Lernens verwendet. Durch diese Algorithmen wird ein Modell, mit Hilfe von historischen Daten eingelernt. Das Modell soll dazu in der Lage sein die gegebenen Daten zu erklären. Dieses Modell kann folgend dazu verwendet werden, um neue Daten prädiktiv (vorhersagend) einzuordnen oder auch Daten deskriptiv (beschreibend) zu beschreiben.

## Data Mining

Unter Data Mining versteht man den Prozess der Wissensgenerierung aus Daten und dessen Darstellung und Anwendung. Die verwendeten Methoden kommen häufig aus der Statistik. Dabei sollen Muster und Zusammenhänge in bestehenden Daten erkannt werden. Die Verfahren des maschinellen Lernens können als ein Teilgebiet des Data Mining interpretieren werden und somit Data Mining als Oberbegriff für sämtliche datengestützten Analyse und Prognoseverfahren betrachtet werden.<sup>44</sup>

---

<sup>43</sup> vgl.(Chang, 2019)

<sup>44</sup> vgl.(V. Brühl, 2019)

## Bedeutung der Datenqualität im BI in Bezug auf Big Data

Durch die Masse der Daten, welche mit dem Begriff Big Data einher gehen, stellt dieses eine Vielzahl neuer Herausforderungen an Business Intelligence Systeme, basierend auf der neuen Vielzahl an Datenquellen. Die Datenwissenschaftler, welche sich mit BI befassen, nutzen die Daten zusammen mit fortgeschrittenen Statistiken und prädiktiven Analysen, um Muster aufzudecken und um zukünftige Muster vorherzusagen. Business Intelligence nutzt dann diese Modelle und Algorithmen, welche die Ergebnisse in eine verständliche Sprache aufschlüsseln und so Unternehmen dabei helfen, die richtigen geschäftsbezogenen Entscheidungen zu treffen, die auf den gesamten gesammelten Daten basieren.<sup>45</sup>

Sensoren beispielsweise, welche immer mehr Verarbeitung mit Lokation, Verhalten der Konsumenten oder Güter und vielem mehr zu tun haben, bringen eine wesentlich höhere Komplexität mit sich.<sup>46</sup>

## Beispiel Predictive Maintenance oder Predictive Analytics

Predictive Analytics ist eine fortschrittliche Analyseverfahren, die sowohl historische als auch neue Daten zur Vorhersage von Verhalten und Trends verwendet. Die Vorhersagen werden auf Grundlage von Mustern, in den historischen Daten getroffen. Dafür werden Methoden des Data Mining, statistische Analysetechniken und automatisierte Machine Learning Algorithmen auf den Datensätzen angewendet, um Vorhersagemodelle zu erstellen. Diese Vorhersagemodelle berechnen schließlich die Wahrscheinlichkeit, ob ein bestimmtes Ereignis eintritt.<sup>47</sup> Die Predictive Analytics kann dabei für Realisierung einer Predictive Maintenance verwendet werden. Die Vorhersage bezieht sich hier auf den Wartungszustand der betrachteten Objekte. Es soll eine Prognose für die Zukunft erstellt werden, um Wartungen planbar zu machen. Die Grundlage für eine Predictive Maintenance sind instandhaltungsrelevante Daten, welches unter anderem Sensordaten und Wartungsprotokolle (time-to-failure-Daten) umfasst.<sup>48</sup>

---

<sup>45</sup> vgl.(247.ai, 2022)

<sup>46</sup> vgl.(Detlef Apel, 2015)

<sup>47</sup> vgl.(V Brühl, 2019)

<sup>48</sup> vgl.(Wuttke, 2022)

## Business Intelligence Systeme

Unter Business Intelligence werden Methoden, Prozesse und Werkzeuge verstanden, um Unternehmensdaten zu transformieren und zu veredeln, um diese zielgerichtet für Analysen und Reporting zur Verfügung zu stellen. Abbildung 8 gibt einen Überblick über den Informationsfluss der BI-Systeme.

Die Datenermittlung, die früher nur den Top-Analytik Experten vorbehalten war, kann jetzt von jedem mit Hilfe von Business Intelligence-Tools durchgeführt werden. Darüber hinaus liefern diese Tools die Informationen, die benötigt werden, um das Wachstum

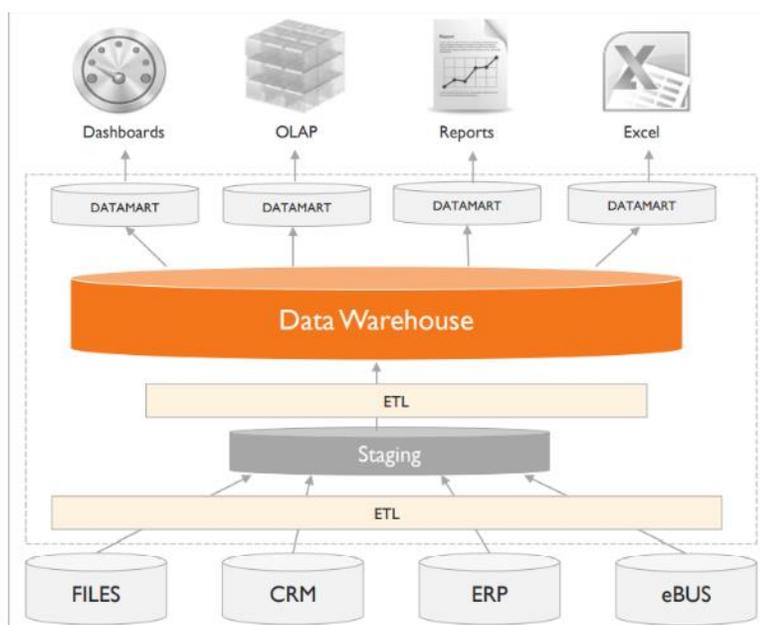


Abbildung 8 - (Enderlein, 2015)

voranzutreiben, dringende Probleme zu lösen, alle Daten an einem Ort zusammenzufassen, zukünftige Leistungen vorherzusagen und vieles mehr.<sup>49</sup>

Was oft nicht erwähnt, aber an anderer Stelle bestätigt wird, ist die Tatsache, dass die meisten Unternehmen ihren primären Analytics- und Business Intelligence-Anbieter betrügen. Einige Unternehmen nutzen 5 oder mehr davon gleichzeitig. Die Medaille hat jedoch zwei Seiten. Auf der einen Seite ist diese Entwicklung verständlich und notwendig. Die Benutzer (und Unternehmen) haben festgestellt, dass kein einziges Tool alle ihre Anforderungen erfüllen kann. Die andere Seite der Medaille ist das Chaos und Verlust des Überblickes über alle Systeme und deren Verwendungszweck.<sup>50</sup>

<sup>49</sup> vgl.(Haije, 2022)

<sup>50</sup> vgl. (Boyer, 2022)

## Experteninterview über die Grenzen der Quellsysteme und Stärken eines Professionellen BI

Das Experteninterview wurde auf Basis praktischer Erfahrungen, eines beratenden Wirtschaftsingenieurs und Business Coach durchgeführt und soll einen Einblick in reale Umstände geben. Dies soll die Grenzen der ERP und CRM Systeme aufzeigen und die realen Voraussetzungen für BI-Systeme identifizieren.

ERP und CRM bieten ebenfalls Auswertungen an, welche ähnlich sind wie auch im BI- Tool, so der Experte. An die Grenzen kämen diese Systeme, sobald der logisch geschlossene Kreislauf der jeweiligen Systeme durchbrochen würde, denn diese müssen miteinander logisch miteinander kommunizieren. Jedoch sei ein einheitliches Reporting nicht möglich und führe zu separaten, statt zu übergreifenden Optimierungen.

### Beispiel des Experten

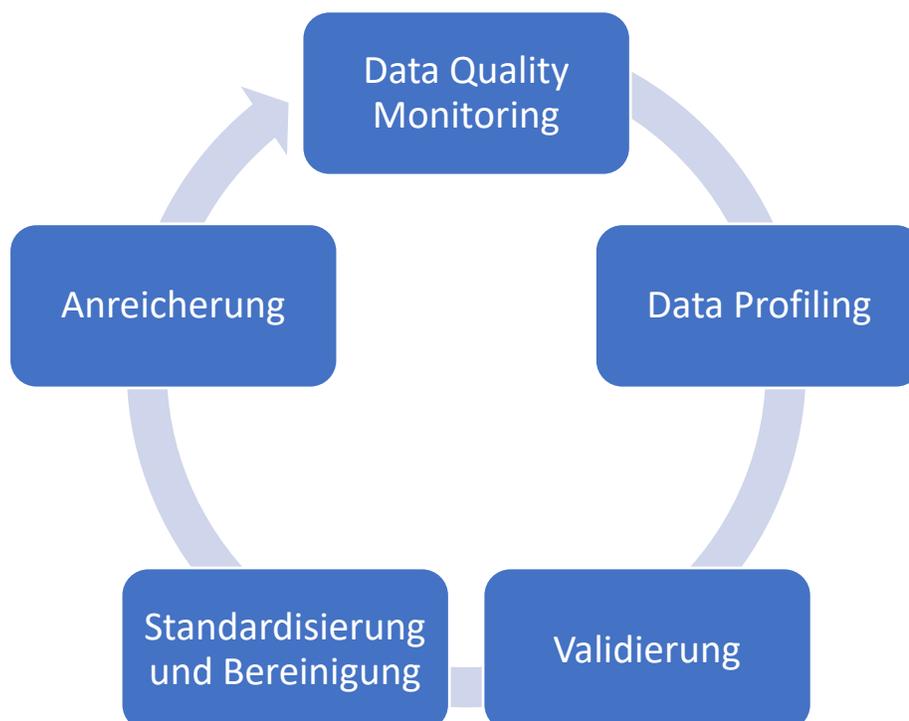
Sobald die folgenden Voraussetzungen erfüllt seien, sei es sinnvoll auf ein professionelles BI zu setzen, um das verschiedene Programm – Chaos vermeiden zu können. Voraussetzungen hierzu sind: Richtige Stücklisten, Materialangaben, Arbeitsplan mit Arbeitsstelle und Bearbeitungszeiten, funktionierende Betriebsdatenerfassung, funktionierender zeitlicher Projektaufwand sowie die entsprechende Kostenstellenstruktur. Eine weitere Voraussetzung sei eine einheitliche Lösung von Artikelnummer zu Produktnummer inklusive Infos aus dem Servicebereich.

## Data Warehouse

Ein Data Warehouse ist ein Datenverwaltungssystem, das zur Erleichterung und Unterstützung von Business Intelligence (BI)-Aktivitäten, insbesondere von Analysen, eingesetzt wird. Data-Warehouses dienen der Abfrage und Analyse und enthalten in der Regel große Mengen an historischen Daten. Eine zugrunde liegende Datenbasis bildet eine wesentliche Komponente auf der Grundlage der verschiedenen Auswertungssysteme. Die Daten innerhalb eines Data Warehouse Systems beziehen sich auf die Kernbereiche der Organisation. Eine wesentliche Eigenschaft ist ein konsistenter Datenbestand, welcher durch die Übernahme der Daten entsteht. Dem Management ist es besonders wichtig diese Daten zu besitzen, um die Entwicklung des Unternehmens über einen bestimmten Zeitraum zu untersuchen. Die Daten in einem Data Warehouse werden in der Regel nur einmal geladen und danach nicht mehr geändert. In aller Regel erfolgt ein Datenzugriff im Allgemeinen nur lesend. Durch einen ETL Prozess (extract transform load) werden die Daten aus verschiedenen Vorsystemen sowie externe Datenquellen in das Data Warehouse transportiert.<sup>51</sup>

## Datenqualitätsverbesserung als Prozess

Abbildung 9 – eigene Abbildung, vgl. (Enderlein, 2015)



<sup>51</sup> vgl.(Hahne, 2014)

## Verbesserung der DQ in Quellsystemen

„Falsche Stammdaten sind doch richtig, nämlich richtig teuer“<sup>52</sup>(Florian, 2021)

Die Grundlage eines jeden Geschäftsprozesses sind die Stammdaten. So gibt es in jedem Unternehmen Stammdaten, jedoch ist nicht jedem Bewusst wie wichtig ein funktionierendes Stammdaten-Management (auch MDM = Master Data Management genannt) ist.

Eine vorgenommene Datenqualitätsverbesserung beschreibt einen iterativen Prozess, wie in Abbildung 9 ersichtlich.

### Beispiele Möglicher Quellsysteme

- ERP
- CRM
- C-MES
- PDM
- SCADA



Abbildung 10 - (cmes, 2022)

Abbildung 10 zeigt wie die Quellsysteme aufeinander aufbauen und soll die Wichtigkeit der Datenqualität in diesem Zusammenhang vermitteln.

Nachfolgend ist eine ausführliche Beschreibung der Quellsysteme und deren Anwendung.

---

<sup>52</sup> (Florian, 2021)

## ERP

Enterprise-Ressource-Planning-Systeme sind vollständig, integrierte Plattformen, die entweder vor Ort oder in der Cloud installiert sind und alle Aspekte eines Produktions- oder Vertriebsunternehmens verwalten. Darüber hinaus unterstützen ERP-Systeme alle Aspekte des Finanzmanagements, des Personalwesens, des Lieferkettenmanagements und der Fertigung mit ihrer zentralen Buchhaltungsfunktion.

ERP-Systeme bieten auch Transparenz in Ihrem gesamten Geschäftsprozess, indem sie alle Aspekte der Produktion, Logistik und Finanzen verfolgen. Diese integrierten Systeme fungieren als zentrale Drehscheibe eines Unternehmens für durchgängige Arbeitsabläufe und Daten, auf die eine Vielzahl von Abteilungen zugreifen können.

ERP-Systeme und -Software haben mehrere wichtige Funktionen in großen, in mittelständischen und kleinen Unternehmen. So stellt sie sicher, dass für die Herstellung der Produkte und Komponenten, alle erforderlichen Materialien an der richtigen Stelle, zur richtigen Zeit und in der richtigen Menge zur Verfügung stehen.

Eingangsdaten sind hierbei beispielsweise Personaldaten (Urlaubstage, etc.) oder bei Büros die benötigten Drucker.<sup>53</sup>

## C-MES (Collaborative manufacturing execution system)

Das C-MES-System wird in der Regel in der digitalen Werkstatt eingesetzt, welche das tägliche Produktionsmanagementsystem der digitalen Fabrik ist, und es trifft zeitnahe Entscheidungen, auf der Grundlage von Echtzeitdaten. Es hat sich zu einem intelligenten System für intelligentes Produktionsmanagement und adaptives Management entwickelt. Dabei dient es zur Überwachung einer Produktionsanlage, in dem verschiedenste Systeme im C-MES vereint, in Echtzeit dargestellt werden können. So können Probleme schnell an der richtigen Stelle ermittelt werden. Als Eingangsdaten werden hierbei Produktionssysteme verwendet, welche sich auf der Maschinen Ebene befinden.<sup>54</sup>

---

<sup>53</sup> vgl.(Oracle, 2022b)

<sup>54</sup> vgl.(Kytka, 2022)

## CRM

Customer Relationship Management (CRM) ist eine Technologie zur Verwaltung aller Beziehungen und Interaktionen eines Unternehmens mit Kunden und potenziellen Kunden. Das Ziel ist die Verbesserung der Geschäftsbeziehungen, um das Unternehmen zu vergrößern. Ein CRM-System hilft dem Unternehmen, mit seinen Kunden in Verbindung zu bleiben, Prozesse zu rationalisieren und die Rentabilität zu verbessern.

Eine CRM-Lösung hilft Ihnen, sich auf die Beziehungen im Unternehmen zu einzelnen Personen zu konzentrieren, z. B. zu Kunden, Dienstleistungsnutzern, Kollegen oder Lieferanten, und zwar über den gesamten Lebenszyklus hinweg, d. h. von der Suche nach neuen Kunden über die Gewinnung von Kunden bis hin zur Bereitstellung von Support und zusätzlichen Dienstleistungen während der gesamten Geschäftsbeziehung.

Eingangsdaten sind hierbei alle Informationen, welche den Kunden betreffen wie etwa Vertragsdaten sowie die Kundenzufriedenheit.<sup>55</sup>

## PDM (Product Data Management)

Ein PDM-System ist also eine Technologie zur Verwaltung der relevanten Produktionsdaten. Es bietet einen Überblick über alle Daten, die während des Produktlebenszyklus und des Herstellungsprozesses erfasst werden.<sup>56</sup> Als Teil des Koordinations- und Informationssystems des Unternehmens legt es die Methoden und Regeln für die Verwaltung der Produktdaten fest. Produktdatenmanagement unterstützt von der Entstehung bis hinein in den Vertrieb eines Produktes. Daher besteht insbesondere für kleine und mittelständische Unternehmen eine sehr große Wertschöpfung im Lebenszyklus der Produkte.<sup>57</sup>

---

<sup>55</sup> vgl.(salesforce, 2022)

<sup>56</sup> vgl.(Collaboration, 2022a)

<sup>57</sup> vgl.(Solutions, 2019)

## SCADA (Supervisory control and data acquisition)

SCADA-Systeme bestehen in der Regel aus Software und Hardware, z. B. speicherprogrammierbaren Steuerungen (SPS) und Remote Terminal Units (RTU). Die Daten werden zunächst gesammelt, um anschließend mit der Produktionsausrüstung des Werks, einschließlich der Produktionsmaschinen und Sensoren, zu kommunizieren. Die gesammelten Daten werden an untergeordnete Ebenen, z. B. Kontrollräume, weitergeleitet, wo die Bediener die SPS und RTUs über eine Mensch-Maschine-Schnittstelle (HMI) steuern können. Die Mensch-Maschine-Schnittstelle (HMI) ist ein wichtiger Teil des SCADA-Systems, da sie den Bedienern die Interaktion mit dem SCADA-System ermöglicht. So können Unternehmen ihre Prozesse wie Pumpen-Steuerungen und Sensoren, lokal oder aus der Ferne bedienen und überwachen. Die Eingangsdaten für das SCADA-System sind aus der Produktionsebene, sowie die Daten aus den einzelnen Maschinen und Sensoren.<sup>58</sup>

## Übersicht der Systeme auf „niedrigerer“ Ebene

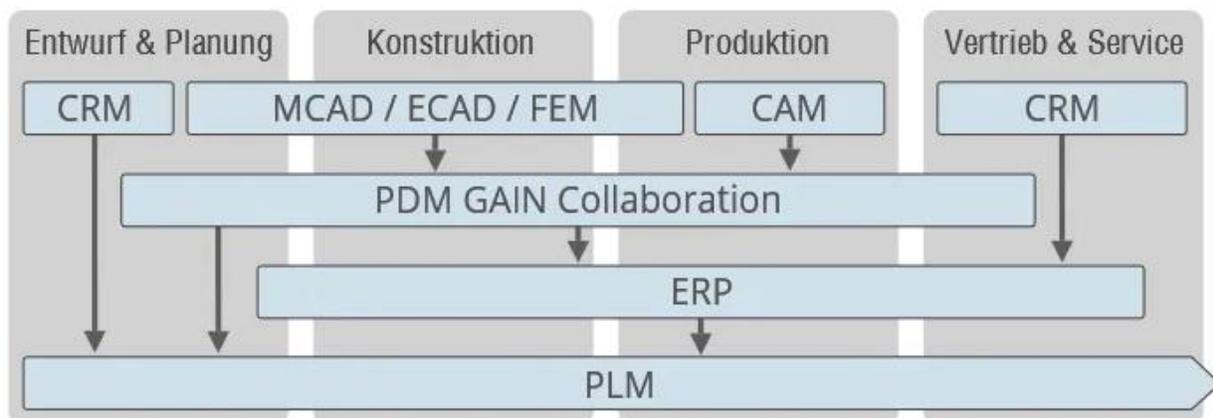


Abbildung 11 - (collaboration, 2022b)

<sup>58</sup> vgl.(Copdata, 2022)

## Datenqualitätsverbesserung in QS

Trotz dieser erfolgreichen und kostengünstigen Methode, Datenqualitätsprobleme direkt in den Quellsystemen zu vermeiden, verwenden nur wenige Unternehmen diese. Bei einer frühzeitigen Fehlervermeidung profitieren alle Systeme, welche auf die Daten der Quellsysteme zugreifen. Wurden bereits vor einer Datenqualitäts- Verbesserung in den Quellsystemen, fehlerhafte Daten in die BI- Anwendung geladen, müssen diese bereinigt werden. Daraus ergeben sich zwei Schritte, um Datenqualität verbessern zu können:

- Fehlerhafte Daten korrigieren
- Neue Datenqualitätsprobleme vorbeugen

Um nicht ständig Daten korrigieren zu müssen, muss der Fehler direkt am Ort des Entstehens behoben werden. Ein Vorgehen hierfür ist:

1. Datenqualitätsprobleme **identifizieren** und unter einer Kosten-Nutzen-Betrachtung **priorisieren**
2. Messbare Zielwerte **definieren**
3. **Gründe** für Qualitätsprobleme **untersuchen** und die verursachenden Prozesse eindeutig identifizieren
4. Betroffene **Prozesse optimieren**, sodass Qualitätsprobleme bereinigt sind
5. Prozessänderung mit anfänglichen **Zielwerten gegenüberstellen**

Sollten die anfänglich definierten Zielwerte nicht erreicht worden sein, gilt es die Schritte zu wiederholen, bis dies der Fall ist. Durch wiederholte Analysen können weitere Fehlerursachen und Prozesse identifiziert werden.<sup>59</sup>

Anschließend werden die einzelnen Schritte nochmals speziell, um den genauen Ablauf zu erklären, erläutert.

---

<sup>59</sup> vgl.(Detlef Apel, 2015)j

### Identifizieren/Priorisieren

Nach einer Auswertung der Fehlerdatenbank, können Fehler priorisiert werden, da aus Ressourcengründen meist keine Behebung aller Fehler auf einmal möglich ist. Als Grundlage für eine Priorisierung soll hier eine Kosten-Nutzen Betrachtung dienen, denn allein die Fehleranzahl entscheidet nicht über die verursachten Kosten. Ergebnis dieses Schrittes, ist eine Liste der zu behebbenden Qualitätsproblemen, inklusive zugehöriger Informationen.<sup>60</sup>

### Zielwerte definieren

Um feststellen zu können ob die durchgeführten Maßnahmen am Ende Wirkung gezeigt haben, werden Kennzahlen und Messpunkte definiert. So werden zu Beginn aktuelle Werte der Kennzahlen notiert und Zielwerte festgelegt, welche mit der Verbesserung erreicht werden sollen. Ergebnis dieses Schrittes, ist eine Liste mit Kennzahlen und zugehörigen Messpunkten, sowie die Zielwerte, welche es zu erreichen gilt.<sup>61</sup>

### Gründe untersuchen

So bald feststeht welche Qualitätsprobleme behoben werden müssen und welche Datenquellen betroffen sind, werden die Gründe für die Probleme untersucht. Am häufigsten entstehen Datenfehler im Dateneingabeprozess, jedoch verursachen System-Upgrades, Prozessautomatisierungen, Datenmigrationen und fehlerhaften Datenkorrekturen ebenfalls Fehler.<sup>62</sup> Workshops sind eine gute Möglichkeit Gründe für die Datenfehler herauszufinden und zu beseitigen. Ergebnis dieses Schrittes ist eine Übersicht der gefundenen Datenqualitätsproblemen, sowie der betroffenen Prozessschritte.<sup>63</sup>

---

<sup>60</sup> vgl.(Detlef Apel, 2015)k

<sup>61</sup> vgl.(Detlef Apel, 2015)l

<sup>62</sup> vgl.(Detlef Apel, 2015)m

<sup>63</sup> vgl.(Detlef Apel, 2015)n

### Prozesse optimieren

Sind Ursachen erst einmal festgestellt, können die Qualitätsprobleme durch geplantes Vorgehen beseitigt werden. In einem Team sollte versucht werden, die beschlossenen Änderungen in einem Projektplan festzuhalten und anschließend umzusetzen. Jedoch ist darauf zu achten, dass die definierten Zielwerte nicht übererfüllt werden, da sonst erhebliche Mehrkosten entstehen. Sind im Prozess nicht nur Maschinen, sondern auch Mitarbeiter involviert, sollten Schulungen und Trainings mit eingeplant werden, um ein größeres Bewusstsein dafür zu schaffen, wie sie selbst die Datenqualität verbessern können und welche Auswirkungen Datenfehler haben. Ergebnis sollte hierbei sein, dass die beschlossenen Änderungen abgeschlossen sind.<sup>64</sup>

### Zielwerte gegenüberstellen

Nach einem Abschluss aller Verbesserungsprozesse, wird durch ein Team überprüft, inwieweit die Zielwerte erreicht wurden. Der Verbesserungsprozess ist beendet, sobald alle Zielvorgaben erreicht sind. Sollte dies nicht der Fall sein, so ist der Prozess ab Schritt 2 (messbare Zielwerte definieren) zu wiederholen. Ergebnis ist eine Zusammenfassung der Erreichung der verschiedenen Kennzahlen, um den Projekterfolg zu beurteilen.

---

<sup>64</sup> vgl.(Detlef Apel, 2015)o

## Data Profiling

Data Profiling ist der Prozess der Untersuchung von Daten, die in einer bestehenden Datenquelle verfügbar sind, des Weiteren ermöglicht das Profiling eine Sammlung von Statistiken und Informationen über diese Daten. Data Profiling umfasst eine breite Palette von Methoden zur Untersuchung von Datensätzen und zur Erstellung von Metadaten. Zu den einfacheren Ergebnissen gehören Statistiken, wie zum Beispiel die Anzahl der Nullwerte, und eindeutige Werte in einer Spalte, ihr Datentyp oder die häufigsten Muster ihrer Werte. Metadaten, die schwieriger zu berechnen sind, betreffen in der Regel mehrere Spalten, wie z. B. Einschlussabhängigkeiten oder funktionale Abhängigkeiten.<sup>65</sup>

Data Profiling ist ein iterativer Prozess, der in Abbildung 12 veranschaulicht wurde.

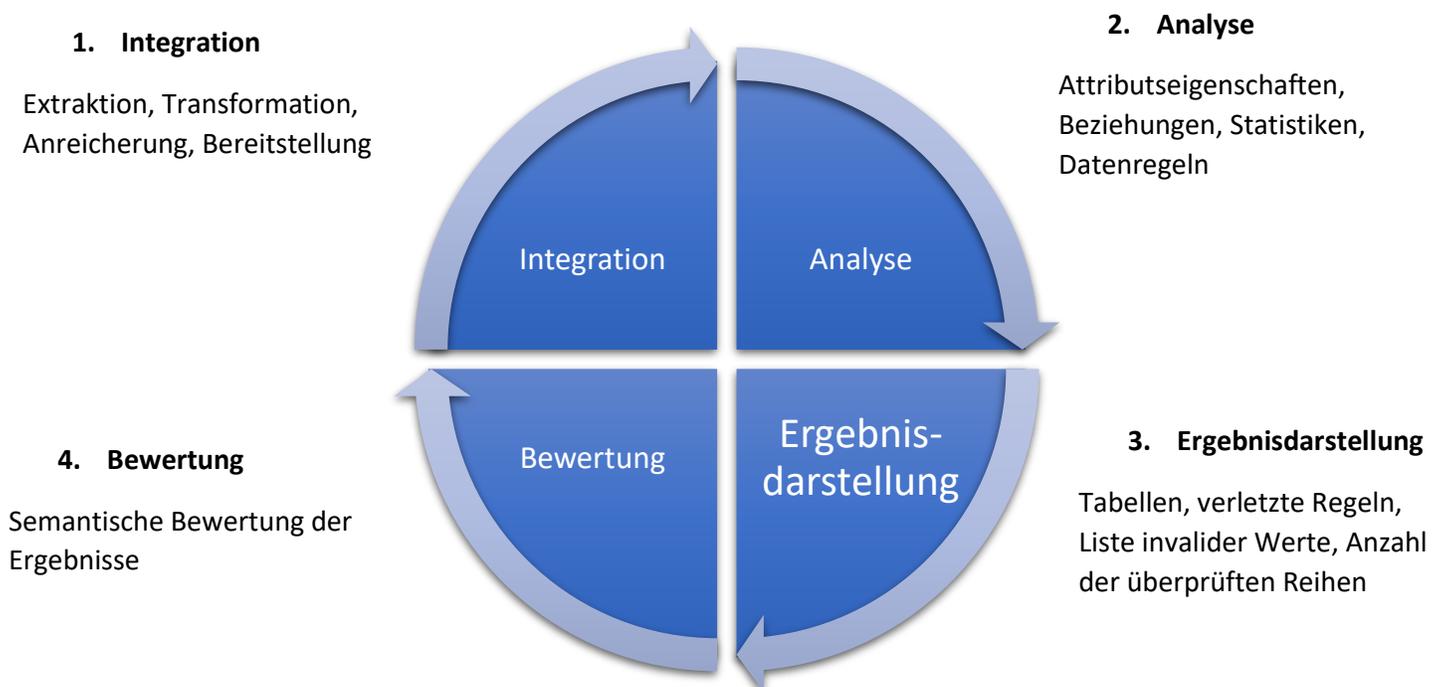


Abbildung 12 – eigene Abbildung – vgl.(Detlef Apel, 2015)

<sup>65</sup> vgl.(Naumann, 2014)

## Beispiel für Data Profiling

Spalten	% von Null-Werten	% eindeutiger Werte	Minimum	Maximum
Artikel Nr.	0.9%	67.2%	10000	222220
Bestell Nr.	7.3%	99.0%	1	999

Es lässt sich beispielhaft eindeutig erkennen, dass die Anzahl leerer Bestellnummern in einigen Datensätzen (7.3%) eindeutig eine maßgebliche Fehlerquelle darstellt.

Einige Artikelnummern sind mehrfach vorhanden ( $100\% - 67.2\% = 32.8\%$ ). Damit ist die Artikelnummer in dieser Qualität kein Mittel, um eine Eindeutige Identifizierung des Artikels durchführen zu können.

## Durchführung eines DQ-Monitorings

Das sogenannte DQ-Assesment beschreibt die Durchführung eines DQ-Monitorings entweder initial oder durch eine Impact Analyse bereits aufgetretener Datenqualitätsprobleme und Bestimmung ihrer Ursachen. Generell gilt nach wie vor, je früher ein DQ-Problem erkannt werden kann, desto besser ist dies für eventuelle weitere auftretende Folgen.<sup>66</sup>

Es gibt 2 Möglichkeiten Datenqualitätsprobleme zu analysieren. Entweder werden Daten vorab analysiert, um Datenqualitätsprobleme aufzudecken, bevor diese in Anwendungen wirksam werden, oder es wird eine Analyse bereits aufgetretener DQ-Probleme durchgeführt, um die Ursache der Probleme nachvollziehen zu können.<sup>67</sup>

Um ein DQ-Monitoring durchführen zu können bietet die folgende Auflistung beispielhaft welche Werkzeuge zu Verfügung stehen könnten:

- SQL-Abfragen
- Analyse durch Kennzahlen
- DQ-Monitoring durch DQ-Werkzeuge
- BI-Frontend-Werkzeuge

---

<sup>66</sup> vgl.(English, 1999a)

<sup>67</sup> vgl.(Olson, 2008)

Der DQ-Monitoring Prozess kann durch die Six Sigma Vorgehensweise stabil gestaltet werden indem wie folgt vorgegangen wird:<sup>68</sup>

- Define
  - Definition der Ziele, des Umfangs und der Ressourcen
- Measure
  - Messung der DQ anhand festgelegter Merkmale in Stichproben
- Improve
  - Priorisierung der gefundenen Verbesserungsmaßnahmen
- Control
  - DQ-Niveau stabilisieren, indem definierte Maßnahmen kontrolliert werden

Um Daten aktiv überwachen zu können, gibt es verschiedenste Methoden, wie etwa durch das initiale Data Profiling und das laufende Data Profiling, wie nachfolgend beschrieben.

### Initiales Data Profiling

Ein initiales Data Profiling ermöglicht ein frühzeitiges Erkennen von Datenqualitätsproblemen, dies führt zu einer Kostensenkung. Eine initiales Data Profiling wird vor allem im Zusammenhang mit BI-Systemen angewendet.

### Laufendes Data Profiling

Die Kontrolle der Datenqualität ist ein maßgeblicher Punkt, um diese in die richtigen Wege lenken zu können. Bei einem laufendem Data Profiling werden laufend Kontrollen periodisch oder anlassgetrieben durchgeführt.

---

<sup>68</sup> vgl.(Schmidt, 2011)

## Experteninterview

Ein Experteninterview dient dem Zweck, ebenso wie eine Umfrage, als eine Methode für die Bachelorarbeit, eine Hypothese zu überprüfen und hilft dabei die Forschungsfrage zu beantworten. Ein Experteninterview eignet sich in diesem Fall bei dieser Forschungsfrage besonders, da ein sehr spezifischer Bereich bzw. ein spezielles Berufsfeld untersucht wird, wie z.B. bei einem Anwender von BI Systemen welcher Entscheidungsträger und Geschäftsführer in einem mittelständischen Unternehmen ist. So werden in diesem Kapitel die Erfahrungen des Experten erläutert, beginnend mit den Erfahrungen und zum Abschluss werden wichtige Erkenntnisse des Interviews zusammengefasst. Während des Interviews entstand ein Transkript, welches im Anhang zu finden ist und auf dieser Basis sind die nachfolgenden Erfahrungen zusammengefasst worden. Inhaltsleere Textpassagen werden hier bereits ausgelassen.

### Vorgehensweise Organisation des Experteninterviews

Um Experteninterviews zu organisieren, wurden hauptsächlich E-Mails und persönliche Kontakte gepflegt. Bei der Firma BMW in München, bei welcher ich bedingt durch eine Werkstudententätigkeit die Möglichkeit hatte, Kontakte innerhalb des Konzerns zu knüpfen, war kein Experteninterview möglich. Die Personen, welche mit einem BI-Tool in Verbindung stehen, sind viel beschäftigt und waren in diesem Zuge nicht bereit ein Interview zu führen, da die gesendete Anfrage meinerseits (in der Abteilung mit dem Namen „FC-250“) unbeantwortet blieb. In höheren Etagen (wie „FC“ – Konzerncontrolling), muss immer erst die persönliche Assistenz kontaktiert werden, um schließlich die Möglichkeit zu haben, ein Gespräch mit einem Teamleiter zu führen.

Nachdem dieses Vorhaben gescheitert war, hatte ich glücklicherweise durch den Kontakt mit Herrn Martini eine Möglichkeit, ein Telefongespräch mit der Geschäftsleitung der Firma Kupil zu führen, welches nachfolgend dokumentiert ist.

### Vorbereitung und Vorgehensweise für das Experteninterview

Durch den zunächst theoretischen Teil der vorliegenden Bachelorarbeit wird eine Grundlage und Wissensbasis geschaffen, um auf einem guten Wissensstand zu sein, bevor ein Experteninterview für eine Bachelorarbeit formuliert wird. Ein Kontakt mit der Interviewten Person konnte durch Herrn Martini hergestellt werden und wird im Folgenden genauer beleuchtet. Dem Experten wurden wie nachfolgend ersichtlich Fragen mit situationsbedingten Nachfragen gestellt.

### Hintergrund des Experten

Gelernter Diplomkaufmann welcher inzwischen bei der Firma Kupil mit Standort in Ehingen als Geschäftsführer tätig ist. Als Standortübergreifendes Tool wird ein BI System verwendet, welches sich auf Cloud basierten Daten stützt. Dies ermöglicht eine Echtzeit Datenabfrage wie beispielsweise der Auftragslage.

## Erfahrung des Experten

Eine der wichtigsten Erkenntnisse, welche der Experte zu Anfang beschreibt, ist die Tatsache, dass er sich auf der Anwenderseite bewege, deshalb mit dem IT-Technischen Hintergrund nur wenig Berührungspunkte habe. Jedoch hat ein Überblick über das Unternehmen gezeigt, dass doch viele Bereiche in das BI-Tool miteinfließen. Wie etwa aus der Vertriebs- und der technischen Abteilung. „Dort kann ich sehen, wie hoch ist der Auftragseingang, von welchem Verkäufer, wie hat er sich entwickelt gegenüber dem Vorjahr.“ Dies ermöglicht eine genaue Analyse des Marktes aber zeigt auch die Entwicklung des Unternehmens. Auch aus Sicht der Projektleitung, sei ein ähnliches Vorgehen möglich. „Das gleiche gilt auch für Projektleiter, wo ich sehen kann welcher Projektleiter, hat wie viele Projekte abgewickelt.“

Ein Quellsystem, welches der Experte beschreibt, ist die Branchensoftware „Klaes“. „Das ist ein relativ veraltetes System, aber es ist das Einzige, was es gerade aktuell auf dem Markt gibt.“ Klaes ist eine Software für alle Prozesse, welche in einem Fensterbauunternehmen ablaufen. Im Bereich Vertrieb, Auftragsbearbeitung, Planung und Beschaffung, Produktion, Versand, Rechnungswesen, Leitung und Verwaltung ist dies ein Werkzeug dafür.<sup>69</sup> In der Praxis funktionieren das indem „Einzelne Aufträge bekommen einen Status danach müssen die Kollegen im Vertrieb und in der Projektierung, diese den Status entsprechend anpassen bzw. verändern. Diese Informationen werden dann in das Monitor Programm eingepflegt

Datenqualitätsprobleme waren ebenfalls ein Thema, wie der Experte übermittelte. „Daten, Qualitätsprobleme die gab es vor allem am Anfang vor allem“. Gelöst wurde das intern, durch Mitarbeiterschulungen und Bewusstseinschaffung gegenüber solchen Datenqualitätsprobleme. Ein Beispiel hierfür sei, „Ist ein Angebot draußen beim Kunden und der sagt ja, ich bestelle, wird der Status verändert, oder wenn der Kunde sagt, nee, ich bestelle nicht wird das Angebot verworfen. Je besser es gepflegt ist, desto besser sind die Daten. [...] Gute Daten kommen aus einer guten Pflege.“ Datenqualitätsprobleme, welche sich im BI durch strategische Fehlentscheidungen ausgewirkt haben, habe es noch nicht gegeben. Jedoch könne es schon mal sein, dass ein Mitarbeiter seinen Status falsch pflegt. „Sollte dies zu einem kritischen Punkt für den Mitarbeiter kommen, wird das persönliche Gespräch gesucht“, so der Experte.

---

<sup>69</sup> (Klaes, 2022)

Nutzung des BI Tools erfolgt in einer Form, um festzustellen wann Preise erhöht werden können. „[...] wann kann ich Preise erhöhen? Steigt die Auftragslage unverhältnismäßig stark an müssen wir Preise erhöhen, oder halt den Auftragseingang stoppen. Wenn der Auftragseingang zurückgehen würde, muss man die Preispolitik anschauen oder verschiedene Maßnahmen ergreifen“, so der Experte.

Das BI Programm wird Standortübergreifend, in Ehingen (Hauptstandort), Ravensburg und Stuttgart durch einen Cloud Service verbunden benutzt. Auch Excel wird als Tool benutzt, um die aktuelle Auftragslage einschätzen zu können, es ist oft schwer sie zu „tracken“ ob beispielsweise die Fenster im Hof stehen oder bereits auf der Baustelle sind, oder sie als halbfertige Produkte zu überwachen. Eine andere Art von Business Tool wird über Addison durchgeführt, mit welchem hauptsächlich betriebswirtschaftliche Auswertungen gemacht werden können, dieses bietet die Möglichkeit Umsätze und Kostenstrukturen zu überwachen. Es wird bei der Firma Kupil einmal im Monat durch Addison die Entwicklung des Unternehmens betrachtet. Das Quellsystem hierbei ist die Buchhaltung.

Ein maßgeblicher Punkt, welcher der Experte angesprochen hatte, stellt die Kalkulation der Materialkosten dar. Diese sind aufgrund aktueller Einflüsse nur schwer kalkulierbar. „[...] da tap ich momentan ziemlich im Dunkeln, da exorbitante Preisveränderungen stattfinden.“ Ein weiteres Problem, welches zum aktuellen Zeitpunkt viele betrifft, sind die Angebote für Kunden zu erstellen, welche meist 3 Monate in der Zukunft liegen, und somit die Materialkosten, aber auch die Kalkulation nicht genau bestimmt werden kann.

## Resümee des Experteninterviews

In diesem Kapitel wird eine inhaltliche Zusammenfassung des Experteninterviews durchgeführt. Praxisbezogene Beispiele, welche sich aus dem Interview ergeben haben, sind ein maßgeblicher Punkt in der Beantwortung der Forschungsfrage und somit zielführend.

Um Kundenentwicklungen beurteilen und auswerten zu können, ist ein Business Intelligence Tool unerlässlich. Es ist nachweislich, wie aus dem Interview herausgehend ebenso wichtig, für die Betrachtung auf die Entwicklung des Unternehmens und die Analyse des Marktes. Die Mitarbeiterperformance kann durch eine Übersicht der Projekte, welche dieser abgeschlossen hat, ermittelt werden. Projekte sind in diesem Zusammenhang große Aufträge wie Fenster, welche für einen Neubau benötigt werden. Die Mitarbeiterperformance gibt ein Verständnis, welchen Wert eine einzelne Person für ein Unternehmen hat. Diese Information kann nützlich sein, um die Arbeitsqualität zu erhöhen.<sup>70</sup>

Als weiteren Punkt beschreibt der Experte die Quellsysteme, welche dem BI System zugrunde liegen. In seiner Fensterbaufirma wird ein Programm mit dem Namen „Klaes“ verwendet. Der Input für dieses Programm entsteht aus folgenden Bereichen: Vertrieb, Auftragsbearbeitung, Planung und Beschaffung, Produktion, Versand, Rechnungswesen und Leitung und Verwaltung.<sup>71</sup>

Im Bereich der Datenqualität entstanden ebenfalls Probleme, wie der Experte beschrieb. Sobald der Auftragsstatus nicht gut genug gepflegt ist, entstehen Datenqualitätsprobleme. Dieses Datenqualitätsproblem wurde vorbildlich, wie auch in der vorliegenden Bachelorarbeit ersichtlich durch eine Schulung der Mitarbeiter in Form eines Workshops bereinigt. Bei schwerwiegenderen Problemen im Bereich der Mitarbeiter, wird auf ein Persönliches Gespräch, zwischen Geschäftsleitung und Mitarbeiter Wert gelegt, was jedoch

---

<sup>70</sup> vgl.(Duvigneau, 2022)

<sup>71</sup> vgl.(Klaes, 2022)

meist nur in mittelständischen Unternehmen in der Form durchgeführt wird. In großen Unternehmen wird meist nur der direkte Vorgesetzte für solch ein Gespräch herangezogen.

Das eingesetzte BI-System wird dafür genutzt, um Preise steuern zu können und zusätzlich unterstützt durch die Excel Tabellenkalkulations- Software von Microsoft.

Der wichtigste Punkt, der aus diesem Gespräch herausging, ist die Glaubwürdigkeit eines BI-Tools, vor allem in dieser aktuellen wirtschaftlichen Lage, in der Materialkosten unverhältnismäßig stark schwanken und sich so nur schwer eine Kostenstruktur bilden lässt. Dazu in der Kritischen Betrachtung der Bachelorarbeit die weitere Ausführung.<sup>72</sup>

---

<sup>72</sup> vgl.(Bundesamt, 2021)

# Entwicklung eines Leitfadens für Unternehmen

## Entwicklung des Leitfadens

Der Nachfolgende Leitfaden wurde auf Basis der theoretischen Grundlage, sowie des Experteninterviews selbst entwickelt und bietet einen Ansatz für Unternehmen, welche Datenqualitätsprobleme in den verschiedensten Bereichen beheben wollen.

## Grundlegendes

Auch im heutigen Wettbewerb ist eine konstante Verbesserung der Qualität der Daten unabkömmlich. Beispielsweise erhöht ein Customer Service die Chance lauernde Unzufriedenheit der Kunden aufzudecken und in langfristigen Beziehungen umzuwandeln, um somit zu verhindern, dass treue und zuverlässige Kunden zur Konkurrenz abwandern.<sup>73</sup>

Bei der Steuerung des Unternehmens müssen sich Mitarbeiter auf optimale Informationen verlassen haben, denn diese wirkt sich direkt auf die Qualität der Entscheidungen im Unternehmen aus. Daten mit größtmöglicher Aussagekraft in diesem Bereich können einen signifikanten Einfluss auf die langfristige und Strategische Ausrichtung eines Unternehmens haben. Durch wiederholte

Analysen können weitere Fehlerursachen und in den laufenden Prozessen identifiziert werden. Eine Datenqualitätsverbesserung ist immer ein iterativer Prozess.

## Ziel

Der Leitfaden soll bei der Bereinigung von Datenqualitätsproblemen unterstützen. Daher richtet sich dieser Leitfaden an Unternehmen, welche Systeme benutzen, in denen sich negative Auswirkungen der unzureichenden Datenqualität gezeigt haben. Der Leitfaden wurde auf Grundlage der theoretischen Forschung im Bereich Datenqualität im BI und das damit verbundene lukrative Kosten - Nutzen-Verhältnis entwickelt.

Durch festgelegte Schritte kann dieser Leitfaden nicht nur für Business Intelligence Systeme Anwendung finden, sondern er kann auch bei anderen IT-Systemen eingesetzt werden, um eine erfolgsversprechende Verbesserung zu ermöglichen.

Nachfolgend werden die einzelnen Schritte genauer erläutert, welche Sie als Geschäftsführer, aber auch Berater mit ihrer Software durchlaufen sollten, um

---

<sup>73</sup> vgl. (Girardi, 2022)

Valide Daten für ein lukratives Kosten-Nutzen Verhältnis zu erzielen.

#### Schritt 1: Identifizierung

Daten untersuchen Mittels Data Profiling (Untersuchung der bestehenden Datenquellen und der damit verbundenen Sammlung von Statistiken und Informationen über diese Daten). Daten Bereitstellen, Analysieren, Ergebnisse Darstellen und danach Bewerten. Nach Auswertung der Fehlerdatenbank, Fehler priorisieren unter Betrachtung der Kosten und Nutzen.<sup>74</sup>

**Ergebnis** ist eine Liste der zu behebenden Qualitätsprobleme, inklusive zugehöriger Informationen.

#### Schritt 2: Zielwerte definieren

Messpunkte definieren, um festzustellen ob die durchgeführten Maßnahmen Wirkung gezeigt haben.

**Ergebnis** in diesem Schritt ist, eine Liste mit Kennzahlen und den dazugehörigen Messpunkten, sowie den Zielwerten, welche es zu erreichen gilt.

#### Schritt 3: Ursachen untersuchen

Nach Feststellung welche Qualitätsprobleme behoben werden müssen und welche Datenquellen betroffen sind, können Fehler behoben

werden. Meist entstehen Datenfehler schon im Dateneingabeprozess, bei den Systemupgrades, der Prozessautomatisierungen, den Datenmigrationen und der fehlerhaften Datenkorrekturen. Daher sollten diese Punkte aus Effizienzgründen und um zusätzliche Kosten zu vermeiden vor allem am Anfang betrachtet werden. Workshops können hierbei helfen, Datenfehler herauszufinden und zu beseitigen.

**Ergebnis** dieses Schrittes zeigt eine Übersicht der gefundenen Datenqualitätsproblemen, sowie der betroffenen Prozessschritte.

#### Schritt 4: Prozesse optimieren

Nach Ursachenfeststellung können Qualitätsprobleme durch geplantes Vorgehen beseitigt werden. Sind im Prozess nicht nur Maschinen sondern auch Mitarbeiter involviert, sollten Schulungen und Trainings mit eingeplant werden, um ein Bewusstsein dafür zu schaffen, wie sie selbst die Datenqualität verbessern können und welche Auswirkungen Datenfehler haben.

**Ergebnis** bei diesem Schritt liegt darin, motivierte Mitarbeiter zu erhalten und Fehlerquellen schnell zu analysieren.

---

<sup>74</sup> vgl. Detlef Apel, 2015p

### Schritt 5: Zielwerte gegenüberstellen

Nach Abschluss aller Maßnahmen, wird vom Team überprüft ob die Zielwerte erreicht wurden. Sollte dies nicht der Fall sein, sollte ab Schritt 2 der Prozess wiederholt werden.

**Ergebnis** ist eine ausführliche Zusammenfassung, sowie die Prüfung ob die verschiedenen Kennzahlen erreicht wurden, um den Projekterfolg zu beurteilen.

### Motivation für gute Datenqualität

Um die Motivation der Mitarbeiter sowie die Anreize steigern zu können, ist extrinsische Motivation eine Methode nach der Self-Determination-Theory (Deci, 2000) um die Datenqualität im Unternehmen, aber auch das Unternehmensklima selbst zu steigern. Es geht hierbei um das Verhalten, dass durch bestimmte Anreize wie Gutscheine, oder Geld gesteigert werden soll.

### Praxisnahes Beispiel für eine Datenqualitätsverbesserung

Die betroffenen Quellsysteme müssen zu Anfang, um einen Istzustand zu erzeugen, eindeutig **identifiziert werden**. Dies geschieht in der Regel durch Interviews oder Workshops. Vorab sollten jedoch Fragebögen ausgefüllt werden, sodass in

Gesprächen bestehende Unklarheiten beseitigt werden können. Bei der Suche nach den Datenverantwortlichen geht es darum die fachlichen als auch die technischen Datenverantwortlichen zu identifizieren. Sobald ein Eindruck der Datenqualität in den Quellsystemen entstanden ist, gilt es dies mit Zahlen zu untermauern mittels Data Profiling (vgl. „Beispiel für Data Profiling“ auf Seite 51). Für eine endgültige Empfehlung hinsichtlich der Kosten und Nutzen, gilt es verschiedene Lösungsmöglichkeiten zu erstellen

Im Entwurf eines Soll-Konzepts können **Zielwerte definiert** werden. Wie exemplarisch durch die Aussage, dass 99% der Artikelnummern vollständig gefüllt sein müssen.

Nachdem feststeht, welche Probleme zu beseitigen sind, gilt es auch deren **Gründe** zu **analysieren**.

**Qualitätsprobleme** können durch geplantes Vorgehen **beseitigt** werden, sollten jedoch im Projektplan festgehalten werden. Methoden hierfür sind bspw. Schulungen.

Nach Abschluss aller Maßnahmen sollten die **Zielwerte** durch ein definiertes Team analysiert und gegenübergestellt werden.

## Limitierungen und zukünftiger Forschungsbedarf

Die vorliegende Arbeit ist mit vielfältigen Limitierungen behaftet, die somit Anstöße für weitere Untersuchungen in diesem Bereich geben. So sind während der Recherche für die Bachelorarbeit folgende Punkte hervorgetreten, welche Faktoren für Limitationen beschreiben.

Die Anzahl der befragten Experten stellt hierbei eine Limitation dar. In größeren Konzernen wird ein BI-Tool sowie damit verbundene Bereiche meist von hierarchisch höheren Positionen bedient. Dies hat zur Folge, dass Personen aus der Führungsebene meist terminlich sehr eingespannt sind und somit nicht die Zeit finden ein Experteninterview oder sonstige Formen von Umfragen durchzuführen. Daher ist die Erfolgswahrscheinlichkeit ein Experteninterview oder auch eine Umfrage bei mittelständischen Unternehmen erhöht.

Weiter sollte angemerkt werden, dass der Informationsschutz ein großes Thema ist, wenn es um preisgeben von beispielsweise strategischen Informationen geht. Diese sind im Umfeld von BI Systemen einhergehend, da beispielhaft Formen der Preisbildung hierbei ein wichtiger Faktor sind.

Insbesondere ist hierbei anzumerken, dass eine umfangreiche Analyse, welche Daten in einem Unternehmen in der Praxis in einem BI System vorhanden sein müssen, nur bedingt möglich war. Auf Basis der theoretischen Grundlagen und praktischen Beispielen aus dem Internet sowie die vereinzelt Informationen aus den Experteninterviews, wurde dieser Leitfaden entwickelt. Zudem ist an dieser Stelle anzumerken, dass es wichtig ist den Leitfaden über die Zeit zu verbessern und gegebenenfalls an neue Umstände anzupassen.

## Kritische Betrachtung der Bachelorarbeit

In diesem Abschnitt sollen die Ergebnisse dieser Arbeit als auch die zugrundeliegenden Informationsquellen kritisch reflektiert werden. Beginnend mit der Aussagekraft der Inhalte und bedingt durch die aktuelle Lage. Anschließend wird diskutiert, dass die Glaubwürdigkeit an manche Informationsquellen in Frage gestellt werden muss, da Produkte zur Qualitätssicherstellung zu kaufen sind und diese den Anspruch daran nicht erfüllen.

Die Identifizierung der Datenqualitätsprobleme und die damit verbundenen Symptome wurden anhand von derzeit verfügbaren Literaturquellen und fachlichen Artikeln erarbeitet. Jedoch ist durch die aktuelle Lage der Preis für Güter aber auch Dienstleistungen stark schwankend, bedingt durch den Ukraine Konflikt, die Corona Pandemie sowie andere äußere Einflussfaktoren. Dies hat unvorhersehbare Folgen in den BI-Systemen, da Kostenstrukturen nicht mehr eindeutig kalkulierbar sind. Preisgleitklauseln werden häufig herangezogen, um vor unvorhersehbaren Kostenschwankungen zu schützen, wie beispielsweise für Aufträge, welche einen Kostenvoranschlag bedürfen.<sup>75</sup> Eine Beurteilung dieser Auswirkungen ist nicht möglich, da die Einflussfaktoren für solche Preisschwankungen unvorhersehbar sind und somit auch nicht in dieser Arbeit thematisiert wurden.

Weiterhin sind manche Informationen, aufgrund der Verkaufsstrategien, der Softwareherstellern einschlägig behaftet und wollen den Leser dazu bewegen ihre Qualitätsverbesserungs- Software beispielsweise zu kaufen. Um solche einschlägigen Quellen zu vermeiden, wurden in dieser Arbeit größtenteils Quellen verwendet, bei denen kein Softwareanbieter im Hintergrund steht, um die Informationen der vorliegenden Arbeit neutral wiedergeben zu können.

Zuletzt sollte auch das Ergebnis der Arbeit kritisch betrachtet werden. Der Leitfaden stützt sich mitunter auf der Analyse des Experteninterviews und theoretischen Grundlagen. Aufgrund von verschiedenen Einflussfaktoren, welche jedes Unternehmen individuell macht, ist es nur schwer möglich ein „Allheilmittel“ gegen schlechte Datenqualität in der Kürze des hier vorliegenden Leitfadens zu finden. Um dies vorzubeugen, sollte die gesamte Arbeit hierbei im Fokus liegen, da außerhalb des Leitfadens mehrere Zusammenhänge erklärt wurden.

---

<sup>75</sup> vgl. (Handelsblatt, 2022)

## Gesamtfazit der Arbeit

In der vorliegenden Arbeit wurden die theoretischen Grundlagen in den Gebieten der Qualität der Daten, die mit einem Business-Intelligence System zusammenhängen bearbeitet. Dabei wurde abgeleitet, dass ein einheitlicher Leitfaden für Unternehmen entwickelt werden kann, um Datenqualitätsprobleme in IT-Systemen beheben zu können. Zum einen, anhand der theoretischen Ansätze, basierend auch auf statistischen Werten, wie beispielsweise die häufigsten Fehlerursachen aber auch Auswirkungen. Zum anderen dienten die Experteninterviews als Basis des Leitfadens.

Eine Erkenntnis dieser Arbeit ist die vielfältige Anwendung des entwickelten Leitfadens. Da Unternehmen, basierend auf statistischen Werten, aber auch aus einem Experten Bericht<sup>76</sup>, sehr oft über dieselben Erfahrungen berichten, wurden hierbei Schritte herausgearbeitet, welche als besonders wichtig erachtet wurden.

Die zweite Erkenntnis beschreibt die Antwort auf die Forschungsfrage:

## **Welche Datenqualität wird hinsichtlich des Aufwandes und Nutzens im Rahmen eines zur Unternehmenssteuerung taugliches BI-System benötigt?**

Es müssen qualitativ hochwertige Quelldaten in einem BI-System vorhanden sein, damit durch den höheren Aufwand, der hier entsteht, durch den bestmöglichen Einsatz der verschiedenen Systeme, der größte mögliche Nutzen erzeugt werden kann. Mit der Verbesserung der fundamentalen Datenqualität steigt der Kosten-Nutzen-Faktor im BI ebenfalls an. Ein angepasstes Bewusstsein der Unternehmen bezüglich der korrekten Datenerfassung kann helfen bereits im Ursprung Fehlerquellen auszubessern (siehe Seite 26). Vergleichbar ist die mit einem Beispiel aus dem Hausbau: Wenn das Fundament auf welchem später ein Gebäude gebaut werden soll auf zu schwachem Untergrund oder auch das Fundament selbst mangelhaft ist, wird das Gebäude später nur wenigen äußerlichen Einflüssen standhalten und am Ende, einen wesentlichen hohen wirtschaftlichen Schaden verursachen.

---

<sup>76</sup> vgl. ("Experteninterview," 2022)

Zusammengefasst wurden in dieser Arbeit, alle als besonders wichtig erachteten Punkte zur Beantwortung der Forschungsfrage thematisiert. Angefangen von Überblicken über den Informationsfluss, warum Datenqualität in einem funktionierenden BI-System ein Faktor ist, der Definitionen und über Messung von Qualität sowie den Kosten und Nutzen Betrachtung der Daten bis hin zu Verbesserung der Qualität in den Quellsystemen und der Durchführung der Experteninterviews mit abschließendem Resümee sowie erarbeitetem Leitfaden. Durch diese qualitative Forschung wurde somit gezeigt, dass Daten verbessert werden müssen, um das Kosten-Nutzen-Verhältnis im BI steigern zu können.

## Anhang

### Transkription

Interviewer: Ich rufe an wegen meiner Bachelorarbeit, da hat mich er Herr Martini schon gestern angekündigt.

Experte: Ja, genau. Ich bin gerade erst heimgekommen aber wir können kurz reden, das dauert so ca. eine viertel Stunde hat er mir gesagt, ist das korrekt?

Interviewer: Ja, genau so etwa in der Richtung.

Experte: Ja mal kucken ob ich helfen kann, aber wenn ich es kann, tue ich es gerne.

Interviewer: Alles klar! Super!

Interviewer: Erst mal vorneweg mein Thema für die Bachelorarbeit heißt:

Welche Daten müssen in einem funktionierendem BI-System vorhanden sein damit dieses eine lukratives Kosten/ Nutzenverhältnis erzeugt.

Das heißt im Klartext Datenqualität im BI-System, so kann man das grob zusammenfassen und wie sich da ein Kosten-Nutverhältnis ergibt.

Experte: Ja.

Interviewer: Hier geht es erst einmal, da habe ich die Frage, ob da Sie sich schon länger mit diesem Thema BI beschäftigen bzw. Datenqualität in ihrem Unternehmen.

Experte: Ich bin mehr der Anwender, ich nutze die Auswertung bin aber nicht der Datenfachmann. Das heißt wir bauen ja Kunststofffenster und haben da so ein Branchensoftware Programm und aus diesem Programm lesen wir Daten aus, in ein Monitor, weiß aber nicht wie der heißt und dann kann ich z.B. sehen, wir haben eine Vertriebs-Abteilung und Technische-Abteilung dann kann ich sehen wie hoch ist der Auftragseingang, von welchem Verkäufer und der gesamte Auftragseingang und wie hat der sich entwickelt zum Vorjahr und das gleiche kann ich auch dann von der Projektleitung sehen, welcher Projektleiter hat wieviel Projekte abgewickelt z.B.

Ist das sowas was sie meinen, oder?

Interviewer: Ja das meine ich schon. Das bringt mich dann auch schon zum nächsten Punkt.

Das sind quasi die Stammdaten für das BI-System.

Was wären da die Eingangsdaten für dieses BI-System. Das was sie gerade genannt haben.

Experte: Es kommt aus der Branche Klaes für uns ein relativ veraltetes System, aber das ist das einzige am Markt das es so richtig gibt und aus diesem System werden sozusagen Daten ausgelesen und das funktioniert so, dass praktisch die einzelnen Aufträge, die bekommen einen Status und dann müssen praktisch die Kollegen im Vertrieb und in der Projektierung, die müssen dann den Status verändern. Anhand diesem Status kann ich diese Daten exportieren in dieses Monitorprogramm hinein.

Interviewer: Gab es da aus ihrer Erfahrung schon mal Probleme mit dem Bereich Datenqualität, dass sie sagen das hat hier und da nicht ganz funktioniert, weil irgendwie manche Leute das falsch eingegeben haben oder es von Programm zu Programm falsch übernommen wurde in den Schnittstellen.

Experte: Ja die Qualität bzw. Probleme die gab es am Anfang vor allem und man muss die Leute natürlich dazu anhalten den Status zu pflegen.

Wenn z.B. ein Angebot wo draußen ist beim Kunden sagt der Kunde ich bestelle dann wird der Status verändert oder wenn der Kunde sagt nein ich bestelle nicht dann wird der Status nochmals verändert, dann ist das Angebot verworfen und natürlich je besser es gepflegt ist desto besser sind die Daten. Der Punkt war einfach nachher, dass wir diese Auswertung den Leuten selber zur Verfügung stellen und dadurch die Leute einen Gewinn daraus ziehen, weil sie selber wissen wollen wie ist denn mein Stand und das animiert sie auch dazu die Daten zu pflegen.

Die guten Daten kommen aus der guten Pflege, klar.

Interviewer: Haben sie allgemein so Datenqualität-Tools eingeführt oder nur solche Methoden, bzw. gibt es verschiedene Programme.

Experte: Das kann ich nichts sagen da müssen sie den Kollegen der das managet fragen.

Wenn sie möchten können sie mit ihm nochmal ein Interview machen, aber er ist zur Zeit im Urlaub, aber der kennt sich hier mehr aus der ist auch in der IT-Ausbildung und der weiß wie er es macht. Ich bin nur der Nutzer, ich schau mir die Daten an und ziehe meine Schlüsse daraus.

Interviewer: Hat es schon mal die Auswirkung im BI sich selber gezeigt dass man, oft ist es ja so wenn da Fehler im BI sind werden da oft die Entscheidungen falsch getroffen, ist das auch schon mal der Fall gewesen, dass da strategische Entscheidungen falsch getroffen wurden die auf falschen Daten basiert haben?

Experte: Das ganze System ist relativ neu, ich nutze es seit ungefähr einem halben Jahr.

Selbst wenn die Daten jetzt nicht zu 100% korrekt wären, alle Entscheidungen die man trifft, was könnte das für eine Entscheidung sein, das könnte z.B. eine Entscheidung sein in dem ich einen Auftrag nicht annehme weil ich schon voll ausgelastet bin oder ich sag der Mitarbeiter ist ein Low-Performer weil er seinen Status nicht gepflegt hat, dann zeigts nichts an, dann sinkt sein Umsatz, aber in Wirklichkeit hat er mehr gemacht, wenn ich sowas bzw. wenn eine kritische Entscheidung anstehen würde, dann würde ich mit dem Mitarbeiter selber sprechen und würde nochmal nachhaken und dann kann man auch wenn da Fehler drin waren oder noch wären, dann kommt das an das Tageslicht und ich würde mal sagen falsche Entscheidungen sind da nicht getroffen worden, aber wahrscheinlich auch weil es noch relativ neu ist und wir noch gar nicht so viele Entscheidungen daraus irgendwo treffen, sondern die Entscheidung trifft hier die Datenbasis ist eine Grundlage für Entscheidungen aber es gibt auch andere so ne Mischung.

Interviewer: Das BI Programm heißt wie nochmal?

Experte: Das weiß ich leider auch nicht wie das heißt. Ich kann vielleicht sagen das ist glaub schon so ein Tool was am allgemeinen im Markt erhältlich gibt das hat man da einfach aufgepflanzt, wir importieren die Daten von der ERP-Software, die importieren wir da rein.

Ich bin gerade nicht in der Firma bin gerade privat sonst könnte ich schnell nachschauen. Dashboard heißt das Programm.

Das haben wir ins Teams integriert und ich war da am Anfang, ich hab da schon gemerkt als darauf umgestellt worden ist, am Anfang war ich ein bisschen skeptisch aber jetzt nutze ich das wirklich sehr viel, um in auch z.B. wann kann ich Preise erhöhen wenn die Märkte Auftragseingang steigen um zu überlegen steigt unverhältnismäßig stark an, dann kannst du sagen OK wir müssen Preise anheben oder auch den Auftragseingang stoppen oder wenn ein Auftragseingang zurück gehen würde stark, da würde man sich die Preispolitik anschauen oder verschiedene Maßnahmen ergreifen. Aber das ist alle noch relativ sehr neu so eine extrem Situation hatten wir noch gar nicht wo das System extrem ausgeschlagen hat.

Interviewer: Das BI-Tool wird das speziell nur lokal an ihrem Standort benutzt oder Standortübergreifend.

Experte: Wir haben 3 Standorte und ja das ist eine Cloud in Teams ist das drin und da gibt's verschiedene Zugänge, nicht jeder kann alles rauslesen, da sind die Zugriffe, ja ich hab glaub alle Zugriffe die es gibt, es gibt andere Kollegen die haben nur, z.B. wo nur ihre eigenen Daten oder Abteilungsleiter von ihrer Niederlassung oder so.

Interviewer: Sie benutzen ein eigenes BI-System, hab ich das richtig verstanden. Ich hab schon gehört es benutzen mehrere gleichzeitig?

Experte: Als ich kann unter BI-System noch gar nicht soviel vorstellen, ich weiß nicht was umfasst das?

Interviewer: Es gibt BI-System im Grunde das Dashboard das sie schon beschrieben haben, das ist die Oberfläche wo man die Auswertungen sieht mit Verkaufszahlen und Umsatzzahlen welche von den Stammdaten gespeist werden, also von Produktion, Absatz, Produktionsdaten usw.

Experte: Ja, also was ich auch noch nutze, das noch so alte Geschichten z.B. was bei uns immer wichtig ist, dass wir die Bestände ermitteln, den Auftragsbestand, das ist gar nicht so leicht, denn so ein Fensterauftrag kann sein ist gerade in der Produktion oder er ist eingebaut, oder halb eingebaut oder die Fenster stehen fertig im Hof, oder sie sind schon auf der Baustelle, aber es fehlen noch die Vorbaurolläden, oder alles ist nur so halb fertig und das ist wichtig wenn ich meine Leistung, wenn ich die Leistung des gesamten Unternehmens bewerten möchte, dann muss ich wissen wie hoch sind meine Umsatzzahlen, das sind relativ harte Zahlen, die kommen aus der Buchhaltung aber die halbfertigen Aufträge die muss ich sozusagen beurteilen und auch einschätzen und das mach ich über Excel, da importiere ich Daten aus dem Programm raus in das Excel rein, ich sag nur so wie man es früher gemacht hat, so hat man das bei den anderen Sachen früher auch gemacht hat so importiert man es ins Excel rein, dann gibt's die kleinen Aufträge werden pauschal bewertet und große Aufträge spreche ich mit den Projektleitern einzeln durch, immer im Quartal, um rauszufinden wie weit sind diese fortgeschritten, dann habe ich am Schluss eine Zahl meiner Auftragsbestände, die beeinflussen natürlich das Ergebnis maßgeblich, diese Bestandszahl und das läuft über Excel das ist auch ein Werkzeug sozusagen welches wir verwenden und dann haben wir eine betriebswirtschaftliche Auswertung, die machen wir über Edison das ist auch eine Art Business-Tool, da sehe ich dann meinen Umsatz, meine ganzen Kostenstrukturen und das ist ein anderes System das kommt aus bzw. speist sich aus der Buchhaltung.

Die Buchhaltung machen wir mit Edison und dann krieg ich einmal, ich könnte das jeden Monat machen aber ich pflege diese Daten, diese Bestandsbewertung auch die Materiallagerbestände die pflege ich einmal im Quartal ein, dann habe ich einmal im Quartal sehe ich dann die Entwicklung unter diesem Punkt des Unternehmens.

Was das Ergebnis angeht natürlich die Auftragseingangsanzahlen sehe ich sofort täglich also minutengenau, wenn der Auftrag angelegt wird erscheint er auch da, es gibt wirklich Sachen die laufen „just in time“ , es gibt andere Sachen die habe ich öfters im Quartal die sind aufwendiger zu ermitteln auch die Daten.

Interviewer: Allgemein kann man sagen Entscheidungen werden nicht nur anhand diesem Dashboard das es gibt getroffen auch dass andere Daten damit einfließen.

Experte: Es gibt verschiedene Entscheidungen, es gibt ja Entscheidungen, hab ich ein Problem mit einem Mitarbeiter oder hab ich ein Problem mit dem Preis, oder habe ich Probleme in einem bestimmten Gebiet und solche Entscheidungen muss ich zeitnah treffen, aber die ganzen Kostenstrukturen und wie hoch ist mein Materialkostenanteil, ist ja momentan Thema, die Preis spielen verrückt die Einkaufspreise, die Materialpreise und wie wirkt sich das auf das ganze Unternehmen aus, das sind jetzt Zahlen die bekommt man nicht schnell, da tappe ich momentan so

ziemlich im Dunkeln weil exorbitante Preisveränderungen sind und wie sich das auswirkt weiß ich erst sozusagen erst im Nachhinein und das ist ein Problem.

Interviewer: Also im Nachhinein nach einem Quartal, oder?

Experte: Nee nicht einmal, denn so ein Auftrag der läuft, ein kleiner Auftrag läuft so in 2-3 Monaten durch, es gibt auch welche die ziehen sich oft über ein Jahr von Auftragserteilung bis nachher bis ein Auftragsobjekt abgeschlossen ist, das dauert und du merkst erst im Nachhinein, ich musst heute ein Angebot abgeben und dann werden die Fenster erst in 3 Monaten produziert und dann in 5 Monaten eingebaut, ich weiß gar nicht wie die Kostensituation bis dahin ist, also das sind die Probleme, mit denen gerade viele kämpfen, dass die Kalkulation einfach sehr schwierig ist, ich weiß nicht ob das noch mit reinspielen soll.

Interviewer: Gut, das hat mir tatsächlich schon sehr weiter geholfen gerade als, das deckt sich also auch teilweise mit meiner Bachelorarbeit die ich schon angefangen hatte, da ist die Frage darf ich die einzelnen Infos aus diesem Gespräch in meiner Bachelorarbeit verwenden?

Experte: Ja, klar war ja nichts indiskretes

Interviewer: Deshalb habe ich es ja relativ so gehalten so dass es nicht kritisch sein könnte, oder sowas, super alles klar, gut, dann war es das schon von meiner Seite, haben sie noch Fragen an mich?

Experte: Ich wünsche ihnen viel Erfolg dabei

Interviewer: Danke, Dankeschön, vielen Dank für ihre Zeit.

Experte: Ja sehr gerne.

Interviewer: Einen schönen Tag noch.

## Literaturverzeichnis

- 247.ai (Producer). (2022, August 17). 247.ai. Retrieved from <https://www.247.ai/insights/business-intelligence-and-big-data#:~:text=Big%20data%20involves%20storing%2C%20processing,extract%20meaningful%20information%20and%20insights.>
- Andreas Krämer, T. B. (2022). Kundenwertzentriertes Management *Value-to-Value-Pricing-Big Data-Controlling* (pp. 67-71): Springer Gabler.
- Bank, D. (1999, 3). Patzer bei Euro Umstellung. *Computerwoche*, 47.
- Boyer, J. (Producer). (2022, Juli 08). Motio. Retrieved from <https://motio.com/de/why-multiple-bi-tools-matter/>
- Brühl, V. (2019). Big Data, Data Mining, Machine Learning und Predictive Analytics: Ein konzeptioneller Überblick (pp. 4). Frankfurt a.M.: Goethe University Frankfurt.
- Brühl, V. (2019). *Big Data, Data Mining, Machine Learning und Predictive Analytics: Ein konzeptioneller Überblick. Big Data , Data Mining, Machine Learning und Predictive Analytics.* Retrieved from Frankfurt a. M.:
- Bundesamt, S. (Producer). (2021, Juli 5). destatis.de. Retrieved from [https://www.destatis.de/DE/Presse/Pressemitteilungen/2021/07/PD21\\_N044\\_61.html#:~:text=Wie%20das%20Statistische%20Bundesamt%20\(Destatis,Bauholz%20um%2038%2C4%20%25.](https://www.destatis.de/DE/Presse/Pressemitteilungen/2021/07/PD21_N044_61.html#:~:text=Wie%20das%20Statistische%20Bundesamt%20(Destatis,Bauholz%20um%2038%2C4%20%25.)
- Chang, W. L., & Grady, N. (2019). Data Interoperability Framework. Volume 1, Definitions Versin 3. *NIST Special Publication*, 6.
- cmes (Producer). (2022, August 22). c-mes. Retrieved from <https://c-mes.co.uk/>
- Collaboration, G. (Producer). (2022a, August 22). gain. Retrieved from <https://www.gain.de/pdm-system/pdm-plm/>
- collaboration, G. (Producer). (2022b, August 22). Gain collaboration. Retrieved from <https://www.gain.de/pdm-system/pdm-plm/>
- Copdata (Producer). (2022, August 22). copdata. Retrieved from <https://www.copadata.com/de/produkt/zenon-software-plattform-fuer-industrie-energieautomatisierung/visualisierung-steuerung/was-ist-scada/>
- DataSelf (Producer). (2022, August 27). DataSelf Informed Decisions. Retrieved from <https://www.dataself.com/six-ways-business-intelligence-changes-customer-service/>
- Deci, R. M. R. a. E. L. (2000). *Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being* Retrieved from Rochester
- Detlef Apel, W. B. (2015). *Datenqualität erfolgreich steuern*. Heidelberg: dpunkt.
- Duvigneau, A. (Producer). (2022, August 10). BEEKEEPER. Retrieved from <https://www.beekeeper.io/de/blog/mitarbeiterengagement-messen/>
- Eckerson, W. W. (2002). *Data Quality and the bottom line*.
- Economist, T. (2012). *The Deciding Factor: Big Data & Decision Making*.
- Ellsworth, M. (Producer). (2019, Juli 09). wiser. Retrieved from <https://blog.wiser.com/de/what-are-the-causes-and-costs-of-bad-data-in-retail/>
- Enderlein, M. (2015). Infomotion GmbH. Retrieved from <https://www.infomotion.de/unternehmen/>
- English, L. (1999a). *Improving Data Warehouse and Business Information Quality - Methods for Reducing Costs and increasing Profits*. New York.
- English, L. (1999b). *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*. New York.
- Eppler M, H. M. (2004). *A Framework for the Classification of Data Quality Costs and Analysis of their Progression*. Paper presented at the Proceedings of 9th International Conference on Information Quality, USA.
- Experteninterview. (2022). Neu-Ulm, Bayern, Deutschland.
- Finnlay, S. (2014). *Predictive analytics, data mining and big data*. Hampshire: Palgrave Macmillan.

- Florian (Producer). (2021, November 30). fact.ai. Retrieved from <https://fact.ai/blog/stammdaten/>
- Förg, M. (Producer). (2019, 12 19). bigdata-insider.de. Retrieved from <https://www.bigdata-insider.de/symptome-schlechter-daten-tipps-fuer-mehr-datenqualitaet-a-887794/>
- Friedman, T. (2010). *Findings from Primary Research Study - Data Quality Issues Create Significant Cost, Yet often to got Unmeasured*. Retrieved from
- Gartner. (2011). *Gartner Special Report Examins How to LEverage Pattern.Based Strategy to Gain Value in Big Data*. Stamford.
- Girardi, J. (Producer). (2022, August 22). dataself. Retrieved from <https://www.dataself.com/six-ways-business-intelligence-changes-customer-service/>
- Hahne, M. (2014). Modellierung von Business-Intelligence Systemen *Leitfaden für erfolgreiche Projekte auf Basis feldxible Data-Warehouse Architekturen* (pp. 1-3). Heidelberg: dpunkt.verlag.
- Haije, E. G. (Producer). (2022, April 06). mopinion. Retrieved from <https://mopinion.com/de/die-15-wichtigsten-business-intelligence-tools/>
- Handelsblatt, R. (Producer). (2022, August 21). Handelsblatt. Retrieved from [https://www.handelsblatt.com/unternehmen/mittelstand/strategie\\_und\\_finanzierung/schwankende-rohstoffpreise-preisklauseln-schuetzen-vor-steigenden-kosten-seite-2/3225392-2.html](https://www.handelsblatt.com/unternehmen/mittelstand/strategie_und_finanzierung/schwankende-rohstoffpreise-preisklauseln-schuetzen-vor-steigenden-kosten-seite-2/3225392-2.html)
- Hannig, D. (Producer). (2022, August 27). Honestly. Retrieved from <https://www.honestly.de/blog/interne-kommunikation/>
- Huifen Wang, L. L. (2014). *A Collaborative Manufacturing Execution System*. Paper presented at the International Conference on Cooperative Design, Visualization and Engineering, Nanjing, Jiangsu 210094, China.
- ITWissen (Producer). (2022, August 27). ITWissen.info. Retrieved from [https://www.itwissen.info/Programmierbare-logische-Steuerung-programmable-logic-controller-PLC.html#:~:text=Programmierbare%20Logik%2D%20Controller%2C%20Programmable%20Logic,von%20industriellen%20Steuerungssystemen%20\(%20ICS\)](https://www.itwissen.info/Programmierbare-logische-Steuerung-programmable-logic-controller-PLC.html#:~:text=Programmierbare%20Logik%2D%20Controller%2C%20Programmable%20Logic,von%20industriellen%20Steuerungssystemen%20(%20ICS))
- Keizer G, G. (Mai 2004). Poor Data Quality Dooms Many IT Projekts. *Information Week*.
- Klaes (Producer). (2022, 08 15). klaes.de. Retrieved from <https://www.klaes.de/klaes-fensterbausoftware>
- Klein, A. (2009). *Datenqualität in Sensordatenströmen*. Retrieved from Dresden:
- Klein D, T.-G. P., Hartmann M. (2013). Big Data- Informatik Spektrum (pp. 319-323).
- Knut Hildebrand, M. G. (2022). Daten- und Informationsqualität (pp. 88). Frankfurt a. M.: Springer.
- Kollmann. (2020). *Handbuch Digitale Wirtschaft* (pp. 12,1038ff). Wiesbaden: Springer Gabler.
- König C, S. J., Wiegand E. (2018). Big Data: Chancen, Risiken, Entwicklungstendenzen. (pp. 8-11). Wiesbaden: Springer Fachmedien.
- kupil (Producer). (2022, August 22). kupil. Retrieved from <https://www.kupil.de/unternehmensprofil.html>
- Kytka, N. (Producer). (2022, August 22). team-con. Retrieved from <https://www.team-con.de/unsere-leistungen/produkte/manufacturing-execution-system>
- Matthias Förg, N. L. (Producer). (2019, Dezember 19). BigData Insider. Retrieved from <https://www.bigdata-insider.de/symptome-schlechter-daten-tipps-fuer-mehr-datenqualitaet-a-887794/>
- Naumann, F. (2014). Data profiling revisited. *SIGMOD Rec.*, 42(4), 40–49. doi:10.1145/2590989.2590995
- Olson, J. E. (2008). Ordnungsrahmen und Anwendungsbeispiele; Integrierte Informationslogistik *Unternehmensweites Datenqualitätsmanagement* (pp. 211-230). Berlin.
- Oracle (Producer). (2022a, August 17). OCI. Retrieved from <https://www.oracle.com/de/autonomous-database/what-is-data-mart/#:~:text=der%20Daten%20kennen.->

- [,Ein%20Data%20Mart%20ist%20eine%20einfache%20Form%20eines%20Data%20Warehouses,weniger%20Quellen%20als%20Data%20Warehouses.](#)
- Oracle (Producer). (2022b, August 22). oracle. Retrieved from <https://www.oracle.com/erp/what-is-erp/>
- planview (Producer). (2022, August 27). planview. Retrieved from <https://www.planview.com/resources/guide/resource-management-software/resource-management-leverage-people-budgets/>
- ProjectBase (Producer). (2022, August 27). ProjectBase. Retrieved from <https://project-base.org/projektmanagement-glossar/qualitaetskriterien/>
- Pütter, C. (2007). Schlechte Datenqualität kostet Banken die Hälfte ihres IT-Budgets. *CIO*.
- Redman, T. C. (2008). *Data Warehouses and Quality - Not just for IT anymore*: Navesink Consulting Group.
- salesforce (Producer). (2022, August 22). salesforce. Retrieved from <https://www.salesforce.com/de/learning-centre/crm/what-is-crm/>
- Schmidt, T. (2011). BI-Spektrum DMAIC-Zyklus: Systematisches Datenqualitätsmanagement zahlt sich aus (pp. 26-28).
- Solutions, a. (Producer). (2019, Oktober 09). applus-erp. Retrieved from <https://www.applus-erp.de/ressourcen/glossar/was-ist-ein-pdm-system-einfach-erklart/>
- statista (Producer). (2022, Januar 21). de.statista. Retrieved from <https://de.statista.com/statistik/daten/studie/946148/umfrage/herausforderungen-im-bereich-digital-analytics-aus-unternehmenssicht/>
- talend (Producer). (2022, August 27). talend. Retrieved from <https://www.talend.com/de/resources/was-ist-datenqualitaet/>
- Tucci, L. (Producer). (2022, August 27). TechTarget. Retrieved from <https://www.techtarget.com/searchcio/definition/strategic-management>
- Vasarhelyi M, K. A., Tuttle B. (2015). Big Data in Accounting *An Overview*. *Account Horiz* (pp. 381-396).
- Wuttke, L. (Producer). (2022, August 19). datasolut. Retrieved from <https://datasolut.com/was-ist-predictive-analytics/#Anwendungen-f%C3%BCr-Predictive-Analytics>