

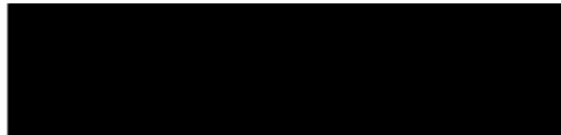
Bachelorarbeit

## **Systematische Untersuchung der Datenqualität und -strukturen in der Supply-Chain**

Zur Erlangung des akademischen Grades  
Bachelor of Science (B.Sc.)

Vorgelegt von:

Büsra Türkmenoglu



Studiengang:

Wirtschaftsingenieurwesen  
193955

Erstgutachter:

Dr. Anne Antonia Scheidler

Zweitgutachter:

M. Sc. Florian Hochkamp

Ort und Datum:

Dortmund, 22.11.2021

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis .....</b>	<b>III</b>
<b>Tabellenverzeichnis .....</b>	<b>IV</b>
<b>Abkürzungsverzeichnis .....</b>	<b>V</b>
<b>1. Einleitung .....</b>	<b>1</b>
<b>2. Datenbegriff .....</b>	<b>3</b>
2.1 Allgemeiner Datenbegriff .....	3
2.1.1 Daten, Information und Wissen .....	5
2.1.2 Datenstrukturen .....	7
2.1.3 Datenqualität .....	10
2.2 Datenbegriff in der Supply Chain .....	20
2.2.1 Supply Chain .....	20
2.2.2 Daten in der Supply Chain .....	22
2.2.3 Datenstrukturen in der Supply Chain .....	25
2.2.4 Datenqualität in der Supply Chain .....	29
<b>3. Datenqualität in der Supply Chain .....</b>	<b>31</b>
3.1 Anforderungen an die Datenqualität .....	31
3.2 Kriterien der Datenqualität .....	45
3.3 Exemplarische Validierung .....	51
3.4 Diskussion und Fazit .....	58
<b>4. Zusammenfassung und Ausblick .....</b>	<b>62</b>
<b>Literaturverzeichnis .....</b>	<b>63</b>
<b>Anhang .....</b>	<b>VI</b>
<b>Eidesstaatliche Versicherung .....</b>	<b>XII</b>

## Abbildungsverzeichnis

Abbildung 1: Wissenspyramide .....	5
Abbildung 2: Abstraktionsebenen von Datenstrukturen .....	7
Abbildung 3: Datenqualitätspyramide.....	11
Abbildung 4: 15 IQ-Dimensionen in 4 IQ-Kategorien .....	16
Abbildung 5: Das organisatorische DQ-Modell mit Fehlerarten und Ansätzen .....	18
Abbildung 6: Beispiel zum relationalen Datenbankmodell.....	26
Abbildung 7: Beziehungen im relationalen und objektorientierten Datenbankmodell	27

# Tabellenverzeichnis

Tabelle 1: Datenqualitätsmerkmale .....	12
Tabelle 2: Qualitätsmerkmale nach Datenwerten.....	12
Tabelle 3: Datenqualitätsmerkmale .....	13
Tabelle 4: Drei Ebenen der Datenqualitätsmerkmale .....	14
Tabelle 5: Ursachen für Datenqualitätsmängel mit Beispielen .....	19
Tabelle 6: Lokationstypen.....	23
Tabelle 7: Arten von Transaktionsdaten .....	25
Tabelle 8: Zentrale Anforderungen an die Datenbanksysteme .....	28
Tabelle 9: Anforderungen an die DQ der Daten .....	31
Tabelle 10: Anforderungen an die DQ der Datenbanken .....	31
Tabelle 11: Kriterien aller Werke .....	33
Tabelle 12: Synonyme innerhalb der DQ-Kriterien .....	37
Tabelle 13: Häufigkeitstabelle gleichbedeutender Kriterien .....	38
Tabelle 14: Mindestkriterien an die Supply Chain .....	49
Tabelle 15: Untersuchte SC-Daten.....	51
Tabelle 16: Lieferantenstammsatz .....	52
Tabelle 17: Stammdaten Gender .....	53
Tabelle 18: Bewegungsdaten Order Date und Ship Date .....	53
Tabelle 19: Genauigkeitsüberprüfung des Postal Codes, State und City Bewegungsdaten Order.....	54
Tabelle 20: Produktinformationen.....	55
Tabelle 21: Redundanzen innerhalb der Datenstruktur.....	55
Tabelle 22: Anwendbarkeit der Kriterien und ihr Einfluss auf die Datenqualität.....	58

## Abkürzungsverzeichnis

ADT	Abstrakter Datentyp
APO	Advanced Planning and Optimization
DQ	Datenqualität
ERM	Entity-Relationship-Modell
IQ	Informationsqualität
PDS	Produktionsdatenstruktur
PPM	Produktionsprozessmodell
SAP	Systemanalyse Programmentwicklung
SC	Supply-Chain
SCM	Supply-Chain-Management

# 1. Einleitung

Im Rahmen des globalen Wettbewerbs werden viele Unternehmen mit Herausforderungen wie dem technologischen Wandel, den sich dauerhaft ändernden Wettbewerbsbedingungen und den vielfältigen Kundenwünschen, konfrontiert. Zunehmend ist ein Wechsel von der Massenproduktion in die kundenindividuelle Produktion bei der Erstellung von Produkten und Dienstleistungen beobachtbar. (Biedermann & Kotzab, 2018) Die dabei entscheidenden Erfolgsfaktoren sind die hohe Flexibilität gegenüber inhomogenen Kundenbedürfnissen, Kosteneffizienz und eine punktgenaue Lieferfähigkeit (Beckmann, 2004). Bei der Erfüllung dieser Kunden- und Marktanforderungen stoßen Unternehmen auf Grenzen, die durch den Zugriff auf externe Ressourcen zu bewältigen sind (Biedermann & Kotzab, 2018). Durch zahlreiche Stufen der Liefer- oder Wertschöpfungskette, der sogenannten Supply-Chain (SC), werden die Unternehmensgrenzen überwunden, in der die Ausgestaltung von der ursprünglichen Kette zur finalen Senke betrachtet wird (Corsten & Gössinger, 2008). Eine dauerhafte Planung, Steuerung und Überwachung der SC verfolgt das Ziel alle Unternehmenspartner wie Zulieferer, Produzenten, Handel und andere Dienstleister effizient zu integrieren und zu koordinieren, um ein erfolgreiches auf den Bedarf des Endkunden abgestimmtes Handeln sicherzustellen (Beckmann, 2004). Für die Planung und Ausführung der SC-Prozesse sind integrierte Informationssysteme notwendig, die neben organisatorischen Funktionen auch die Infrastruktur aufrechterhalten und bei der Entscheidungsfindung unterstützen (Becker & Geimer, 2001). Mittels dieser Informationssysteme findet innerhalb der SC ein Wissensaustausch statt, der Datenbanken mit umfangreichen Daten erfordert. Der dabei erzeugte Informationsfluss ist als eine unerlässliche Voraussetzung anzusehen, der das Gelingen der Planung, Optimierung und Steuerung des Materialflusses innerhalb der SC erst ermöglicht (Hildebrand, 2004). Darunter befinden sich Daten, die zusätzlich zu quantitativen Eigenschaften, wie der Datenmenge, qualitative Eigenschaften, wie die der Datenqualität und -struktur, in SC-Datenbanken aufweisen (Hildebrand, 2004).

Vor diesem Hintergrund ist das Hauptziel dieser Arbeit, die Erstellung eigener Datenqualitätskriterien zur Bewertung der Datenqualität (DQ) in der SC. Im Hinblick darauf ist ein Teilziel ein grundlegendes Verständnis für die Datenstrukturen und DQ in der SC zu schaffen. Auf dieser Grundlage können die nächsten Teilziele, wie das Ableiten der Anforderungen an die DQ in SC-Datenbanken realisiert werden. Mithilfe dieser

Anforderungen erfolgt schließlich die Umwandlung in Kriterien an die DQ. Zur Bestätigung ihrer Umsetzbarkeit wird exemplarisch eine Validierung durchgeführt und die gewonnenen Erkenntnisse werden gegenübergestellt.

Die Arbeit wird in vier Kapitel gegliedert. Im zweiten Kapitel teilt sich der Stand der Technik in zwei Grundlagengebiete auf; Grundlagen der Daten und Grundlagen der Daten im SC-Kontext. In dem ersten Grundlagenteil wird der Begriff Daten erklärt und im nächsten Schritt von den oftmals synonym verwendeten Begrifflichkeiten Information und Wissen abgegrenzt. Weiter werden die zu untersuchenden Datenqualitäten und -strukturen näher erläutert. Da die DQ nicht einheitlich definiert ist, werden einige Definitionsansätze aus der Literatur vorgestellt und es wird auf bekannte Datenqualitätsmängel (DQ-Mängel) eingegangen. Es wird die Bedeutung des Supply-Chains thematisiert und die dafür üblichen Daten vorgestellt. Anschließend werden die Grundlagen der SC aufgegriffen, um die Schnittstelle DQ und Datenstrukturen in der SC darstellen zu können. Im Anschluss hierzu beginnt in Kapitel drei die hauptsächliche Untersuchung der DQ, bei der die Anforderungen an diese abgeleitet werden. Basierend auf diesen Anforderungen findet die Umwandlung in Kriterien an die DQ in der SC statt, die daraufhin exemplarisch überprüft und in der Diskussion näher bewertet werden. Daraufhin wird im Fazit eine Zusammenfassung über die gewonnenen Erkenntnisse gegeben und die Bedeutung dieses Ansatzes für die SC abgewägt. Letztlich beendet Kapitel vier die Arbeit mit einer Zusammenfassung und einem Ausblick über künftige Chancen und Entwicklungen.

## 2. Datenbegriff

In diesem Kapitel findet vor der Spezifikation der Daten im SC-Kontext eine allgemeine Betrachtung der Daten statt. Diesbezüglich wird der allgemeine Datenbegriff vorgestellt, um im nächsten Schritt von den beiden Begrifflichkeiten *Information* und *Wissen* abgegrenzt zu werden. Hiernach wird Bezug zu allgemeinen Datenstrukturen und zur DQ einschließlich der DQ-Mängel genommen und das ganzheitliche Verständnis hinsichtlich der Daten vervollständigt. Daraufhin werden die Daten im SC-Kontext eingeordnet. Hierzu wird zunächst die SC thematisiert, woraufhin sowohl die Daten als auch die Datenstrukturen in der SC betrachtet werden. Zum Ende des Kapitels werden die DQ und gleichzeitig die DQ-Mängel in der SC aufgezeigt.

### 2.1 Allgemeiner Datenbegriff

Im Folgenden wird die Bedeutung der Daten näher erklärt. Gemäß DIN 44300 sind Daten ein Gebilde aus Zeichen oder kontinuierliche Funktionen, die beruhend auf bekannten oder unterstellten Vereinbarungen, Informationen repräsentieren. Der Zweck hierbei sei die Verarbeitung und somit das dazugehörige Ergebnis. Weiter wird in der Norm angemerkt, dass die Verarbeitung der Daten die Durchführung mathematischer, umformender, übertragender und speichernder Operationen umfasst. (Deutsches Institut für Normung, 1988)

In der Literatur befinden sich eine Reihe verschiedener Definitionen von Datenarten (Piro & Gebauer, 2021; Schulte, 2016). Hierbei eignet sich die Arbeit von Piro und Gebauer (2021) besonders, da sich ihre Definition der Datenarten auf die konsistente Kommunikation im Unternehmen bezieht. Demnach lassen sich ihre Beschreibungskriterien in die „Eigenschaften der Daten“ und der „Kontextinformationen“ unterteilen (Piro & Gebauer, 2021). Zu den Eigenschaften der Daten zählen das Format, die Struktur, der Inhalt, die Stabilität, die Verarbeitung sowie das Geschäftsobjekt. Bei den Kontextinformationen sind die Angaben zu den Prozessen und zu den Verwendungszwecken relevant. (Piro & Gebauer, 2021)

Unter der ersten Eigenschaft Format wird verstanden, dass ein Datenfeld als alphanumerisch, numerisch, integer etc. bestimmt ist. Diese Klassifikation nach der Zeichenart findet auch nach Mertens et al. (2017) statt. Bei der zweiten Eigenschaft Struktur erfolgt eine Einteilung der Daten in strukturierte, semistrukturierte und unstrukturierte Daten (Piro & Gebauer, 2021). Während strukturierte Daten Metadaten enthalten,



welche die vorliegenden Daten definieren, werden unstrukturierte Daten erst indirekt durch die Interpretation durch den Informationsempfänger zu strukturierten Daten und somit eindeutig (Piro & Gebauer, 2021).

Eine Schnittstelle bilden semistrukturierte Daten, welche vereinzelt strukturiert vorliegen können, aber in der Gesamtheit betrachtet keine einheitliche eindeutige Struktur aufzeigen. Bei dem nächsten Kriterium, dem Inhalt, werden Daten nach den Inhaltsdaten und den Metadaten unterschieden. Die Inhaltsdaten geben hierbei das Objekt direkt an. Metadaten jedoch stehen für strukturgebende Informationen, deren Aufgabe die Zuordnung der Inhaltsdaten zu einer feststehenden Struktur ist, um diese dann beschreiben zu können. (Piro & Gebauer, 2021)

Das Kriterium Stabilität betrachtet die Zeitdauer, in der die Daten unverändert bleiben und trennt dementsprechend fixe Stammdaten von variablen Bewegungsdaten. Zu den Stammdaten gehören Daten, die sich nach ihrer Eingabe im System selten ändern und aus diesem Grund zentral gespeichert sowie prozessbedingt ausgelesen werden, wie zum Beispiel eine Artikelbezeichnung. Im Gegensatz dazu ändern sich Bewegungsdaten häufig und müssen den Geschäftsumständen entsprechend neu eingegeben oder ermittelt werden, wie zum Beispiel eine Bestellmenge. (Piro & Gebauer, 2021; Schuh & Stich, 2012, Schwemm, 2009)

Zum vorletzten Kriterium, der Verarbeitung, ist die Stellung der Daten bei ihrer Verarbeitung von Bedeutung. Zur ersten Stellung gehören Daten an, die nach ihrer Eingabe in das System, Eingabedaten sind. Verweilen diese Daten weiterhin im System werden sie in der zweiten Stellung, als Speicherdaten bezeichnet. Die dritte Stellung sind Daten, die in einem System verarbeitet werden und dann als Ausgabedaten vorliegen können. (Piro & Gebauer, 2021)

Dies wird auch durch die Arbeit von Mertens et al. (2017) bestätigt, welche die Ein- und Ausgabedaten als die Stellung im Verarbeitungsprozess klassifiziert. Das letzte Kriterium beruft sich darauf, dass Daten gleichzeitig Objekte darstellen und somit dauerhaft sowie eindeutig zum Geschäftsobjekt zugeordnet werden können. Hierbei bringt das Geschäftsobjekt den fachlichen Verwendungszweck mit der inhaltlichen Verantwortung in Verbindung und ermöglicht beispielsweise durch den Zugriff auf mehrere Geschäftsobjekte die Geschäftsabwicklung in einem Gesamtprozess. (Piro & Gebauer, 2021)

Nach den Eigenschaften der Daten ist ein Einblick in die beiden Kriterien, Prozess und Zweck wichtig, da sie den Daten einen Kontext geben (Piro & Gebauer, 2021).

Bei dem ersten Kriterium Prozess, werden die beiden Fragen behandelt, in welchem Prozess das Datum entstanden sei und in welchem die Informationen genutzt werden. Weiter werden den Daten bei dem Kontext-Kriterium, Zweck, unterschiedliche Zwecke innerhalb der Gesamtprozesse gegeben. Innerhalb beider Kriterien sowie bei dem zuvor beschriebenen Kriterium der Dateneigenschaften, Verarbeitung, sind Mehrfachnennungen möglich. (Piro & Gebauer, 2021)

### 2.1.1 Daten, Information und Wissen

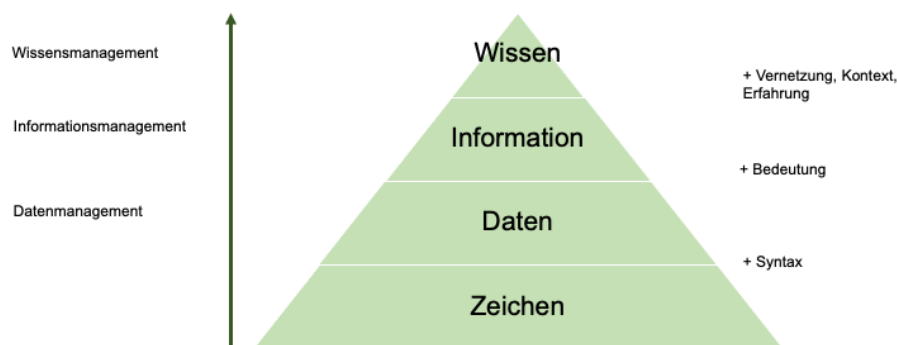


Abbildung 1: Wissenspyramide  
(eigene Darstellung in Anlehnung an Bodendorf, 2006, S.1)

In diesem Abschnitt wird deutlich, dass der Begriff Daten nicht mit den beiden Begriffen Information und Wissen gleichzusetzen ist und eine klare Abgrenzung vorgestellt werden kann. Wie in 2.1 beschrieben, stellen Daten ein Gebilde aus Zeichen dar (Deutsches Institut für Normung, 1988). Zeichen aus einem Zeichenvorrat werden mithilfe von feststehenden Syntaxregeln zu Daten (siehe Abbildung 1). Erst mit der Zuordnung einer Bedeutung (Semantik) werden aus den Daten, Informationen (Bodendorf, 2006). Mithilfe der Semantik verknüpft der Empfänger die Daten, „mit einem Begriff, einer Vorstellung aus der Realität oder theoretischer Art“ (Bodendorf, 2006, S. 1). Die Semantik allein reicht aber nicht aus. Piro und Gebauer (2021) beschreiben Daten als „das inhaltliche Element einer Information und der Kontext ist das beschreibende Element“ (Piro & Gebauer, 2021, S. 145). Der Kontext, in welchem die Daten vorliegen, wird zwingend benötigt, um die Daten zu interpretieren und daraus Informationen zu gewinnen (Piro & Gebauer, 2021). Anhand dieses Interpretationsschemas (Kontext) liest eine Person die Daten in einem Sinnzusammenhang (Köhler & Oswald, 2009). Die

Verknüpfung der gewonnenen Informationen ermöglicht dann die Entstehung des Wissens (Bodendorf, 2006). Hierzu werden Kenntnisse benötigt, um zusammenhängende Beziehungen der Informationen und ihre Vernetzung herzustellen. So definieren Köhler und Oswald (2009) das Wissen als „gelernte Information“, da für den Umgang mit den Informationen ein Lernprozess stattgefunden haben muss. Demnach wird Wissen erarbeitet, in dem gemeinsame Kontexte geschaffen und sich neues Wissen, in Form der „Verinnerlichung neuer Information“, angeeignet wird (Köhler & Oswald, 2009, S. 15). Dabei liegt der Fokus auf den Informationen, die einen Zustand sowie dessen Entstehung und Änderung beschreiben (Köhler & Oswald, 2009). Bei dem letzten Aspekt wird ein Bezug zum Wissen über Zusammenhänge hergestellt. Aus der Literatur wird ersichtlich, dass diese Verknüpfung der Informationen pragmatisch sei (Bodendorf, 2006). Auch Wittmann (1959) definiert Informationen als zweckorientiertes Wissen, da somit der Zweck, Handlungen durchzuführen und Entscheidungen zu treffen, erfüllt wird.

## 2.1.2 Datenstrukturen

Nach dem ein grundlegendes Verständnis über die Daten-Bedeutung und deren Abgrenzung aufgebaut werden konnte, werden in diesem Teil die, für diese Arbeit essenziell wichtigen Datenstrukturen thematisiert. Grundlegend ist relevant, was Datenstrukturen sind, in welchem Zusammenhang die Strukturen mit den Daten stehen und welchen Zweck sie verfolgen. Hierbei ist in der Literatur auffallend, dass keine einheitliche Definition für Datenstrukturen existiert (Ehrig et al., 2001; Schäfer, 1989; Dieker & Güting, 2018). Deshalb sollen zwei Definitionsansätze weiterhelfen und in den nächsten Schritten herangezogen werden. Laut Ansätzen aus der Literatur bestehen Datenstrukturen aus einem oder mehreren Datenbereichen und enthalten außerdem eine Reihe von Operationen auf diesen Datenbereichen (Ehrig et al., 2001). Die Modellierung der Datenbereiche erfolgt durch Mengen und die der Operationen durch Abbildungen (Ehrig et al., 2001). Es können sich Datenstrukturen als beispielsweise unterschiedliche Zahlenbereiche mit Rechenoperationen wie INT (ganze Zahlen) usw. oder als z.B. Zeichenketten (Wörter) darstellen lassen. Zudem sind sie auch als Operationen zur Zeichenverarbeitung STRING (z. B. Wörter) sowie als Graphen abbildbar. Des Weiteren sind Datenstrukturen auch durch Algorithmen definiert. (Ehrig et al., 2001)

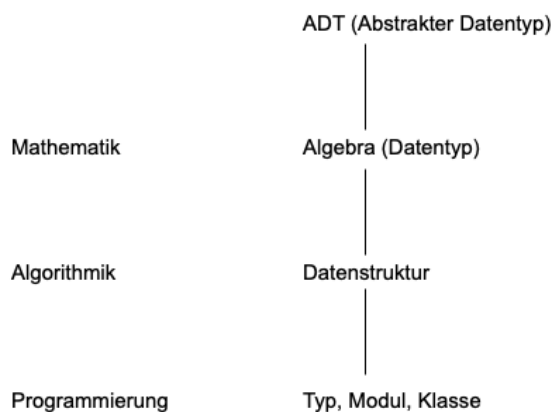


Abbildung 2: Abstraktionsebenen von Datenstrukturen  
(eigene Darstellung in Anlehnung an Dieker & Güting (2018), S. 22)

Dies wird auch durch Dieker und Güting (2018) bestätigt, für sie sind die Algorithmen und Datenstrukturen eng miteinander verbunden und nicht zu trennen. Demnach

bilden die Datenstrukturen für die Algorithmen die Grundlage, auf welcher sie arbeiten können und die Algorithmen sind die Komponenten der Datenstrukturen (Dieker und Güting, 2018). Anzumerken ist, dass sich diese Arbeit nur auf die Datenstrukturen konzentriert und nicht näher auf die Algorithmen eingeht.

Aufgrund der uneinheitlichen Definierbarkeit haben Dieker und Güting (2018) ebenfalls Definitionsansätze aufgestellt, mit dem Bestreben, die Datenstrukturen aus ihrer Sicht zu beschreiben. Der dabei zum Einsatz kommende Erklärungsansatz basiert auf einem Diagramm, das die drei Abstraktionsebenen von Datenstrukturen umfasst. Diese drei Abstraktionsebenen sind die Mathematik, die Algorithmik und die Programmierung (siehe Abbildung 2). Anhand dieses Diagramms leiten sie ab, dass eine „Datenstruktur eine Implementierung einer Algebra oder eines abstrakten Datentyps (ADT) auf algorithmischer Ebene“ ist (Dieker und Güting, 2018, S. 2). Darüber hinaus besteht die Möglichkeit ihrer Implementierung in einer Programmiersprache, da Datenstrukturen sich den Begriffen des Typs, der Klasse oder des Moduls ähneln (siehe Abstraktionsebene drei) (Dieker und Güting, 2018).

Wie zuvor angedeutet, bestätigt die Abstraktionsklasse Mathematik, dass eine Datenstruktur mehrere Klassen von Objekten mit den sich darin befindenden infrage kommenden Werten aufweist, auf denen Operationen stattfinden können (Dieker und Güting, 2018). Die Realisierung dieser Operationen erfolgt durch Algorithmen (Dieker und Güting, 2018).

Es liegen unterschiedliche Arten von Datenstrukturen vor (Schäfer, 1989; Steyer, 2019; Mehta & Sahni, 2018). Einige davon sind lineare Strukturen, zirkulare Strukturen, mehrdimensional indizierte Strukturen, Bäume, Netze und Hash-Strukturen (Schäfer, 1989; Steyer, 2019). Zu den linearen Strukturen gehören beispielsweise Listen, Warteschlangen, Stapel und eindimensional indizierte Strukturen wie Felder (Schäfer, 1989; Mehta & Sahni, 2018). Das Gegenteil bilden die mehrdimensional indizierten Strukturen wie die Indexstrukturen (Schäfer, 1989). Weiter werden der Datenstruktur, Bäume, der allgemeine Baum und der Binärbaum zugeordnet (Schäfer, 1989). Netze hingegen beinhalten allgemein Graphen und Hash-Strukturen (Schäfer, 1989; Steyer, 2019). Diese sind Datenstrukturen, die die Implementierung dynamischer Mengen veranlassen (Steyer, 2019).

Die zuvor genannten Datenstrukturen lassen sich in eine nächsthöhere Art der Datenstrukturen, in dynamische und statische Datenstrukturen, unterteilen (Dieker & Güting, 2018). Datenstrukturen haben allgemein die Aufgabe, die programmierte Organisation

der Daten darzulegen. Bei der Datenorganisation ist die Position des Elements im Speicher sowie die Relation zu anderen Elementen wichtig. (Schäfer, 1989)

Unter statischen Datenstrukturen versteht man die bisher verwendeten Variablen und Felder wie Arrays, Integer, String usw. Nach ihrer Deklaration in einem Programm oder innerhalb einer Funktion ändert sich der Speicherplatz-Zustand nicht mehr und bleibt fest. (Dieker & Güting, 2018)

Im Gegensatz dazu existieren dynamischen Datenstrukturen, die auf Grundlage der Standardtypen und deklarierten Strukturen zur Laufzeit neue Speicherplätze bereitstellen. Neue Elemente der Datenstrukturen werden initialisiert und mit bereits vorhandenen Elementen in Relation gebracht. Für die Speicherplatzfreigabe können auch andere Elemente entfernt werden. (Lange & Stegemann, 1987)

Dies hat zu bedeuten, dass die dynamischen Datenstrukturen durch Entfernen und Einfügen von Elementen, schrumpfen oder wachsen werden. Zu den dynamischen Datenstrukturen zählen sowohl lineare Datenstrukturen wie Felder, Listen und Warteschlangen etc. als auch nicht-lineare Datenstrukturen wie Bäume und Graphen. (Dieker & Güting, 2018; Saake & Sattler, 2021)

### 2.1.3 Datenqualität

Für das Datenqualitätsverständnis sind, aufgrund der hohen Komplexität des Begriffs Qualität, nähere Einblicke in diesen notwendig. Sowohl die Begriffsbeschreibung als auch die daraus resultierenden für diese Arbeit relevanten Qualitätssichten werden hierfür offengelegt. Allgemein beschreiben DIN- oder ISO-Normen Qualität als „die Gesamtheit von Eigenschaften und Merkmalen eines Produkts oder einer Dienstleistung, die sich auf deren Eignung zur Erfüllung festgelegter oder vorausgesetzter Erfordernisse beziehen“ (Deutsches Institut für Normung e. V., 1995, S. 212). Eine andere aber ähnliche Definition findet sich in dem „Handbuch Qualität“ von Geiger und Kotte (2008) wieder, in der die Qualität als „realisierte Beschaffenheit bezüglich geforderter Beschaffenheit“ gefasst wird (Geiger und Kotte, 2008, S. 68). Diese Begriffsdefinition stützt sich auf vier weiteren Begriffsdefinitionen wie die der Beschaffenheit, der Forderung, der Einheit und der Anspruchsklasse, die für ein Definitionsverständnis herangezogen werden. Unter der Einheit wird das einzeln betrachtete bzw. beschriebene verstanden. Liegt eine Gesamtheit der Merkmale oder der Merkmalswerte vor, die einer Einheit angehören, wird sie als Beschaffenheit bezeichnet. Zusammengefasst stellt die Forderung an die Beschaffenheit, die „Gesamtheit der Einzelforderungen an die Beschaffenheit der Einheit“ dar (Geiger und Kotte, 2008, S. 69).

Im Folgenden werden die zwei Qualitätssichten, Design- und Ausführungsqualität näher erläutert. Die Aufgabe der Designqualität obliegt der Umformung von anwenderbezogenen Qualitätsforderungen in eine Spezifikation (Helfert, 2002). Bei ihrer Auswahl von geeigneten Eigenschaften an die Qualitätsforderungen sind die Ziele, die Anwenderbedürfnisse zu verwirklichen, um die Kundenzufriedenheit sicherzustellen (Juran, 1999). Sobald die Spezifikationen feststehen, ist die Ausführungsqualität für die Einhaltung dieser verantwortlich. Dies bedeutet, dass der primäre Fokus nicht auf den Anwenderwünschen, sondern viel mehr auf der fehlerfreien Realisierung der Anforderungen in den Spezifikationen liegt. (Juran, 1999)

Im vorherigen Teil wurde der Begriff Qualität erläutert und eine Basis für den genauso mehrfach definierten Begriff DQ und deren Kriterien geschaffen. Eine Definition kann folgendermaßen lauten: Die „Datenqualität ist die Gesamtheit der Ausprägungen von Qualitätsmerkmalen eines Datenbestandes bezüglich dessen Eignung, festgelegte und vorausgesetzte Erfordernisse zu erfüllen“ (Gebauer und Windheuser, 2021, S. 88).

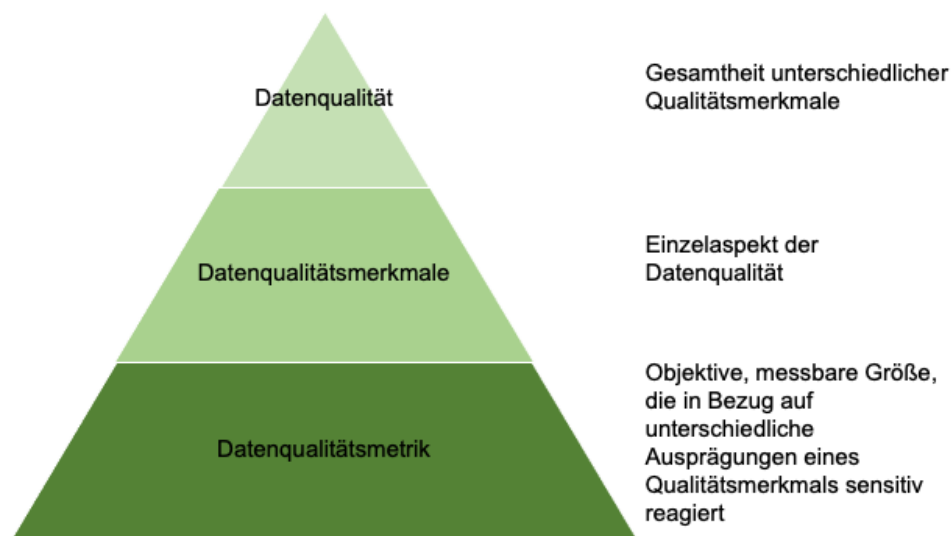


Abbildung 3: Datenqualitätspyramide  
(eigene Darstellung in Anlehnung an Gebauer & Windheuser (2021), S. 88)

Für das Ziel einer Operationalisierung wird die Datenqualitätspyramide als Hilfsmittel herangezogen (siehe Abbildung 3) (Gebauer & Windheuser, 2021). Die vorliegende Pyramide veranschaulicht die drei Stufen einer erfolgreichen Operationalisierung. Dabei bildet der Begriff DQ die Obermenge der sich in der zweiten Stufe befindenden Datenqualitätsmerkmale. Um diese Datenqualitätsmerkmale zu bestimmen, sind Datenqualitätsmetriken der dritten Stufe erforderlich. Sie stellen das Qualitätsmaß dar und ermöglichen somit eine quantitative Aussage. Zudem schaffen sie die operationelle Grundlage, um die Datenqualität zu bestimmen. (Gebauer & Windheuser, 2021) In der Literatur sind die Datenqualitätsmerkmale nicht eindeutig definiert, weshalb mehrere Definitionsansätze vorliegen (Otto & Österle, 2016). Aus dem breiten Portfolio der Datenqualitätsansätze und der Kriterien wurden einige, für diese Arbeit relevante, ausgewählt und in den nächsten Schritten erläutert.

Bei der Auswahl der Werke wurde darauf geachtet, dass ihre Kontexte zu dem Anwendungsgebiet der SC passen und die Ansätze anwenderorientiert sind, da alle SC-Beteiligten als Datenempfänger auf Basis der vorliegenden Daten ihre Interessen umsetzen werden.

Als Vorreiter vieler Arbeiten eignet sich die empirische Arbeit von Wang und Strong (1996), die in einer umfangreichen zweistufigen schriftlichen Befragung das Ziel hatte, Datenqualitätsmerkmale zu bestimmen.



Tabelle 1: Datenqualitätsmerkmale  
 (eigene Darstellung in Anlehnung an Helfert (2002), S. 74 und in Anlehnung an Wang & Strong (1996), S. 20)

Kategorie	Datenqualitätsmerkmale
<i>Innere Datenqualität</i>	Glaubwürdigkeit, Genauigkeit, Objektivität, Vertrauenswürdigkeit
<i>Kontextabhängige Datenqualität</i>	Zusatznutzen, Relevanz, Aktualität, Vollständigkeit, Angemessenes Datenvolumen
<i>Darstellungsqualität</i>	Interpretierbarkeit, Verständlichkeit, Konsistente Darstellung, Knappe Darstellung
<i>Zugangsqualität</i>	Zugriffsmöglichkeit, Zugriffssicherheit

Dabei entstanden im ersten Schritt mithilfe aller teilnehmenden Personen Datenqualitätskriterien, die im zweiten Schritt von weiteren Teilnehmerinnen und Teilnehmern gewichtet wurden. Die Datenanalysen sowie Untersuchungen brachten die vier Kategorien der Datenqualitätsmerkmale hervor (siehe Tabelle 1). (Wang & Strong, 1996) Einen weiteren DQ-Definitionsansatz liefert English (1999), der bei der DQ zwischen der Datendefinitions- und Architekturqualität unterscheidet.

Die Datendefinitionsqualität spiegelt die Qualität der Datenwerte wider, die er ähnlich wie bei Wang und Strong (1996) in die inneren Datenqualitätskriterien und zusätzlich in die pragmatischen Datenqualitätskriterien unterteilt (English, 1999). Hierbei sind die Eigenschaften innerer Datenqualitätskriterien allgemeingültig sowie unabhängig von der Datennutzung. Anders als bei den pragmatischen Datenqualitätskriterien, diese hängen vom vorliegenden Anwendungsfall ab (siehe Tabelle 2). (English, 1999)

Die Architekturqualität umfasst die Qualität der Datenpräsentation, die sich auf die Datendarstellung sowie die Datenbereitstellung bezieht. Dazu gehören Kriterien wie die Interpretierbarkeit, Rechtzeitigkeit und Zugriffsfähigkeit. (English, 1999)

Tabelle 2: Qualitätsmerkmale nach Datenwerten  
 (eigene Darstellung in Anlehnung an English (1999), S. 142f.)

Innere Datenqualitätskriterien
Übereinstimmung mit der Datendefinition
Vollständigkeit der Datenwerte
Plausibel und mit den Geschäftsregeln übereinstimmend
Genauigkeit zu Referenzdaten (Vergleichsdaten)
Genauigkeit zur Realität
Granularität der Datenwerte
Eindeutigkeit (Keine Duplikate)
Übereinstimmung von redundanten und verteilten Datenbeständen
Zeitliche Übereinstimmung von redundanten und verteilten Datenbeständen
Allgemeine Zugriffsfähigkeit

<b>Pragmatische Datenqualitätskriterien</b>
Aktualität und Pünktlichkeit Klarheit / Interpretierbarkeit für den Datenanwender Übereinstimmung zwischen abgeleiteten (berechneten) Daten zu den Ursprungsdaten Nützlichkeit Vollständigkeit bezogen auf den Informationsbedarf

Im betrachteten Anwendungskontext der Informationssysteme wird auch der Ansatz von Redman et al. (1996) eingeordnet. Basierend auf dem Schemakzept von Datenmodellen werden Datenqualitätskriterien für die konzeptionelle Sicht, die Dateninhalte und die Datenrepräsentation aufgestellt (siehe Tabelle 3) (Redman, 1996).

Hierfür wird zunächst zwischen der Datendefinition und den genauen Datenwerten unterschieden.

Die Definition folgt dem zuvor erwähnten Konzept, das heißt ein Datum  $\langle e, a, v \rangle$  mit dem Wert  $v$  aus dem Wertebereich des Attributs  $a$  einer Entität  $e$  wird bestimmt (Redman, 1996). Nach der Festlegung dieses Datenschemas verweist die konzeptionelle Sicht auf die Qualitätsmerkmale der darin enthaltenen Attribute und Entitäten und die Dateninhalte beschreiben, wie die Eintragungen der Datenwerte sein sollten, z. B. korrekt, vollständig, konsistent und aktuell. Die letzte Einteilung, die Datenrepräsentation, erfolgt durch die Formate und die physikalische Speicherung und beschreibt deren Qualitätsmerkmale (siehe Tabelle 3). (Redman, 1996)

Tabelle 3: Datenqualitätsmerkmale  
(eigene Darstellung in Anlehnung an Redman (1996), S. 267)

<b>Konzeptionelle Sicht</b>	Inhalt	<i>Relevanz, Zugriff, Klarheit der Definitionen</i>
	Umfang	<i>Vollständigkeit, Wesentlich</i>
	Detaillierungsgrad	<i>Attributgranularität, Domaingenauigkeit</i>
	Struktur	<i>Natürlich, Identifizierbar, Homogen, Minimum an Redundanz</i>
	Konsistenz	<i>Semantische Konsistenz, Strukturelle Konsistenz</i>
	Reaktionen auf Veränderungen	<i>Stabil, Flexibel</i>
<b>Dateninhalte</b>	<i>Korrekt, Vollständig (Entitäten und Attribute), Konsistenz, Aktualität</i>	
<b>Datenrepräsentation</b>	Format	<i>Angemessen, Formatgenauigkeit, Effiziente Speichernutzung, Interpretierbarkeit, Formatflexibilität, Übertragbarkeit, Fähigkeit Nullwerte abzubilden</i>
	Physikalische Speicherung	<i>Darstellungskonsistenz</i>

Auf Grundlage vorheriger Werke wie von dem Werk von Redman (1996), hat Müller (2000) mit dem Ansatz der Semiotik als Strukturierungshilfe die Datenqualität untersucht und die Datenqualitätsmerkmale in den drei Ebenen, die syntaktische Ebene, die semantische Ebene und die pragmatische Ebene, unterschieden (siehe Tabelle 4).

Tabelle 4: Drei Ebenen der Datenqualitätsmerkmale  
(eigene Darstellung in Anlehnung an Müller (2000), S. 15)

<p><b>Syntaktische Ebene</b></p>	<p><i>Technische Verfügbarkeit der Daten, Technische Verwendbarkeit der Daten, Datensicherheit, Einheitlichkeit von Formaten und Darstellungsformen</i></p>
<p><b>Semantische Ebene</b></p>	<p>Aussagegehalt der Daten: <i>Präzision, Detailliertheit, Validität, Quantifizierbarkeit</i></p> <p>Empirischer und logischer Wahrheitsgehalt: <i>Vertrauenswürdigkeit, Fehlerfreiheit, Konsistenz, Prüfbarkeit</i></p>
<p><b>Pragmatische Ebene</b></p>	<p><i>Zeitliche Eignung für den gegebenen Zweck, Sachliche Eignung für den gegebenen Zweck, Vollständigkeit der Daten für den gegebenen Zweck</i></p>

Zu Beginn der Arbeit wurden die Begriffe Daten und Information voneinander abgegrenzt und festgestellt, dass diese nicht synonym zu verwenden sind und die Daten hauptsächlich die Grundlage für die Informationen schaffen (siehe 2.1.1). Nach dem gleichen Prinzip gilt dies auch für die Begriffe Daten- und Informationsqualität. Da beide Begrifflichkeiten sich nahestehen, werden die zehn Dimensionen der Informationsqualität von Miller (1996) ebenfalls aufgeführt. Die zehn Dimensionen bilden die Eigenschaften: Relevanz, Genauigkeit, Aktualität, Vollständigkeit, Stimmigkeit, Format, Zugänglichkeit, Kompatibilität, Sicherheit und Gültigkeit. Während Millers Artikel sich in erster Linie auf die Kunden bezieht, welche die Informationen erhalten, wird bei den Dimensionserklärungsansätzen dies allgemeiner gehalten und somit auf den Datenempfänger bezogen. Die erste Dimension, Kernkomponente der Informationsqualität, Relevanz zielt darauf ab, dass die vorliegende Information den Bedürfnissen des Datenempfängers entspricht. So können die anderen neun Dimensionen erfüllt sein, dennoch würde sich bei einer irrelevanten Information, diese als unzureichend erweisen. Der Datenempfänger kann die Information, welche auf seinen Nutzen gerichtet ist, besser verstehen und einsetzen (Miller, 1996). Unter der zweiten Dimension, einer

genauen Information wird die Wiedergabe der Realität verstanden. Da in der Praxis verschiedene Anwendungszwecke bezüglich der Information existieren, sind unterschiedliche Grade der Genauigkeit möglich. (Miller, 1996) Eine zu genaue Information, das heißt eine zu präzise Information, hat zur Folge, dass diese die Verarbeitungskapazität des Datenempfängers übersteigt und weitere Kosten in Informationssystemen auslöst. Die dritte Dimension, Aktualität, basiert auf der Erkenntnis, dass sich die Informationen in einem dynamischen Prozess befinden und nur so lange aktuell sind, bis neue Informationen generiert werden und ihre Plätze einnehmen. Dies hat zu bedeuten, dass sich der Status Aktualität, von Datenempfänger-Erkenntnissen abhängig, ständig ändert und neu definiert wird. Bei der vierten Dimension, Vollständigkeit, ist nennenswert, dass unvollständige Informationen den Datenempfänger möglicherweise verwirren. Weiterhin ist anzumerken, dass die Informationen je nach Wissensstand vollständig für die einen und unvollständig für die anderen sein können. Die fünfte Dimension Stimmigkeit beschreibt, dass die Informationen mit sich selbst vereinbar und gut miteinander zusammenhängen müssen. Als sechste Dimension, das Format, wird die dem Datenempfänger vorliegende Präsentation der Informationen verstanden. Hier zählen die Formatform und der Interpretationskontext zu den beiden Komponenten des Formats, die auch die Rahmenbedingungen bilden. (Miller, 1996) Die Formatform hängt hierbei vom Datenempfänger und von seinem Nutzen ab. Bei Bedarf von Informationen müssen diese abrufbar, das heißt verfügbar sein. Diese Verfügbarkeit der Informationen stellt die siebte Dimension dar. Für die Zusammenarbeit der Systeme beschreibt die achte Dimension, dass die Informationen kompatibel mit anderen Informationen sein müssen. Die neunte Dimension Sicherheit legt fest, wie die Informationen sowohl vor unbefugten Menschen durch Passwörter etc. als auch vor natürlichen Schäden mittels Backups geschützt werden. Sind Informationen unsicher, liegt eine Abnahme des Vertrauens in diesen Informationen vor und es fehlt die vollständige Ausschöpfung der Informationen. Die letzte Dimension Gültigkeit wird bestätigt, wenn die Informationen verifiziert und die zuvor beschriebenen Dimensionen Genauigkeit, Aktualität, Vollständigkeit und Sicherheit erfüllt werden. (Miller, 1996) Die zehn Dimensionen von Miller (1996) können von jüngeren Arbeiten wie von Rohweder et al. (2021) abgelöst werden, die auf der Grundlage von Wang und Strong (1996), ein Katalog von 15 IQ-Dimensionen und Definitionen für Datenqualitätsmerkmale aufgestellt haben. Unter den 4 IQ-Kategorien, systemunterstützt, inhärent,

darstellungsbezogen und zweckabhängig, werden die 15 IQ-Dimensionen zugeordnet (siehe Abbildung 4).

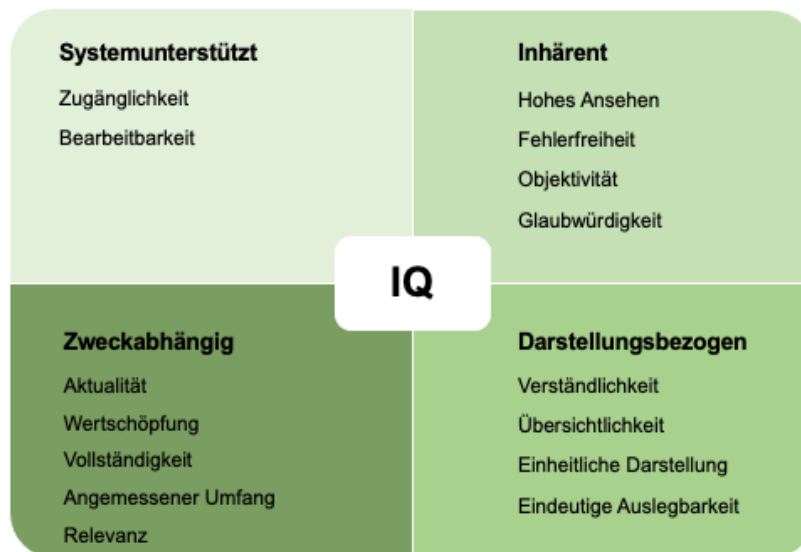


Abbildung 4: 15 IQ-Dimensionen in 4 IQ-Kategorien  
(eigene Darstellung in Anlehnung an Rohweder et al. (2021), S. 28)

Dabei repräsentieren die vier Kategorien ein Ordnungskonzept und beschreiben den jeweiligen zur IQ-Dimension gehörenden Untersuchungsgegenstand für die Überprüfung der DQ. Bei der Untersuchung der systemunterstützten Kategorie sind nur das datenverarbeitende System und die Benutzeroberfläche zu betrachten. Demnach sei in diesem Fall das System zur Datenverarbeitung der Untersuchungsgegenstand (Rohweder et al., 2021). Die inhärente Kategorie setzt bei ihrer Überprüfung auf die Inhalte der Daten. Zur Beobachtung der darstellungsbezogenen Dimensionen steht die eigentliche Darstellung der Daten im Vordergrund. Bei der letzten IQ-Kategorie, der zweckabhängigen Kategorie, wird nur die hauptsächliche Nutzung in den Unternehmensprozessen berücksichtigt. (Rohweder et al., 2021)

Ergänzend zu den zehn Dimensionen von Miller (1996) werden die nicht genannten bzw. neu dazu kommenden Dimensionen beschrieben, um die IQ auf den neuesten Stand zu bringen. Beide Modelle unterscheiden sich zum Großteil und haben nur nachfolgende Dimensionen gemeinsam. Diese sind aus der systemunterstützten Kategorie von Rohweder et al. (2021) die Zugänglichkeit und aus der zweckabhängigen Kategorie; die Relevanz, die Vollständigkeit und die Aktualität. Zu der systemunterstützten Kategorie zählt zusätzlich die Bearbeitbarkeit, die sich auf die leichte Bearbeitbarkeit von Informationen konzentriert, indem diese leicht zu ändern sowie für verschiedene

Zwecke nutzbar sind. Vier weitere, von Miller (1996) nicht erwähnte, Dimensionen sind aus der inhärenten Kategorie das hohe Ansehen, die Fehlerfreiheit, die Objektivität und die Glaubwürdigkeit. Informationen sind dann hoch anzusehen, wenn die Informationsquelle, das Transportmedium und das verarbeitende System, aus einer hohen Vertrauenswürdigkeit und Kompetenz stammt. Ähnlich wie die Dimension Genauigkeit von Miller (1996), steht die Dimension Fehlerfreiheit für fehlerfreie Informationen, wenn eine Übereinstimmung mit der Realität vorliegt. Für die Erfüllung der Dimension Objektivität müssen die Informationen streng sachlich und wertfrei sein und für die Dimension Glaubwürdigkeit sind Zertifikate notwendig, die hohen Qualitätsstandards entsprechen und die Informationsgewinnung und -verbreitung aufwändig betrieben wurde. Laut Rohweder et al. (2021) können die IQ-Begriffe, das hohe Ansehen und die Glaubwürdigkeit, als redundant erscheinen, weshalb eine technische Begriffsbestimmung nötig sei. Demnach wird die Glaubwürdigkeit technisch definiert als „eine Eigenschaft, die von der Aufbereitung der Information bestimmt wird, während hohes Ansehen durch wiederholte positive Erfahrungen mit ähnlichen Informationen (z. B. in einer Zeitreihe) erlangt wird“ (Rohweder et al., 2021, S. 25). Die nächsten vier IQ-Dimensionen sind aus der Kategorie darstellungsbezogen und schließen die Dimensionen Verständlichkeit, Übersichtlichkeit, einheitliche Darstellung und die eindeutige Auslegbarkeit ein. Unter der einheitlichen Darstellung sind Informationen inbegriffen, die fortlaufend auf dieselbe Art und Weise dargestellt werden. Übersichtliche Informationen hängen von einem Format ab, das passend zu den Informationen ist und eine leichte Erfassung dieser ermöglicht. Bei der eindeutigen Auslegbarkeit gilt das Prinzip, dass sie „in gleicher, fachlich korrekter Art und Weise begriffen werden“ (Rohweder et al., 2021, S. 27). Voraussetzung für die Dimension Verständlichkeit sind verständliche Informationen, die direkt von den Anwendern verstanden und für ihre Zwecke zum Einsatz kommen. Aus der zweckabhängigen Kategorie liegen zusätzlich zu den von Miller (1996) bereits thematisierten, zwei Dimensionen die Wertschöpfung und der angemessene Umfang vor. Ein angemessener Umfang ist dann vertreten, wenn die Menge der Informationen ausreichend für die gestellten Anforderungen ist. Die letzte Dimension Wertschöpfung beschreibt Informationen als wertschöpfend, sobald diese zur quantifizierbaren Steigerung einer monetären Zielfunktion beitragen. (Rohweder et al., 2021)

Die Bedeutung der Datenqualität wird durch die Folgen schlechter Datenqualität für ein Unternehmen offensichtlich. Werden Daten schlechter Qualität nicht identifiziert

sowie korrigiert, können sie erhebliche negative Auswirkungen auf die Wirtschaftlichkeit und Sozialität eines Unternehmens haben (Ballou et al., 2004). Dazu zählen aus Unternehmenssicht eine Reihe von Folgen wie ineffiziente Entscheidungsprozesse ausgelöst von bedenklichen Daten, hohe Betriebskosten durch die Korrektur von Datenfehlern, geringere Leistung sowie abnehmende Arbeitszufriedenheit der Mitarbeiter und die geringere Kundenzufriedenheit (Kahn et al., 2002; Lee et al., 2002; Redman, 1998). Zunehmend tritt ein Misstrauen und mangelnde Akzeptanz in die Unternehmensdaten aus Anwendersicht auf, was sich negativ auf die Unternehmenskultur auswirkt (Ryu et al., 2006). Für eine gute DQ ist es deshalb von Bedeutung bereits bei dem Design der Applikation oder des Prozesses Maßnahmen zu ergreifen (Wolf, 2021).



Abbildung 5: Das organisatorische DQ-Modell mit Fehlerarten und Ansätzen (eigene Darstellung in Anlehnung an Wolf (2021), S. 238)

Vor diesem Hintergrund sind für diese Arbeit neben den Erläuterungen zu der Datenqualität Untersuchungsergebnisse bezüglich vorhandener Datenqualitätsmängel unabdingbar und werden als Nächstes thematisiert. Bei den Untersuchungsergebnissen liegt nach Helfert (2000) der Fokus auf den ausschlaggebenden Problemen und Ursachen von Datenqualitätsmängeln. Aus organisatorischer Sicht können Datenqualitätsmängel ihren Ursprung in den klassischen IT-Funktionen haben (Wolf, 2021). Diese sind beispielsweise bei der Ein- und Ausgabe, der Verarbeitung und bei der technischen Speicherung der Daten (Wolf, 2021). Neben den technischen Fehlerquellen

kommen menschliche Fehler wie zu Beginn durch die Erhebung und am Ende durch die Interpretation hinzu (siehe Abbildung 5) (Wolf, 2021).

Als mögliche Ursache werden inkorrekte sowie fehlende Werte, fehlende Referenzen und Duplikate aufgeführt (Helfert, 2000). Weiter sind inkonsistente Daten zwischen unterschiedlichen Anwendungssystemen eine Ursache für Datenqualitätsmängel (Helfert, 2000). Modellierungsfehler sowie redundante Daten verursachen ebenfalls unerwünschte Datenqualitätsmängel. Daraus ergeben sich eine Vielzahl von Problemen wie die Falschbuchungen, Falscherfassungen und systemtechnische Probleme (Helfert, 2000). In den operativen Systemen fehlen Plausibilitätskontrollen sowie die Dokumentation und Bekanntgabe von Änderungen wie beispielsweise von Stammdaten (Helfert, 2000).

Tabelle 5: Ursachen für Datenqualitätsmängel mit Beispielen  
(eigene Darstellung in Anlehnung an Wolf (2021), S. 238)

<b>DQ-Mängel</b>	<b>Mögliche Ursachen (Auswahl)</b>
Prozessfehler	<i>Doppelerfassungen, Daten werden zu einem zu frühen oder zu späten Zeitpunkt verlangt, Ergebnisse, welche den Datenempfänger nicht erreichen</i>
Anwenderfehler	<i>Ungenügende Erhebung, Tippfehler, Verlegenheitseinträge</i>
Programmierfehler	<i>Mehrfachvergabe von Schlüsseln, Datensätze verschwinden oder man kann nicht mehr darauf zugreifen, Auswertungen werden missverständlich umgesetzt, Ungenügende oder fehlerhafte Speicherung von Daten</i>
Kundenfehler (Datenbezügler)	<i>Falsch oder ungenau angeforderte Daten und Auswertungen, Missverständliche Definitionen</i>

Die Tabelle 5 gibt sowohl zuvor genannte als auch ergänzende Ursachen für Datenqualitätsmängel wieder und teilt sie in vier Fehlerarten ein (Wolf, 2021).



## 2.2 Datenbegriff in der Supply Chain

Ziel dieses Kapitels ist es, die Daten im SC-Kontext einzuordnen. Hierzu wird zunächst die SC thematisiert, woraufhin sowohl die Daten als auch die Datenstrukturen in der SC betrachtet werden. Zum Ende des Kapitels werden die DQ und die DQ-Mängel in der SC aufgezeigt.

### 2.2.1 Supply Chain

Nach dem ein grundlegendes Verständnis über das Gebiet der Daten aufgebaut wurde, wird als nächstes ein Einblick in das Gebiet der Supply-Chain und die sich darin befindenden Daten gegeben. Der Begriff „Supply-Chain“ kommt aus dem Englischen und bedeutet übersetzt Versorgungs- oder Lieferkette. Die Bedeutung dieses Begriffs geht aber über die Übersetzung hinaus. (Beckmann, 2004) Bei der Betrachtung der SC-Funktion wird ersichtlich, dass viel mehr ein Wertschöpfungsnetzwerk vorliegt. Denn an der SC sind alle Unternehmen beteiligt, die „an der Entwicklung, Erstellung und Lieferung eines Erzeugnisses (Produkt oder Dienstleistung)“ arbeiten (Beckmann, 2004, S. 1). Die Zusammenarbeit im Rahmen der SC erfolgt vom Rohstofflieferanten bis zum Endverbraucher. In der SC werden dementsprechend alle Unternehmen in einen Wertschöpfungsverbund eingegliedert, „bei dem die Beschaffungs-, Produktions- und Absatzaktivitäten mit denen der vor- und nachgelagerter Unternehmen verbunden werden“ (Pollmeier und Steven, 2008, S. 50). Innerhalb dieses Wertschöpfungsnetzwerkes findet ein sogenannter Leistungsprozess statt, in der die Leistungsobjekte wie Information, Material und Finanzmittel durch das Netzwerk aller Wertschöpfungspartner, alle beteiligten Unternehmen, fließen (Beckmann, 2004). Die dabei entstandenen Beziehungen zwischen den Partnern sind dynamisch und können sich durch die veränderten Anforderungen sowie Strategien der SC ändern (Beckmann, 2004). Es kann zwischen der internen und externen SC unterschieden werden (Wannenwetsch, 2021). Unter der internen SC wird die interne Wertschöpfungskette des Unternehmens wie Personal, Entwicklung, Einkauf, Lagerhaltung, Produktion, Qualitätssicherung, Marketing, Kommissionierung, Verpackung, Transport, Vertrieb und Rechnungswesen verstanden (Wannenwetsch, 2021). Die externe SC beinhaltet die unternehmensübergreifende Vernetzung von Lieferant-Produzent-Handel-Konsument sowie die gesamte Lieferantenkette über mehrere Lieferanten (Wannenwetsch, 2021).

Houlihan (1985) führt weitere Kennzeichen zum Charakterisieren der SC auf und hebt als wesentliches Merkmal die Dokumentation aller Prozesse von der Rohstoffbeschaffung bis zu den Leistungen des Endkunden hervor.

Zu den Prozessen innerhalb der SC zählen Entwicklungs-, Beschaffungs-, Produktions- und Distributionsprozesse und die Interaktion aller Beteiligten wird als ein ganzheitliches System gesehen. Die Prozesse benötigen für ihre Ausführung Ressourcen. Diese Ressourcen sind beispielsweise Standorte, Rohstoffe, Materialien, Transportmittel, Mitarbeiter oder Lager. (Simchi-Levi et al., 2008) Wie der Kern dieser Arbeit ist auch das, von allen Beteiligten zugängliche, Informationssystem für die Ermöglichung der Koordination aller Teilnehmer unerlässlich (Houlihan, 1985). Weiter beschreibt Houlihan (1985) das Hauptziel, den Kundennutzen zu maximieren und dabei ein ausgewogenes Verhältnis zu Kosten und Gewinn zu schaffen. Für die Erfüllung der individuellen Ziele aller beteiligten Partner muss die Leistungsfähigkeit der gesamten Kette gewährleistet sein (Houlihan, 1985). Die SC kann kurzgefasst als eine organisatorische Einheit bezeichnet werden, „die entweder agiert oder reagiert und dadurch eine Leistung (Output) erzielt“ (Biedermann & Kotzab, 2018, S. 57).

Für viele Unternehmen ist im globalen Wettbewerb die SC zum wichtigen Wettbewerbsinstrument geworden, das zum Erreichen vieler Ziele für den Unternehmenserfolg zum Einsatz kommt. Die SC ermöglicht eine schnellere Lieferfähigkeit der Unternehmen als ihre Konkurrenten, die Einhaltung ihrer Liefertermine und die Senkung der SC-Management-Kosten und der Bestände. (Becker & Geimer, 2001) Damit dies gelingen kann, betrachtet beispielsweise der Lieferant des Lieferanten die Datenbank des Kunden seines Kunden und stellt mit Unterstützung der frühen Bedarfsweitergabe eine genaue Planung sicher (Hildebrand, 2004).

Damit der SC-Erfolg sichergestellt werden kann, sind eine Reihe von grundlegenden Anforderungen zu erfüllen. Zu den SC-Anforderungen gehören Anforderungen, welche für die Planung, Beschaffung, Produktion und Lieferung eines Produkts von der SC zu berücksichtigen sind. (Grimm, 2011) Die Planungsanforderung umfasst alle Anforderungen an den zu planenden Bedarf für die SC-Prozesse wie Beschaffung, Produktion und Lieferung. Beispielhaft geht es um die Planung von vorhandenen Ressourcen, Lagerkapazitäten, Mitarbeiterfähigkeiten sowie maschinellen Ressourcen. (Grimm, 2011) Auch die Kombination mit externen Partnern ist möglich. Bei der Beschaffungsanforderung sind alle Anforderungen vorhanden, welche auf das zur Produktion

notwendige Material, dessen Beschaffung, Lagerung und die Auswahl der einsatzbereiten Lieferanten verweisen (Grimm, 2011).

Zur Auftragserfüllung müssen möglicherweise auch Handelswaren ohne Be- oder Verarbeitung beschafft werden. Die Produktionsanforderungen enthalten Anforderungen, die zur Herstellung und zum Testen eines Produkts realisiert werden müssen. Hierzu zählen Anforderungen an Maschinen, Kapazitäten, Testumgebungen oder Mitarbeiterkompetenzen. Bei den Lieferanforderungen geht es um die Erfüllung der Anforderungen an die Lieferung des Produkts. Dabei beziehen sie sich möglicherweise auf die Spedition, die Lieferung sowie Teillieferung oder deren Erlaubnis. (Grimm, 2011)

### 2.2.2 Supply Chain Daten

Im vorherigen Teil standen die Bedeutung, die Anforderungen sowie die Charakteristika der SC im Vordergrund. Nun wird der wesentliche Bestandteil, die Daten, thematisiert und ihre Bedeutung und Funktion im Rahmen der SC eingeordnet. Als einige Ziele der SC liegen kurze Lieferzeiten sowie eine hohe Termintreue vor. Für deren Realisierung ist eine „exakte Disposition und Koordination von Aufträgen sowie Personal- und Betriebsmittelkapazitäten“ erforderlich (Schulte, 2016, S. 125). Hierbei müssen sowohl die Planung als auch die Abwicklung auf Daten beruhen, die korrekt, aktuell und genügend detailliert sind (Schulte, 2016). Zudem kommt beispielsweise die zunehmende Nachfrage nach dem Status der Bestellung durch den Kunden hinzu, die eine zeitverzugslose Datenerfassung nach jeder Teilleistung im Auftragserstellungsprozess erfordert (Schulte, 2016). Das korrekte Erfassen sowie Speichern von Daten stellt die Rückverfolgbarkeit des Produktionsprozesses sowie der Produktkomponenten sicher, die wichtig für die Rechtsprechung bezüglich der Produkthaftung ist (Schulte, 2016). Mit dem Ziel für die Planung und Steuerung des Materialflusses eine geeignete Datenbasis zu sichern, werden nach Schulte (2016) folgende Anforderungen an die logistischen Daten gestellt. Die Daten müssen die fünf Anforderungen Aktualität, Verfügbarkeit, Zuverlässigkeit, Sicherheit und Genauigkeit erfüllen (Schulte, 2016).

Die Supply Chain Daten werden in Stammdaten und Bewegungsdaten (Transaktionsdaten) unterteilt (Kurbel, 2021). Als Stammdaten können die Lokation, das Produkt, die Ressource, das Produktionsprozessmodell (PPM), die Produktionsdatenstruktur (PDS), die Transportbeziehung, Fremdbeschaffungsbeziehung, Quotierung,

Hierarchien, das Modell und die Version gesehen werden (Dickersbach, 2009, S. 17 ff.; Hoppe 2007, S.215 ff.; Wood, 2007, S.63 ff.).

Unter den Stammdaten, Lokationen, „sind physische oder logistische Orte, an denen Produkte hergestellt, gelagert oder transportiert bzw. Ressourcen verwaltet werden“ zuzuordnen (Kurbel, 2021, S. 487). Beispielhaft sind diese das Produktionswerk, das Distributionszentrum, der Kunde, der Lieferant usw. (siehe Tabelle 6) (Kurbel, 2021). Bedeutende Attribute einer Lokation sind die Zeitzone und die Kalender wie Produktions-, Lager- und Versandkalender (Kurbel, 2021).

Tabelle 6: Lokationstypen  
(eigene Darstellung in Anlehnung an Kurbel (2021), S.487)

<b>Lokationstypen</b>		
<i>Produktionswerk</i>	<i>Distributionszentrum</i>	<i>Umlagerungspunkt</i>
<i>Kunde</i>	<i>Lieferant</i>	<i>Transportdienstleister</i>
<i>Transportzone</i>	<i>Lagerort-Dispositionsbereich</i>	<i>Terminal</i>
<i>Geografisches Gebiet</i>	<i>Filiale</i>	

Als Produktstammdaten gelten Materialien sowie Dienstleistungen. Diese umfassen mehr als hundert relevante Attribute und sind somit sehr umfangreich. Des Weiteren werden sie zwischen globalen und lokationsabhängigen Stammdaten unterschieden. Während die globalen Stammdaten die Produkteigenschaften wie Gewicht, Volumen etc. und die Mengeneinheiten beinhalten, bilden die Beschaffungsart, Losgrößen-, Wareneingangs- und Warenausgangseinstellungen die lokationsabhängigen Attribute. (Kurbel, 2021)

Kennzeichen der Ressourcen als Stammdaten sind, dass sie Einheiten bilden, an denen gearbeitet wird und sie dementsprechend über Kapazitäten verfügen. Diese sind beispielsweise Produktionsressourcen wie Arbeiter und Maschinen, Transportressourcen wie Transportmittel und Lagerressourcen wie Lager etc. (Kurbel, 2021)

Ähnliche Datenstrukturen sind das PPM und die PDS, die sich in der Anwendung und in den Details voneinander abgrenzen lassen (Kurbel, 2021). Das PPM verfügt über Kopfdaten wie Lokation und zeitliche Gültigkeit sowie einen PPM-Plan. Dieser Plan ist für die allgemeine Beschreibung der Ressourcen, Materialien und Arbeitsschritte ohne Auftragsbezug zuständig (Hoppe, 2007). Weiter bestimmt er den Produktionsprozess für ein Produkt basierend auf Stücklisten-, Arbeitsplan- und Ressourcendaten (Kurbel, 2021).

Die sich dem PPM ähnelnde PDS weist Unterschiede in ihrer Entstehung auf, aber auch hier liegt eine Kombination von Stücklisten- und Arbeitsplandaten vor (Kurbel, 2021). Die produktorientierten Stammdaten, Transportbeziehungen, beschreiben, welche Lokationen (siehe Tabelle 6) miteinander verbunden und welche Transportmittel hierfür genutzt werden, um ein Produkt vom Sender A zum Sender B transportieren zu können (Kurbel, 2021). Im Fall, dass Informationen über einen Rahmenvertrag mit einem Lieferanten (Vertrag oder Lieferplan) vorliegen, stehen diese Informationen gemeinsam mit einer Transportbeziehung auch in einer Fremdbeschaffungsbeziehung (Kurbel, 2021). Quotierungen, welche ebenfalls zu den Stammdaten gehören, bestimmen allgemein, welcher prozentuale Anteil einer Gesamtmenge von welcher Lokation beschafft oder an welche Lokation geliefert werden soll (Kurbel, 2021). Die Hierarchien als Stammdaten finden Verwendung, wenn die Planung von Produkten oder Lokationen auf einer höheren Aggregationsebene stattfinden soll. Dabei werden die Produkte zu Produktgruppen und die Lokationen zu Lokationsgruppen zusammengefügt und basierend auf der Produkt- und Lokationshierarchie wird die Bildung der Lokationsprodukt Hierarchien erzeugt. (Kurbel, 2021)

Die letzten Stammdaten, Modell und Version, beziehen sich auf die Planung vonseiten des Supply Chain Managements (SCM) mit der Systemanalyse Programmentwicklung (SAP) (Kurbel, 2021) und werden für diese Arbeit nicht weiterhin vertieft.

Bei den nachfolgenden SC-Daten handelt es sich um Bewegungsdaten (Transaktionsdaten), die sich in vier Kategorien unterteilen lassen. Diese Kategorien sind Bestände, Wareneingänge, Bedarfe und Vorhersagen. Innerhalb der Bestände sind verschiedene Varianten wie frei verwendbare, gesperrte, Qualitätsprüf-, Sicherheitsbestände etc. vorhanden. (Kurbel, 2021)

Die Transaktionsdaten treten in verschiedenen Auftragsarten wie in Planaufträgen, Fertigungsaufträgen, Vertriebsaufträgen etc. auf und weiter werden beispielsweise in Tabelle 7 weitere Transaktionsdatenarten unterschieden (Kurbel, 2021). Durch die unterschiedlichen Arten (Bestände, Aufträge usw.) und ihre Zusammenhänge erfolgt eine Auswertung und durch die Verkettung verschiedener Auftragsarten die Entstehung der Lieferketten (Kurbel, 2021).

Tabelle 7: Arten von Transaktionsdaten  
(eigene Darstellung in Anlehnung an Kurbel (2021), S.494)

<b>Transaktionsdatenarten</b>		
<i>Bestand</i>	<i>Vorhersage</i>	<i>Umlagerungsreservierung</i>
<i>Vertriebsauftrag</i>	<i>Planauftrag</i>	<i>Fertigungsauftrag</i>
<i>Bestellanforderung</i>	<i>Unabhängiger Bedarf</i>	<i>Abhängiger Bedarf</i>
<i>Bestellung</i>	<i>Transportauftrag</i>	<i>Lieferplan</i>

### 2.2.3 Datenstrukturen in der Supply Chain

Im vorherigen Teil wurde ein umfassender Einblick in die SC-Daten gewonnen, welche die Datenqualität repräsentieren. Jedoch bevor die „Daten in einer Datenbank gespeichert werden können, muss die aufzunehmende Datenstruktur festgelegt werden“ (Schulte, 2016, S. 110). Mithilfe des logisch-konzeptionellen Datenmodells erfolgt die Erfassung der Objekte mit ihren Eigenschaften und den zwischen ihnen existierenden Beziehungen (Schulte, 2016). Das Entity-Relationship-Modell (ERM) nach Chen (1976) ist dabei eine verbreitete Methode zur Konstruktionsunterstützung von logischen Datenstrukturen die von Schulte (2016) näher beschrieben wird.

Zunächst wird ein zu beschreibendes, echtes Objekt (z. B. Material M1) als Entity definiert und durch seine Eigenschaften (Attribute) beschrieben. Mögliche Attribute des Materials sind beispielsweise die Materialnummer, der Lagerbestand und der Beschaffungspreis. Der Entitytyp umfasst alle vorhandenen gleichartigen Entities und steht demnach für die Materialien. Wenn Entities sich logisch zuordnen lassen, wird dies als Relationen (Beziehungen) bezeichnet. Bei gleichartigen Beziehungen liegen nach ihrer Zusammenfassung Beziehungstypen vor. (Schulte, 2016)

Im Anschluss an die Konstruktion der Datenstrukturen werden die Objekte und die Beziehungen in ein Datenbankmodell-Schema integriert, um die Datenbanksysteme gestalten zu können. Im Wesentlichen existieren zurzeit vier Datenbankmodell-Arten. Diese sind hierarchische, netzförmige, relationale und objektorientierte Datenbankmodelle. (Schulte, 2016)

Hierarchische Datenbankmodelle kommen aufgrund der zunehmenden Komplexität und der schwierigen Darstellung von allen Datenabhängigkeiten im Unternehmen, in neueren Datenbankverwaltungssystemen nicht mehr zum Einsatz (Stahlknecht, 1995) und werden hier nicht weiter behandelt.

Zudem eignen sie sich eher gemeinsam mit den ähnlichen Netzwerkdatenbanken gut für „Routineanwendungen mit häufigen, unveränderlichen Datenbankabfragen“ (Schulte, 2016, S. 112).

Als gängige Datenbankmodelle in der SC finden sich häufig relationale Datenbanken wieder (Scheidler, 2017). Das relationale Datenbankmodell braucht für die Erstellung eines Datenbankmodells nur das Strukturelement Relation und liegt in tabellarischer Darstellungsform vor (siehe Abbildung 6) (Schulte, 2016).

<b>Relation „Artikel“</b>			
<u>Artikel_NUMMER</u>	ARTIKEL_NAME	WAREN_GRUPPE	ARTIKEL_PREIS
15003	QE 1300	A	598,00
37111	CDP 100 A	B	898,60
34590	Sound 7	C	193,70
23676	QE 1700	D	715,50
40400	Quattro B	E	5100,00

Abbildung 6: Beispiel zum relationalen Datenbankmodell (eigene Darstellung in Anlehnung an Schulte (2016), S. 113 und in Anlehnung an Mertens et al. (1996), S. 66)

Durch die moderne Standardsoftware können mithilfe des Relationenmodells die betriebswirtschaftlich-organisatorischen Daten wie „Rohstoffe bzw. Materialien, Halbfertigfabrikate und Enderzeugnisse (Bestände, Bestellungen, Aufträge etc.)“ sowie die „technischen Daten aus der Konstruktion (Stücklisten, Teileverwendungsnachweise etc.), Arbeitsplanung (Arbeitspläne, Bearbeitungszeiten etc.) und Fertigung (Maschinenbelegungen, Anfangs- und Endtermine etc.)“ in einer relationalen Datenbank zusammengefasst werden (Schulte, 2016, S. 112).

Mit den Jahren werden die relationalen Datenbanken zunehmend durch objektorientierte Datenbanken ersetzt (Schulte, 2016). Innerhalb der objektorientierten Datenbanken findet eine Verwaltung der „(Geschäfts-)Objekte bzw. Klassen (Auftrag, Material, Erzeugnis usw.) gemeinsam mit den auf sie auszuübenden Methoden (annehmen, bestellen, fertigen usw.)“ statt (Schulte, 2016, S. 113). Sie unterscheiden sich somit von relationalen Datenbanken, die nur die Eigenschaften der Objekte beinhalten, indem sie auch über die, auf die Objekte anwendbaren, Funktionen verfügen (Schulte, 2016). Zudem wird die Beziehung zu einem anderen Objekt nicht in einem eigenen Beziehungstyp, sondern als Bestandteil des Objektes selbst, abgebildet (siehe Abbildung 7) (Schulte, 2016).

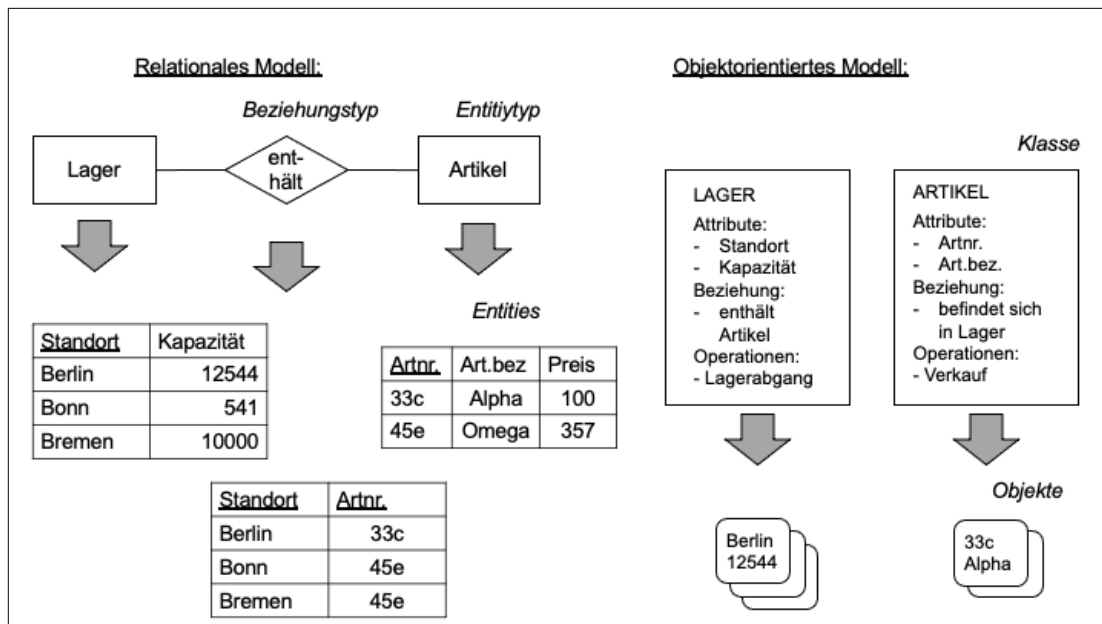


Abbildung 7: Beziehungen im relationalen und objektorientierten Datenbankmodell (eigene Darstellung in Anlehnung an Schulte (2016), S. 113 und in Anlehnung an Mertens et al. (1996), S. 69)

Nach Stahlknecht (1995) werden eine Reihe von zentralen Anforderungen an die Datenbanksysteme gestellt. Diese sind beispielsweise die Datenunabhängigkeit, die Redundanzfreiheit, die Datenintegrität und die Benutzerfreundlichkeit etc. Weitere Anforderungen und ihre Bedeutungen befinden sich in der Tabelle 8.



Tabelle 8: Zentrale Anforderungen an die Datenbanksysteme  
 (eigene Darstellung in Anlehnung an Schulte (2016), S. 108f. und in Anlehnung an Stahlknecht (1995), S. 217f.)

<b>Anforderungen</b>	<b>Bedeutung</b>
Datenunabhängigkeit	Drei Dimensionen: Unabhängigkeit vom Anwendungsprogramm (anwendungsneutrale Speicherung der Daten) Unabhängigkeit der logischen von der physischen Datenorganisation Physische Datenunabhängigkeit
Redundanzfreiheit	Jedes Datenelement sollte möglichst nur einmal gespeichert werden
Datenintegrität	Vollständigkeit, Korrektheit, Widerspruchsfreiheit, exakte und aktuelle Wiedergabe der Realität  (besonders bei verteilten Datenbanken durch unterschiedliche Produktions- und Lagerstandorte)
Benutzerfreundlichkeit	Einfache Datenhandhabung durch einfach zu erlernende Benutzersprachen
Mehrfachzugriff	Jeder Berechtigte darf im Mehrbenutzerbetrieb auf die gespeicherten Daten zugreifen (auch gleichzeitig mit anderen)
Flexibilität	Flexible Verknüpfbarkeit der Daten hinsichtlich beliebiger Attribute aus beliebigen Objekten
Leistung	Kurze Antwortzeiten für die Datenabfrage und -verarbeitung sowie für Änderungen und Ergänzungen des Datenbestandes
Datenschutz	Vermeidung unbefugter Zugriffsmöglichkeiten durch Überprüfen der Zugriffsberechtigung eines Teilnehmers
Datensicherheit	Sicherung gegen Programmfehler und Hardware-Ausfälle

## 2.2.4 Datenqualität in der Supply Chain

An die Daten in der SC werden eine Reihe von Anforderungen bezüglich der Qualitätseigenschaften gestellt, um die Planung, Optimierung und Steuerung des Materialflusses sicherzustellen (Hildebrand, 2021). Gemäß der jüngsten Arbeit von Hildebrand (2021) müssen Daten in der SC „korrekt, aktuell, zeitnah und periodengerecht verbucht, konsistent, vollständig, relevant sowie redundanzfrei sein“ (Hildebrand, 2021, S. 310). Bei Betrachtung der vorliegenden heterogenen und unternehmensübergreifenden Informationssysteme und der Datenbanken-Anwendung sind weitere Eigenschaften von hoher Relevanz. Diese sind die Verfügbarkeit, die Zuverlässigkeit des Datenzugriffs, die Aktualität (regelmäßige Update-Frequenz), die Antwortzeit und die Genauigkeit. (Hildebrand, 2021)

Das Hauptproblem bei der Entstehung von Datenqualitätsmängeln liegt in den Stammdaten, da sie in allen Geschäftsprozessen, dispositiven Anwendungen und Auswertungen auftreten und mit ihrer Qualität einen großen Einfluss auf die Geschäftsabläufe, Bestände und Bedarfe haben (Hildebrand, 2021). Aus diesem Grund werden innerhalb des SCM qualitativ hochwertige Stammdaten angestrebt (Hildebrand, 2021). Durch einige fehlerhafte Stammdaten wie falsche Dispo-Parameter, Losgrößen, Wiederbeschaffungszeiten und Sicherheitsbestände entstehen eine Vielzahl unerwünschter Folgen (Hildebrand, 2021).

Beispielhaft können zu hohe oder zu geringe Bestände und somit Bestandswerte in der Buchhaltung vorliegen. Bei dem Dispo-Lauf können zu viele frühe und späte Bedarfe und Bestellungen entstehen. Auch falsche Aussagen im Hinblick auf die Verfügbarkeit der Menge und des Termins sind möglich. Mit den zuvor aufgeführten Fehlinformationen entsteht eine Unzufriedenheit bei den Kunden und Lieferanten und im Rahmen der SC werden diese Fehleinstellungen weitergegeben. (Hildebrand, 2021)

Neben den Stammdaten können auch Belege (physische und dispositive Bewegungsdaten) zu Datenqualitätsproblemen beitragen. Unter den Belegen sind Aufträge, Bestellungen, Warenein- und -ausgänge, Umlagerungen sowie Planbedarfe zu verstehen. Die in den Belegen enthaltenen Informationen finden täglich sowie jederzeit Verwendung bei der Verfügbarkeitsprüfung, der Bestandsführung und der Disposition. (Hildebrand, 2021)

Sobald ungenaue und verspätet gebuchte Belege vorliegen, sind mit Fehlern in den Beständen sowie Auswertungen zu rechnen. Zu diesen Fehlern zählen keine

zeitnahen Buchungen, falsche Artikel oder Menge, falsche Periode, fehlerhafte Beleg- sowie Bestandarten usw. Daraus resultierend können unterschiedliche Folgen aus mehreren Bereichen zustande kommen. Einige dieser Folgen betreffen die Bestände, die Falschaussagen bei der Materialverfügbarkeit (Menge und Termin), Inventurdifferenzen sowie fehlerhafte Bedarfe aufzeigen. Weitere Folgen machen sich beispielsweise bei dem Dispo-Lauf durch falsche Ergebnisse und bei den Kennzahlen durch falsche Aussagen bemerkbar. Analog zu den Folgen durch die Stammdaten haben Fehler in den Belegen auch die Unzufriedenheit der Kunden und Lieferanten zur Folge. Gleichzeitig sind fehlerhafte Prognosen unabdingbar, da bei der Planung zukünftiger Materialbedarfe vergangene Verbrauchsdaten verwendet werden. (Hildebrand, 2021) Auf Basis früherer Fehlinformationen werden dann durch falsch gebuchte Belege, falsche Materialnummern sowie eine verspätete Erfassung einige Folgefehler in den kommenden Jahren im System bestehen bleiben. Auch die XYZ-Analyse der Materialien wird infolgedessen Fehler bei der Ermittlung des Variationskoeffizienten ( $Variationskoeffizient = \frac{Standardabweichung}{arithmetisches\ Mittel}$ ) enthalten. Weitere Folgen in der Distribution und Produktion sind, dass die Fehler auf der einen Seite keine Lieferung, Versand sowie den Warenausgang und auf der anderen Seite keine Folgeproduktion durch bspw. fehlendes oder falsches Stücklistenmaterial ermöglichen. (Hildebrand, 2021)

### 3. Datenqualität in der Supply Chain

Hauptziel dieser Arbeit ist es, Anforderungen für die DQ in der SC abzuleiten und diese in Kriterien umzuwandeln. Für dieses Vorgehen werden als Nächstes die in 2.1 und 2.2 beschriebenen DQ-Modelle miteinander verglichen und kritisch gegenübergestellt, um daraufhin die Anforderungen an die DQ in der SC abzuleiten. Infolgedessen wird in 3.2 eine Tabelle erstellt, welche die aus den Anforderungen bestimmten Kriterien an die DQ beinhaltet. Mithilfe eines Datensatzes kann in Form einer exemplarischen Validierung in 3.3 die Tabelle, samt der festgelegten DQ-Kriterien, auf ihre Anwendbarkeit hin überprüft und diskutiert werden. Zum Schluss fasst das Fazit alle Ergebnisse zusammen und präsentiert Empfehlungen für eine mögliche weiterführende Forschung.

#### 3.1 Anforderungen an die Datenqualität

Im folgenden Teil werden die Anforderungen an die DQ abgeleitet. Als Einstieg werden mit den beiden Tabellen 9 und 10 ein Überblick über alle zuvor thematisierten Anforderungen sowie Kriterien an die DQ (2.1.3), DQ in der SC (2.2.4) und die Anforderungen an die allgemeinen Datenbanken (2.2.3) sowie die SC-Datenbanken (2.2.4) gegeben.

Tabelle 9: Anforderungen an die DQ der Daten  
(eigene Darstellung in Anlehnung an 2.1.3)

<b>DQ-Kriterien</b>					
<b>Zu untersuchende Werke</b>					
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
Wang und Strong (1996): Datenqualitätsmerkmale	English (1999): Datenqualitätsmerkmale nach Datenwerten	Redman et al. (1996): Datenqualitätsmerkmale	Müller (2000): Drei Ebenen der Datenqualitätsmerkmale	Miller (1996): 10 Dimensionen der Informationsqualität	Rohweder et al. (2021): 15 IQ-Dimensionen in 4 IQ-Kategorien
Vergleich mit Hildebrand (2021): Anforderungen an die SC-Daten					

Tabelle 10: Anforderungen an die DQ der Datenbanken  
(eigene Darstellung in Anlehnung an 2.2.3 & 2.2.4)

<b>Datenbank-Anforderungen</b>	
Stahlknecht (1995): Anforderungen an Datenbanksysteme	Vergleich mit Hildebrand (2021): Anforderungen an die Datenbanken

Aus der jüngsten Arbeit wie von Hildebrand (2021) wird ersichtlich, dass ein funktionierender Informationsfluss im SCM mindestens folgende Qualitätseigenschaften erfordert (siehe 2.2.4):

- Korrektheit
- Aktualität
- Relevanz
- Zeitnahe und periodengerechte Verbuchung
- Konsistenz
- Vollständigkeit
- Redundanzfreiheit

Wie wichtig die oben genannten Kriterien sind und wie sie um weitere Mindestanforderungen ergänzt werden können, werden durch weitere Untersuchungen deutlich. Verschiedene Werke von sämtlichen Autoren in 2.1.3 zeigen unterschiedliche Einteilungen, Ebenen, Dimensionen sowie Kategorien von Kriterien. Durch eine einheitliche Betrachtung dieser DQ-Modelle werden wichtige erkennbare Kriterien hervorgehoben mit dem Ziel die oben genannten Kriterien zu bestätigen sowie zu erweitern. In erster Linie werden quantitative Eigenschaften wie die Häufigkeiten der genannten Kriterien beobachtet. Hierzu werden im ersten Schritt tabellarisch alle erwähnten Kriterien aufgelistet, ohne sie näher nach ihrer vorgesehenen Kategorie oder Einteilung zu spezifizieren (siehe Tabelle 11). Um die Kriterien hinsichtlich ihrer Häufigkeit in dem jeweiligen Modell eins bis sechs zu überprüfen, müssen zunächst alle exakten Doppel- und Mehrfachnennungen eliminiert werden (siehe Anhang B). Hierzu wurden innerhalb aller Kriterien die identischen aussortiert und eine erste Häufigkeitstabelle erstellt (siehe Anhang C). Für die genaue Bestimmung der finalen Häufigkeitstabelle wird dann mit der Semantik gearbeitet, da einige Kriterien als Synonyme auftreten, die bei der Überprüfung der Häufigkeiten zusammengefasst werden können. Die Gesamtheit von ihnen können für das bessere Verständnis näher beschrieben werden.

Tabelle 11: Kriterien aller Werke  
(eigene Darstellung in Anlehnung an 2.1.3)

Kriterien	Untersuche Werke (1 - 6)					
	Wang und Strong (1996)	English (1999)	Redman (1996)	Müller (2000)	Miller (1996)	Rohweder et al. (2021)
	Glaubwürdigkeit	Übereinstimmung mit der Datendef.	Relevanz	Technisch. Verfügbarkeit	Relevanz	Zugänglichkeit
	Genauigkeit	Vollständigkeit der Datenwerte	Zugriff	der Daten	Genauigkeit	Bearbeitbarkeit
	Objektivität	plausibel u. mit den Geschäftsregeln	Klarheit der Definitionen	Technisch. Verwendbarkeit	Aktualität	Hohes Ansehen
	Vertrauenswürdigkeit	übereinstimmend	Vollständigkeit	der Daten	Vollständigkeit	Fehlerfreiheit
	Zusatznutzen	Genauigkeit zu Referenzdaten	Wesentlich	Datensicherheit	Stimmigkeit	Objektivität
	Relevanz	Genauigkeit zur Realität	Attributgranularität	Einheitlichkeit von Formaten	Format	Glaubwürdigkeit
	Aktualität	Granularität der Datenwerte	Domaingenauigkeit	Präzision	Zugänglichkeit	eindeutige Auslegbarkeit
	Vollständigkeit	Eindeutigkeit (keine Duplikate)	Natürlich	Detailliertheit	Kompatibilität	einheitliche Darstellung
	Angemessenes Datenvolumen	Übereinstimmung von redundanten	Identifizierbar	Validität	Sicherheit	Übersichtlichkeit
	Interpretierbarkeit	und verteilten Datenbeständen	Homogen	Quantifizierbarkeit	Gültigkeit	Verständlichkeit
	Verständlichkeit	Zeitliche Übereinstimmung von "	Minimum an Redundanz	Vertrauenswürdigkeit		Relevanz
	Konsistente Darstellung	Allgemeine Zugriffsfähigkeit	Semantische Konsistenz	Fehlerfreiheit		Angemessener Umfang
	Knappe Darstellung	Aktualität	Strukturelle Konsistenz	Konsistenz		Vollständigkeit
	Zugriffsmöglichkeit	Klarheit für den	Stabil (Reaktionen auf Veränderungen)	Zeitliche Eignung für		Wertschöpfung
	Zugriffssicherheit	Datenanwender	Flexibel (")	den gegeb. Zweck		Aktualität
		Übereinstimmung zwischen abgeleit.	Korrekt	Sachliche Eignung für		
		Daten zu den Ursprungsdaten	Vollständig (Entitäten/Attribute)	den gegeb. Zweck		
		Nützlichkeit	Konsistenz	Vollständigkeit der Daten für		
		Vollständigkeit bezogen auf den In-	Aktualität	den gegeb. Zweck		
		formationsbedarf	Angemessen	Prüfbarkeit		
		Pünktlichkeit	Formatgenauigkeit	Einheitlichkeit von Darstellungs-		
		Interpretierbarkeit für den	Effiziente Speichernutzung	formen		
		Datenanwender	Interpretierbarkeit			
			Formatflexibilität			
			Übertragbarkeit			
			Fähigkeit Nullwerte abzubilden			
			Darstellungskonsistenz			

In den DQ-Modellen eins bis sechs taucht das Kriterium Genauigkeit beispielsweise nur zwei Mal auf (siehe Anhang A & B). Bei näherer Betrachtung der Erklärung der Genauigkeit von Miller (1996) (siehe 2.1.3) wird aber festgestellt, dass sie mit dem Kriterium, Fehlerfreiheit, von Rohweder et al. (2021) übereinstimmt und bedeutungsgemäß für die Wiedergabe der Realität steht. Das Kriterium, Fehlerfreiheit, taucht auch in dem Werk von Müller (2000) auf (siehe 2.1.3). Auch in der Arbeit von English (1999) ist das Kriterium, Genauigkeit zur Realität, gegeben, was ebenfalls unter dem allgemeinen Kriterium, Genauigkeit, verstanden werden kann (siehe 2.1.3). Anders als bei Redman (1996), der zwar in seiner Arbeit auf die Domaingenaugkeit sowie Formatgenaugkeit eingeht, aber explizit nicht die Genauigkeit der Daten erwähnt, die das Reale wiedergeben sollen. Zusätzlich werden in Müllers (2000) Arbeit zwei weitere Kriterien, Detailliertheit und Präzision genannt, die zur Bestätigung der Genauigkeit herangezogen und somit auch unter dem Kriterium Genauigkeit eingeordnet werden (siehe 2.1.3). Redmans (1996) Kriterium Korrektheit kann ebenfalls unter dem Kriterium Genauigkeit zugeordnet werden (siehe 2.1.3).

Das nächste Kriterium Konsistenz gehört auch zu den Kriterien, welches in unterschiedlicher Schreibweise vorkommt, aber im Prinzip das Gleiche zur Bedeutung hat. So sind unter konsistenten Daten, Daten zu verstehen, die mit sich selbst vereinbar sind und gut mit anderen Daten zusammenhängen (siehe 2.1.3). Sowohl das Kriterium, konsistente Darstellung, von Wang und Strong (1996) als auch das zusätzliche Kriterium, semantische Konsistenz und die strukturelle Konsistenz der konzeptionellen Sicht, von Redman (1996) stimmen mit der Definition der Konsistenz von Miller (1996) überein. Zudem beschreibt English (1999) als weitere wichtige Kriterien die Übereinstimmung der Daten mit der Datendefinition und den Geschäftsregeln sowie der Plausibilität, welche auch in den allgemeinen Rahmen der Konsistenz eingeordnet werden können.

Eines der offensichtlichsten Kriterien, ist das Kriterium Aktualität, welches mit dem Kriterium zeitliche Eignung für den gegebenen Zweck von Müller (2000), in allen DQ-Ansätzen vorliegt (siehe Anhang A). Auch das Kriterium Relevanz, was sowohl inhaltlich als auch pragmatisch unter den beiden Kriterien, Nützlichkeit und die sachliche Eignung für den gegebenen Zweck, zugeordnet wird (siehe 2.1.3) ist in allen beschriebenen Modellen vorhanden (siehe Anhang A). Das Kriterium, Zugänglichkeit, schließt auch sinngemäß die Kriterien Zugriffsmöglichkeit von Wang und Strong (1996),

allgemeine Zugriffsfähigkeit von English (1999) und den Zugriff von Redman (1996) ein (siehe 2.1.3). Laut Rohweder et al. (2021) sind Informationen als zugänglich definiert, wenn sie mittels einfachen Verfahren und für den Anwender direkt abrufbar sind. Diese Definition kommt dem Kriterium von Müller (2000) sehr nahe, welches als Kriterium die technische Verfügbarkeit der Daten vorsieht (siehe 2.1.3). Neben der Zugriffsmöglichkeit zählt auch die Zugriffssicherheit zu den wichtigen Kriterien, die für den Schutz der Daten vor Unbefugten vorliegen muss (siehe 2.1.3). Die dazugehörigen Synonyme sind die Datensicherheit und allgemein die Sicherheit, die von Miller (1996) beschrieben wird (siehe 2.1.3).

Das nächste untersuchte Kriterium ist die Vertrauenswürdigkeit, welches von den Pionieren der DQ-Arbeiten Wang und Strong (1996) aufgestellt und von Müller (2000) ebenfalls bestätigt wurde (siehe 2.1.3). Bei näherem Vergleichen mit der IQ-Dimension von Rohweder et al. (2021) wird erkennbar, dass die Dimension, hohes Ansehen, ebenfalls die Vertrauenswürdigkeit der Informationsquelle und somit der Information voraussetzt (siehe 2.1.3).

Bei der ersten Betrachtung der Kriterien ist das Kriterium, die Vollständigkeit, ebenfalls auffällig, welches um weitere Kriterien wie die der Vollständigkeit der Datenwerte sowie bezogen auf den Informationsbedarf, vollständige Attribute oder Entitäten sowie die Vollständigkeit der Daten für den gegebenen Zweck erweitert wird (siehe 2.1.3). Der Grund für die Erweiterung ist die gleiche Bedeutung, nämlich nicht fehlende, zur Verfügung stehende, vollständige Informationen aufzuweisen. Des Weiteren ist das Kriterium, Verständlichkeit, gleichermaßen wichtig, da Anwender nur das verstandene Wissen für ihre eigene Zwecke verwenden können (siehe 2.1.3). Dies hat große Ähnlichkeiten mit Englishs (1999) Kriterium, die Klarheit sowie Interpretierbarkeit für den Datenanwender und Redmans Kriterium (1996), Klarheit der Definitionen, weil sie alle auf dem Verständnis basieren. Weiterführend listen neben der Verständlichkeit drei weitere Werke (Wang und Strong (1996), Redman (1996) und English (1999)) die Interpretierbarkeit als weiteres wichtiges Kriterium für die DQ auf (siehe Anhang A).

Ein weiteres relevantes, aber nicht auf dem ersten Blick erkennbares Kriterium, ist das angemessene Format sowie die Formatgenauigkeit von Redman (1996) (siehe 2.1.3). Es befindet sich unter anderem unter den Synonymen Format von Miller (1996), welches unter dem Format die Formatform und den damit verbundenen Interpretationskontext beschreibt, welches das angemessene Format sowie die Formatgenauigkeit voraussetzt. Auch die Einheitlichkeit von Formaten und Darstellungsformen von Müller



(2000) hängen mit dem Format zusammen, da ein einheitliches Format sowie die einheitliche Darstellungsform eine leichtere und übersichtliche Interpretation der Daten erlaubt. In der Arbeit von Rohweder et al. (2021) liegt der Bezug zum Format bei dem Kriterium, Übersichtlichkeit, dieses untermauert die zuvor beschriebene Annahme, da Informationen erst dann übersichtlich sind, wenn die benötigten Informationen in einem passenden sowie leicht fassbaren Format vorliegen (siehe 2.1.3).

Nachfolgend geht es um den Umfang von Daten. Die Autoren von drei von sechs Werken sind der Ansicht, dass sich ein wichtiges Kriterium um den Umfang handeln muss. Der Datenumfang sollte nach Rohweder et al. (2021) angemessen sein, d. h. nur die Menge der Informationen beinhalten, die zur Erfüllung der Anforderungen erforderlich sind (siehe 2.1.3). Auch Redman (1999) ist derselben Ansicht, da er in seinen Kriterien den Umfang als wesentlich festlegt. Dies wird ein weiteres Mal durch das Kriterium, knappe Darstellung, von Wang und Strong (1996) bestätigt.

Unter Millers (1996) Kriterium, Kompatibilität sind Informationen zu verstehen, die kompatibel mit Informationen aus anderen Systemen sind (siehe 2.1.3). In English (1999) DQ-Modell beschreibt er die Übereinstimmung von redundanten und verteilten Datenbeständen sowie die Übereinstimmung zwischen den abgeleiteten Daten zu den Ursprungsdaten und setzt somit auch eine Kompatibilität zwischen den Daten voraus. Weiter sind gleichbedeutende Kriterien die Eindeutigkeit, d. h. keine Duplikate und daraus folgend das Kriterium, Minimum an Redundanz. Auch die Bearbeitbarkeit ist mit der technischen Verwendbarkeit der Daten gleichzusetzen, da somit Operationen wie leichte Änderungen an den Daten sowie Verwendungen an unterschiedlichen Zwecken vorgenommen werden können (siehe 2.1.3). Nach Miller (1996) ist von großer Bedeutung, dass die Daten verifiziert werden (Kriterium Gültigkeit) (siehe 2.1.3). Für die Feststellung dieser Verifizierung müssen weitere Kriterien wie die Prüfbarkeit und die Validierung erfüllt werden, die zusammengefasst werden können. Zudem können auch die Granularität der Datenwerte und die Attributgranularität zusammengeführt werden, da Attribute auch zu den Daten zählen.

Die einheitliche Darstellung ist gleichbedeutend mit der Einheitlichkeit von Darstellungsformen, der homogenen Struktur und der Darstellungskonsistenz, denn die einheitliche Darstellung beschreibt die fortlaufend gleiche Abbildung der Informationen (siehe Rohweder et al., 2021).

Diese, in ihrer Bedeutung übereinstimmenden und allgemein zusammenpassenden, Kriterien sind in der Tabelle 12 abgebildet.

Tabelle 12: Synonyme innerhalb der DQ-Kriterien

<b>Kriterium</b>	<b>Synonyme</b>		
<i>Genauigkeit</i>	Genauigkeit zur Realität	Fehlerfreiheit	Korrektheit
<i>Konsistenz</i>	Konsistente Darstellung	Übereinstimmung mit der Datendefinition/Geschäftsregeln	
<i>Zugänglichkeit</i>	Zugriff	Allgemeine Zugriffsfähigkeit	Technische Verfügbarkeit der Daten
<i>Vollständigkeit</i>	Vollständigkeit der Daten	Vollständigkeit der Attribute u. Entitäten	Vollständigkeit bezogen auf den Informationsbedarf
<i>Format</i>	Angemessenes Format	Formatgenauigkeit	Einheitlichkeit von Formaten
<i>Zugriffssicherheit</i>	Sicherheit	Datensicherheit	
<i>Relevanz</i>	Nützlichkeit	Sachliche Eignung für den gegebenen Zweck	
<i>Verständlichkeit</i>	Klarheit für den Datenanwender	Klarheit der Definitionen	
<i>Einheitliche Darstellung</i>	Homogene Struktur	Einheitlichkeit von Darstellungsformen	Darstellungskonsistenz
<i>Angemessener Umfang</i>	Wesentlich	Knappe Darstellung	
<i>Gültigkeit</i>	Prüfbarkeit	Validität	
<i>Kompatibilität</i>	Übereinstimmung von redundanten u. verteilten Datenbeständen	Übereinstimmung zwischen den abgeleiteten Daten zu den Ursprungsdaten	
<i>Aktualität</i>	Zeitliche Eignung für den gegebenen Zweck		
<i>Minimum an Redundanz</i>	Eindeutigkeit (keine Duplikate)		
<i>Bearbeitbarkeit</i>	Technische Verwendbarkeit der Daten		
<i>Attributgranularität</i>	Granularität der Datenwerte		
<i>Glaubwürdigkeit</i>	Hohes Ansehen		

Nach der Überprüfung aller DQ-Modelle eins bis sechs ergeben sich insgesamt 35 DQ-Kriterien, die mit ihrer Häufigkeit näher bewertet werden. Hierbei steht 1 für die minimale und 6 für die maximale Häufigkeit (siehe Tabelle 13). Mithilfe der Synonyme und sinngemäß identischen Kriterien erfolgt eine Auswertung der Kriterienhäufigkeit.

Tabelle 13: Häufigkeitstabelle gleichbedeutender Kriterien

Nr.	Kriterium	Werke						Häufigkeit
		1	2	3	4	5	6	
1	Genauigkeit	1	1	1	1	1	1	6
2	Vollständigkeit	1	1	1	1	1	1	6
3	Relevanz	1	1	1	1	1	1	6
4	Aktualität	1	1	1	1	1	1	6
5	Zugänglichkeit	1	1	1	1	1	1	6
6	Konsistenz	1	1	1	1	1		5
7	Verständlichkeit	1	1	1			1	4
8	Format			1	1	1	1	4
9	Zugriffssicherheit	1			1	1		3
10	Interpretierbarkeit	1	1	1				3
11	Angemessener Umfang	1		1			1	3
12	Einheitliche Darstellung			1	1		1	3
13	Vertrauenswürdigkeit	1			1		1	3
14	Glaubwürdigkeit	1					1	2
15	Minimum an Redundanz		1	1				2
16	Objektivität	1					1	2
17	Kompatibilität		1			1		2
18	Bearbeitbarkeit				1		1	2
19	Granularität (der Datenwerte)		1	1				2
20	Gültigkeit				1	1		2
21	Wertschöpfung						1	1
22	Eindeutige Auslegbarkeit						1	1
23	Quantifizierbarkeit				1			1
24	Domaingenauigkeit			1				1
25	Genauigkeit zu Referenzdaten		1					1
26	Zeitliche Übereinstimmung von redundant. Und verteilt. Datenbest.		1					1
27	Pünktlichkeit		1					1
28	Natürlich			1				1
29	Identifizierbar			1				1
30	Stabil			1				1
31	Flexibel			1				1
32	Effiziente Speichernutzung			1				1
33	Formatflexibilität			1				1
34	Übertragbarkeit			1				1
35	Fähigkeit Nullwerte abzubilden			1				1

Die höchste Häufigkeit sechs weisen die Kriterien Genauigkeit, Aktualität, Relevanz, Zugänglichkeit und Vollständigkeit auf. Diese fünf Kriterien sind in allen untersuchten Werken vertreten. Danach folgt das zweithäufig genannte Kriterium, die Konsistenz und mit der Häufigkeit vier knüpfen die Kriterien; die Verständlichkeit und das Format an. Die Kriterien Vertrauenswürdigkeit, die Interpretierbarkeit, die Zugriffssicherheit, der angemessene Umfang und die einheitliche Darstellung gehören zu den Kriterien mit der Häufigkeit drei an. Wichtige, aber in ihrer Häufigkeit mit zwei, auftretende Kriterien sind die Glaubwürdigkeit, das Minimum an Redundanz, die Objektivität, die Gültigkeit, die Kompatibilität, die Bearbeitbarkeit und die Granularität der Daten. Die Tabelle 13 bietet einen zusammenfassenden Überblick der DQ-Untersuchungen.

Nach der Erstellung der Häufigkeitstabelle können die Ergebnisse mit den zuvor erwähnten Mindestanforderungen von Hildebrand (2021) verglichen werden. Sowohl in den untersuchten sechs Arbeiten und den von ihm genannten DQ-Kriterien stimmen folgende am häufigsten aufgeführten Kriterien überein; die Genauigkeit (Korrektheit nach Hildebrand (2021)), die Aktualität, die Relevanz und die Vollständigkeit. Die festgestellten Beobachtungen heben die Wichtigkeit dieser DQ-Kriterien besonders hervor. Auch das Kriterium, die Konsistenz, welches am zweithäufigsten genannt wird, ist unter den Mindestanforderungen an die DQ von Hildebrand (2021) vorhanden. Auffällig sind seine beiden letzten Kriterien, die innerhalb der Häufigkeitstabelle gar nicht oder sehr wenig auftreten. Hierbei handelt es sich um die beiden Kriterien, zeitnahe und periodengerechte Verbuchung und die Redundanzfreiheit. Während das Kriterium hinsichtlich der Verbuchung nicht in der Häufigkeitstabelle präsent ist, zählt das Minimum an Redundanz, identisch mit der Redundanzfreiheit, mit der Häufigkeit zwei zu den am wenigsten erwähnten DQ-Kriterien. Hieraus sind weitere Beobachtungen notwendig, um eine genaue Aussage über diese beiden Kriterien zu treffen.

Als weiteres grundlegendes Thema dieser Arbeit ist die Untersuchung der Datenstrukturen wie die der Datenbanken, um ihre Sicht der Anforderungen an die DQ der Datenbanksysteme kennenzulernen und gleichzeitig die Kriterien um die Daten zu erweitern. Hierzu soll die Tabelle 10 betrachtet werden. In heterogenen sowie unternehmensübergreifenden Informationssystemarchitekturen und Datenbankapplikationen sind die Verfügbarkeit, die Zuverlässigkeit des Datenzugriffs (Zugänglichkeit, Lesbarkeit), die Update-Frequenz, die Antwortzeit und die Genauigkeit relevant (siehe 2.2.4). Unter ihnen sind Kriterien vorhanden, welche ebenfalls in der beobachteten Häufigkeitstabelle vorkommen. Diese sind mit der maximalen Häufigkeit sechs das Kriterium Genauigkeit, die Verfügbarkeit (synonym für die Zugänglichkeit) und mit der Häufigkeit drei, die Zuverlässigkeit des Datenzugriffs (bedeutungsgemäß wie die Vertrauenswürdigkeit).

Ältere und weitere Anforderungen an die Datenbanksysteme liegen von Stahlknecht (1995) vor, die ebenfalls erwähnenswert sind. Analog zu Hildebrand (2021) nennt er Anforderungen wie die Redundanzfreiheit, die Datenintegrität (stellvertretend für die Genauigkeit), den Mehrfachzugriff (ähnlich wie die Zugänglichkeit) und die Leistung, welche für die kurze Antwortzeit steht und ebenfalls von Hildebrand (2021) aufgeführt wird (siehe 2.2.3). Obwohl das Kriterium, Redundanzfreiheit, nicht von allen Autoren genannt wird, wird die Bedeutung von Stahlknecht (1995) ein weiteres Mal bestätigt

und lässt Rückschlüsse darüber ziehen, dass dieses Kriterium eine durchaus entscheidende Rolle spielt. Ergänzend ordnet er weitere wichtige Anforderungen wie die Datenunabhängigkeit, die Benutzerfreundlichkeit, die Flexibilität, den Datenschutz und die Datensicherheit ein (siehe 2.2.3), die teilweise in der zuvor bestimmten Häufigkeitstabelle ebenfalls einen Platz einnehmen. Beispielhaft gehört der Datenschutz mit dem Kriterium der Datensicherheit von Miller (1996) dazu, welcher in seiner Arbeit als die IQ-Dimension Sicherheit zusammengefasst wird und in der Häufigkeitstabelle als das Kriterium Zugriffssicherheit mit der Häufigkeit drei auftritt.

Infolge der Häufigkeitstabellenanalyse werden die DQ-Mängel näher untersucht, um weitere bedeutende Anforderungen an die DQ zu erkennen und diese möglicherweise in der Häufigkeitstabelle oder in den Anforderungen an die Datenbanksysteme wiederzufinden. DQ-Mängel können bei ihrer Nichterkennung sowie Nichtbehebung eine Vielzahl von Problemen in einem Unternehmen verursachen (siehe 2.1.3). Besonders ungenaue bzw. fehlerhafte Daten tragen zu hohen Kosten bei, wenn sie korrigiert werden müssen (siehe 2.1.3). Auch aus Anwendersicht entstehen Anwenderfehler durch Ursachen wie die ungenügende Erhebung, Tippfehler und Verlegenheitseinträge, welche die Ungenauigkeit von Daten zur Folge haben. Dies zeigt wieder, wie wichtig das Kriterium, die Genauigkeit, der Daten ist. Entscheidend ist auch, dass die Daten für die Anwender vertrauenswürdig sind, da sonst bei bedenklichen Daten ein Misstrauen in diese Daten besteht, die Akzeptanz zunehmend verloren geht sowie eine Unzufriedenheit aller Beteiligten entsteht (siehe 2.1.3). Eine weitere Ursache für die DQ-Mängel nach Helfert (2000) ist das Fehlen der Referenzen. Bei näherer Betrachtung der aufgestellten Kriterien in der Häufigkeitstabelle (siehe Tabelle 11) ist ersichtlich, dass als Einziger bereits English (1999) mit dem Kriterium, Genauigkeit zu Referenzdaten (Vergleichsdaten), Bezug hierzu genommen hat (siehe 2.1.3). Somit kann das Kriterium, Genauigkeit zu Referenzdaten, in ihrer Wichtigkeit höhergestuft werden, um die DQ sicherstellen zu können. Als Nächstes wird unter der Fehlerart Prozessfehler, als mögliche Ursache die Doppelerfassungen von Wolf (2021) und zusätzlich die Duplikate von Helfert (2000) aufgeführt, die beide das gleiche Problem kenntlich machen und somit die Notwendigkeit des Kriteriums, das Minimum an Redundanz bzw. die Redundanzfreiheit, veranschaulichen (siehe 2.1.3). Weitere Ursachen aus dieser Fehlerart sind zudem, dass die Daten zu einem zu frühen oder zu späten Zeitpunkt verlangt werden oder der Empfänger die Daten gar nicht erhält (siehe 2.1.3). Dies bezieht sich insbesondere auf die Lösung dieser Probleme durch das Kriterium, die zeitnahe

und periodengerechte Verbuchung von Hildebrand (2021) (siehe 2.2.4) und das in der Häufigkeitstabelle aufgelistete Kriterium, die Zugänglichkeit (siehe Anhang F). Ein weiteres Problem ist die Inkonsistenz der Daten zwischen unterschiedlichen Anwendungssystemen, welches daraus schlussfolgernd durch die Beachtung und Einhaltung des Kriteriums, Konsistenz, gelöst wird (siehe 2.1.3). Neben den Prozess- sowie Anwenderfehlern, gibt es auch die Programmierfehler, die einige Ursachen wie z. B. die Mehrfachvergabe von Schlüsseln und die missverständliche Umsetzung von Auswertungen haben (siehe 2.1.3), aber die Lösung in den drei Kriterien Genauigkeit, Verständlichkeit und Interpretierbarkeit liegt. Bei den anderen beiden aufgeführten Ursachen für dieselbe Fehlerart handelt es sich um die ungenügende oder fehlerhafte Speicherung von Daten und das Verschwinden oder keine Zugriffsmöglichkeit mehr auf die Datensätze zu haben (siehe 2.1.3). Hierbei spielt das Kriterium, die Zugriffssicherheit, eine entscheidende Rolle, welches die Abrufbarkeit von Daten sicherstellt und durch Backups Schutz vor Datenverlusten bietet (siehe 2.1.3). Die vierte und letzte Fehlerart Kundenfehler umfasst die falschen oder ungenau angeforderten Daten und Auswertungen sowie die missverständlichen Definitionen als mögliche Ursachen (siehe 2.1.3) und erneut sind die drei Kriterien Genauigkeit, Verständlichkeit und Interpretierbarkeit von hoher Relevanz, um die genannten Ursachen zu umgehen.

Nach der allgemeinen Betrachtung werden bei näherer Beobachtung der SC-Daten und deren möglicher Probleme, welche in den Stammdaten liegen, dieselben Ursachen und Fehlerarten erkennbar, die lösbar durch die gleichen Kriterien sind. Insbesondere die Genauigkeit (Korrektheit) der Daten muss gewährleistet werden, um keine falschen Dispo-Parameter, Losgrößen, Wiederbeschaffungszeiten, Sicherheitsbestände usw. zu erhalten (siehe 2.2.4).

Zusammenfassend werden alle aus der Analyse sichtbaren Erkenntnisse gesammelt, und Anforderungen an die DQ abgeleitet. Der Vergleich aller Werke, der Anforderungen der Datenbanken und der Ursachen für die Entstehung der DQ-Mängel werden primär herangezogen, um aussagekräftige Anforderungen zu erhalten. In Anbetracht der Untersuchungsergebnisse lassen sich nachfolgende Anforderungen an die DQ ableiten.

Auffällig ist hierbei die Anforderung, die *Genauigkeit*, welches in allen Werken sowie in den Anforderungen an die Datenbanken auftritt. Auch spielt dieses Kriterium bei der Vorbeugung vieler DQ-Mängel eine große Rolle, weil es erst die Entstehung vieler Ursachen der Fehlerarten verhindert. Die Aussagen an diese Anforderung,

Genauigkeit, kann man mittels untersuchter Werke näher beschreiben. Die Daten sind dann genau, wenn sie die Realität exakt wiedergeben und mit dieser übereinstimmen, die innere DQ bzw. die Dateninhalte korrekt sind und auf der semantischen Ebene, der empirische und logische Wahrheitsgehalt fehlerfrei ist. Diese Anforderung wird insbesondere in der inhärenten Kategorie abgebildet, weil es sich nur auf die Untersuchung des Daten- und Information-Inhaltes beschränkt (siehe 2.1.3). Somit bildet die Anforderung, Genauigkeit, den Anfang aller noch zu beschreibenden Anforderungen. Die nächste, an die DQ gestellte, Anforderung ist, dass die Daten in allen Werken aus zweckabhängiger bzw. pragmatischer Sicht, *vollständig* bezogen auf den Informationsbedarf sein müssen und zu festgelegten Zeitpunkten in den benötigten Prozessschritten verfügbar sind. Weiter müssen sie auch nach English (1999) und Redman (1996) hinsichtlich der Dateninhalte vollständig sein und dürfen nicht fehlen. Unvollständige Daten führen allgemein dazu, dass die Datenempfänger verwirrt werden, die Daten nicht verstehen und dies sich durch die zunehmende Unzufriedenheit in den Daten bemerkbar macht.

Die Daten in der SC müssen dem Zweck dienen, unternehmensintern sowie -übergreifend dem Datenempfänger bei seiner Entscheidungsfindung zu unterstützen. Deshalb ist auch die nächste Anforderung, die Relevanz, aus zweckabhängiger Sicht besonders entscheidend. Wie Miller (1996) bereits beschreibt, können die Daten alle Anforderungen erfüllen, wenn sie aber nicht relevant sind, liegt kein Nutzen vor (siehe 2.1.3). Relevante Informationen sind Informationen, die auf die Bedürfnisse des Empfängers abgestimmt sind und ihn mit den notwendigen Informationen versorgen, um seinen Nutzen zu steigern.

Neben der genauen Wiedergabe der Realität ist die *aktuelle* Wiedergabe der Realität als weitere Anforderung unausweichlich. Besonders aus pragmatischer Sicht spielen aktuelle Daten eine wichtige Rolle und stehen für Daten, die zeitlich für den gegebenen Zweck geeignet sind und den tatsächlichen Stand der Daten zeitnah darstellen. Auf Grundlage dieser Daten kann der Datenempfänger diese für seine eigenen Zwecke nutzen.

Eine wichtige Anforderung, die indirekt die Daten betrifft, aber für ihre Datenerhebung unerlässlich ist, umfasst die *Zugänglichkeit* der Daten. Aus der systemunterstützten Sicht wird diese Anforderung bei der Betrachtung des datenverarbeitenden Systems deutlich. (siehe 2.1.3) Hierbei muss der Zugriff auf die Daten und den Dateninhalt ermöglicht bzw. sichergestellt werden, um Daten einfach und direkt für den

Datenempfänger verfügbar bzw. abrufbar zu machen. Die Zugänglichkeit betrifft auch den Mehrfachzugriff, da die Vernetzung innerhalb der SC mehrere Anwender anbelangt und allen Berechtigten, ob zeitgleich oder nicht, einen Zugriff auf die Daten ermöglichen muss. Daten, die den Anwender nicht erreichen, weil sie nicht verfügbar sind, rufen aus pragmatischer Sicht erhebliche Probleme hervor und führen wie zuvor beschrieben zu Prozessfehlern, die gerade in der SC ineffizient sind. (siehe 2.2.3)

Im Hinblick auf die unternehmensübergreifenden und verteilten Datenbanken innerhalb der SC ist die Anforderung an die Daten, dass sie *konsistent* sind, besonders von großer Bedeutung. Konsistente Daten sind Daten, die mit sich selbst vereinbar sind und gut mit anderen Daten zusammenhängen (siehe 2.1.3). Während einige Werke sich auf die inhaltliche Konsistenz beziehen, kann eine ganzheitliche Betrachtung der Konsistenz als Anforderung an die DQ angenommen werden. So soll die einheitliche Darstellung gleichbedeutend mit der Darstellungskonsistenz, die semantische und die strukturelle Konsistenz ebenfalls unter dem Kriterium Konsistenz Berücksichtigung finden.

Als eine unverzichtbare und ebenfalls mehrheitlich benannte Anforderung liegt die *Verständlichkeit* vor, die nicht als selbstverständlich angesehen werden sollte und nähere Beachtung verdient. Gerade bei DQ-Mängeln wie Kundenfehler sind Missverständnisse eine mögliche Ursache. Zur Vermeidung dieser Komplikationen sind Daten erforderlich, die für die Datenempfänger verständlich sind bzw. eine inhaltliche Klarheit schaffen und darauf basierend ihre Verwendung ermöglichen.

Die nächste Anforderung, welche darstellungsbezogen ist, schließt das *Format* ein. Notwendige Daten sollten durch ein angemessenes Format repräsentiert werden, welches passend und leicht zu erfassen ist und somit den Interpretationskontext stützt.

Wenngleich die Anforderung *Redundanzfreiheit* in den meisten früheren Werken kaum erwähnt wird, wird ihr innerhalb der SC-Strukturen ein hoher Wert beigemessen. Durch redundante Daten wie z. B. Doppelerfassungen entstehen unangenehme Prozessfehler, deren Beseitigung mit hohem Aufwand (Zeit und Kosten) verbunden ist (siehe 2.1.3). Aus diesem Grund müssen die Daten eindeutig sein, keine Duplikate besitzen und ein Minimum an Redundanz anstreben.

Unter der nächsten Anforderung, *Sicherheit*, sind die Datensicherheit und der Datenschutz zu berücksichtigen. Den Daten ist somit vor internen als auch vor externen Faktoren Schutz zu gewähren. Der Datenschutz soll verhindern, dass Unbefugte Zugriff auf die Daten haben. Daten sollten demnach mit Passwörtern geschützt und nur



Berechtigten den Datenzugriff erlauben. Wie zuvor in 2.1.3 beschrieben lösen ungeschützte, bedenkliche Daten ein Misstrauen bei den Datenempfängern und eine unerwünschte Unzufriedenheit in die Daten aus. Ebenso ist die Datensicherheit für den Schutz vor Ausfällen, Programmfehlern etc. verantwortlich und sichert die Daten gegen mögliche Schäden wie Verluste oder Beschädigungen ab (siehe 2.1.3).

Im Rahmen der SC streben die Beteiligten einen Datenaustausch an. Insbesondere die Anforderung, *Vertrauenswürdigkeit*, muss hierbei erfüllt sein, um das Vertrauen der Partner nicht zu verlieren. Für die Partner muss ersichtlich sein, dass die Datenverarbeitung aus Quellen hoher Vertrauenswürdigkeit und Kompetenz stammt. Auch auf der semantischen Ebene müssen die Daten einen vertrauenswürdigen empirischen und logischen Wahrheitsgehalt aufweisen (siehe 2.1.3).

Eine von Hildebrand (2021) und Stahlknecht (1995) gestellte Anforderung an die Datenbanken ist die *Antwortzeit*, die in den zuvor behandelten Werken nicht thematisiert wird, aber innerhalb der SC aufgrund der Zeiteffizienz für die SC-Prozesse besonders wichtig ist. Demnach sollte die Antwortzeit für die Bearbeitung der Daten wie z. B. die Datenabfrage und -verarbeitung, die Änderungen und Ergänzungen des Datenbestandes möglichst kurz sein (siehe 2.2.3).

Die Hauptaufgabe dieser Arbeit war es unter anderem die Mindestanforderungen an die SC-Daten von Hildebrand (2021) zu erweitern. Eine kaum unbehandelte, aber von ihm übernehmbare Anforderung ist die *zeitnahe und periodengerechte Verbuchung*, die eine wichtige Rolle für die Termintreue des Datenerhalt spielt.

## 3.2 Kriterien der Datenqualität

Nachdem im vorherigen Teil die Anforderungen an die DQ hergeleitet wurden, erfolgt nun ihre Umwandlung in die DQ-Kriterien in der SC. Vor der Tabellarisierung dieser Kriterien findet eine Kategorisierung der aufgestellten Anforderungen statt. Basierend auf den vier Kategorien von Rohweder et al. (2021) werden die 13, in 3.1 genannten, Anforderungen entsprechend eingeteilt. Die Art der Kategorisierung wird bevorzugt, da die Kategorien eine gute und passende Einteilung der DQ-Kriterien erlauben und ein übersichtliches Ordnungskonzept repräsentieren. Diese vier Kategorien mit dem jeweiligen Untersuchungsgegenstand aus 2.1.3 sind:

1. IQ-Kategorie:	inhärent	→	Inhalt
2. IQ-Kategorie:	zweckabhängig	→	Nutzung
3. IQ-Kategorie:	systemunterstützt	→	System
4. IQ-Kategorie:	darstellungsbezogen	→	Darstellung

Die Untersuchung der jeweiligen Kategorie hängt, wie in 2.1.3 beschrieben, von den inhaltlichen Zusammenhängen bezüglich des Untersuchungsgegenstandes ab. Innerhalb der inhärenten Kategorie steht der Inhalt der Daten bzw. Informationen im Untersuchungsfokus. Bei der zweckabhängigen IQ-Kategorie ist die hauptsächliche Nutzung in den Unternehmensprozessen relevant. Die systemunterstützte Kategorie bezieht sich bei der Beurteilung der IQ bzw. DQ auf das System, welches die Daten und somit die Informationen verarbeitet. Als Untersuchungsgegenstand in der letzten Kategorie, der darstellungsbezogenen IQ-Dimension, liegt die eigentliche Darstellung der Daten vor. (siehe 2.1.3)

Nach der Wiederholung der vier Kategorien werden die gestellten Anforderungen in Kriterien umgewandelt und in der dazugehörigen Kategorie eingeordnet. Für jede Einordnung ist jedoch eine Begründung notwendig, da das Kriterium zur passenden Kategorie zugeteilt werden muss, um dem Datenanwender eine gute Bewertung der DQ zu ermöglichen. Gemäß der obigen Reihenfolge liegt als erstes Kriterium die *Genauigkeit* vor. Bei genauer Betrachtung aller Modelle wird die Genauigkeit mit der inneren Datenqualität bzw. Dateninhalten in Verbindung gesetzt und beschreibt, welches inhaltliche Merkmal, die Daten aufweisen müssen. Demnach ist an dieser Stelle der Untersuchungsgegenstand, der Inhalt der Daten, welcher zur Kategorie der inhärenten

Dimensionen angehört. Hiernach folgt das in der Kategorie umstrittene Kriterium *Vollständigkeit*, welches sowohl für die vollständige Verfügbarkeit der Daten aus pragmatischer Sicht als auch der Vollständigkeit in den Dateninhalten zugeteilt werden kann. Der Datenanwender braucht die Daten innerhalb der Arbeitsabläufe in ihrer Vollständigkeit für seinen Nutzen, d. h. innerhalb der zweckabhängigen Kategorie. Dennoch ist auch der inhaltliche Aspekt von hoher Bedeutung, da inhaltlich unvollständige Daten nicht von Nutzen sind. Somit kann das Kriterium Vollständigkeit in beiden Kategorien vorkommen, um die Anforderung in der Gesamtheit zu erfüllen.

Die Zugehörigkeit zur entsprechenden IQ-Kategorie des Kriteriums Relevanz ist aufgrund der Einstimmigkeit aller Werke sehr deutlich. Daten sind erst dann für den Nutzer relevant, wenn sie notwendige Informationen für die Erfüllung des eigentlichen Verwendungszwecks hervorbringen (siehe 2.1.3). Somit gehört das Kriterium die Relevanz, mit dem Ziel des Nutzens, der zweckabhängigen Kategorie an. Bei dem nächsten Kriterium *Aktualität*, ist analog zur Relevanz, das Bestreben aktuelle Daten zu erhalten, ebenfalls damit begründbar, den eigenen Nutzen zu realisieren. Zeitnah abgebildete Daten helfen insbesondere bei der Planung von Arbeitsprozessen und sind demgemäß zweckabhängiger Natur.

Die *Zugänglichkeit*, wie in 3.1 beschrieben, thematisiert den Zugriff auf ein System, um die darin befindlichen Daten einzusehen sowie zu verarbeiten. Aus diesem Grund kann dieses Kriterium eindeutig in die Kategorie der systemunterstützten Dimensionen eingeteilt werden.

Ein ebenfalls mehrfach einsetzbares Kriterium ist die *Konsistenz*. Aus der vorherigen Ableitung dieser Anforderung (siehe 3.1) wird ersichtlich, dass eine ganzheitliche Konsistenz im Sinne der SC-Datenbetrachtung angezielt wird. Dahin gehend wird neben der inhaltlichen Konsistenz, welche der inhärenten Kategorie entspricht, die Darstellungskonsistenz in der darstellungsbezogenen Dimensionskategorie ihren Platz einnehmen.

Nachfolgend wird der Untersuchungsgegenstand des Kriteriums *Verständlichkeit* näher betrachtet. Beobachtungen aus allen Werken zufolge existiert eine Uneinigkeit bei der Einordnung dieses Kriteriums. Während Wang und Strong (1996) und Rohweder et al. (2021) eine darstellungsbezogene Einteilung auswählen, bezieht sich Redman (1996) bei der Verständlichkeit auf den Dateninhalt und somit der inhärenten Kategorie (siehe 2.1.3). Zusätzlich liefert English (1999) einen weiteren Ansatz, indem er die Verständlichkeit, Klarheit für den Datenanwender, in den pragmatischen

Datenqualitätskriterien einordnet (siehe 2.1.3). Um die Einteilung zu spezifizieren wird die widersprüchliche Dimensionsdefinition von Rohweder et al.(2021) näher beobachtet, in der heißt, dass Daten verständlich seien, wenn sie unmittelbar von den Anwendern verstanden und für deren Zwecke zum Einsatz kommen (siehe 2.1.3). In dieser Erklärung steht viel mehr die Nutzung der Daten im Vordergrund als die Datendarstellung. Zudem kommen die Daten für den Nutzen zum Einsatz, wenn sie inhaltlich verstanden werden. Folglich sind die beiden Untersuchungsgegenstände, primär der Inhalt und sekundär die Nutzung, welche dieses Kriterium in die inhärente und zweckabhängige Kategorie eingliedern.

Wie in 3.1 bereits beschrieben handelt es sich bei dem Kriterium *Format* um darstellungsbezogene Merkmale hinsichtlich des Formates. So wird bei der Bewertung der DQ überprüft, wie das Format dargestellt ist und ob es die dazugehörigen Voraussetzungen Einfachheit und leichte Erfassbarkeit erfüllt.

Weiter bezieht sich das Kriterium *Redundanzfreiheit*, welches das Ziel hat, Duplikate innerhalb der Daten zu umgehen, auf die innere DQ bzw. innerhalb der Datenstruktur. Demzufolge kommt dies der inhärenten Kategorie gleich, wenn der bereits bestehende Untersuchungsgegenstand, Inhalt der Daten, um ein weiteres Untersuchungsobjekt, den Inhalt der Datenstruktur, erweitert wird.

Wie das Kriterium *Zugänglichkeit*, verfolgt das Kriterium *Sicherheit* ebenfalls einen systemunterstützten Ansatz, da das System sowohl die Datensicherheit als auch den Datenschutz gewährleisten muss, um das Abrufen sowie Bearbeiten der Daten zuzulassen.

Bei der Überprüfung des nächsten Kriteriums, der *Vertrauenswürdigkeit*, liegt die Betrachtung auf dem Inhalt der Daten. Der Datenanwender überprüft die Daten inhaltlich danach, ob sie auf der semantischen Ebene vertrauenswürdig sind. Neben der inhärenten Kategorieneinteilung besteht, anders als in den zuvor thematisierten Arbeiten, eine zusätzliche Einteilung in die systemunterstützte Kategorie. Der Grund hierfür ist, dass zusätzlich das zu verarbeitende System untersucht wird, um die Vertrauenswürdigkeit der Datenquelle zu kontrollieren.

Ein weiteres, in die systemunterstützte Kategorie passendes, Kriterium stellt die *Antwortzeit* dar, die nur von Stahlknecht (1995) und Hildebrand (2021) benannt, aber nicht nach einer bestimmten Kategorie spezifiziert wurde (siehe 2.2.3 & 2.2.4). Diese Einteilung bietet sich besonders gut an dieser Stelle an, da die kurze Antwortzeit beispielsweise bei der Datenverarbeitung angestrebt wird, die von dem jeweiligen System

erlaubt wird. Demnach besteht bei diesem Kriterium das hohe Interesse das System zu untersuchen.

Das letzte Kriterium, die *zeitnahe und periodengerechte Verbuchung*, ist ebenfalls ausgehend vom jeweiligen System, in der die Daten verbucht werden. Somit handelt es sich hier gleichermaßen um die Zuteilung in der Kategorie der systemunterstützten Dimensionen.

Bei näherer Betrachtung der DQ-Kriterien in der SC und ihrer Zuordnung in der dazugehörigen passenden Kategorie wird erkennbar, dass sich die 13 Kriterien auffällig auf die vier Kategorien verteilen. Ein Großteil der Kriterien verlagert sich auf die inhärente Kategorie, unmittelbar gefolgt von der systemunterstützten (siehe Tabelle 14). Dies hat zur Bedeutung, dass sowohl der Dateninhalt als auch das jeweilige System mehrheitlich im Hauptfokus bei der Untersuchung der DQ in der SC liegen. Darauf folgt die Zweckabhängigkeit der Daten, bei der die Nutzung im Mittelpunkt der DQ-Überprüfung steht. Schließlich beinhaltet die letzte Kategorie, darstellungsbezogen, die wenigsten Kriterien und rückt die Darstellung der Daten im Vergleich zu den anderen Untersuchungsgegenständen in den Hintergrund.

Die geleistete Vorarbeit wird im nächsten Kapitel 3.3, der exemplarischen Validierung, herangezogen, da sie anhand der Kategorien mit den jeweiligen Untersuchungsgegenständen, die Grundlage für die Untersuchung der ermittelten DQ-Kriterien in der SC schafft.

Tabelle 14: Mindestkriterien an die Supply Chain  
(eigene Darstellung in Anlehnung an die Untersuchung und Ergebnisse in 2.1.3, 2.2.3, 2.2.4 & 3.1)

Nr	Kriterium	Bedeutung	Kategorie
1	<i>Genauigkeit</i>	Das Kriterium <i>Genauigkeit</i> beschreibt Daten als genau, wenn sie die Realität exakt wiedergeben und mit ihr übereinstimmen. Zudem sind <i>genaue</i> Daten korrekt und fehlerfrei.	inhärent
2	<i>Vollständigkeit</i>	Das Kriterium <i>Vollständigkeit</i> bezieht sich darauf, dass die Dateninhalte <i>vollständig</i> sind und Daten allgemein nicht fehlen dürfen. Gleichzeitig sind <i>vollständige</i> Daten auf den Informationsbedarf abgestimmt und jederzeit für den Gebrauch verfügbar.	inhärent, zweckabhängig
3	<i>Relevanz</i>	Das Kriterium <i>Relevanz</i> definiert Daten als relevant, wenn sie nach den Bedürfnissen des Datenempfängers ausgerichtet sind und für den Empfängernutzen notwendige Informationen beinhalten.	zweckabhängig
4	<i>Aktualität</i>	Das Kriterium <i>Aktualität</i> steht für Daten, welche für den gegebenen Zweck zeitlich geeignet sind und den tatsächlichen Stand der Daten zeitnah abbilden.	zweckabhängig
5	<i>Zugänglichkeit</i>	Das Kriterium <i>Zugänglichkeit</i> beschreibt den reibungslosen Zugriff auf die Daten. <i>Zugängliche</i> Daten müssen für den Datenempfänger einfach und direkt verfügbar sowie abrufbar sein. Dies gilt auch für den <i>Mehrfachzugriff</i> , bei der mehrere Anwender eine Zugriffsmöglichkeit haben.	systemunterstützt
6	<i>Konsistenz</i>	Das Kriterium <i>Konsistenz</i> bezieht sich auf die Daten, die mit sich selbst sowie anderen Daten sowohl semantisch als auch strukturell vereinbar sind.	inhärent, darstellungsbezogen
7	<i>Verständlichkeit</i>	Das Kriterium <i>Verständlichkeit</i> definiert Daten als verständlich, wenn sie für den Datenempfänger <i>verständliche</i> Daten liefern, die sich in interpretierbare Informationen umwandeln und für den Nutzen der Datenanwender einsetzen lassen.	inhärent, zweckabhängig
8	<i>Format</i>	Das Kriterium <i>Format</i> beinhaltet die Darstellung der Daten in einem angemessenen, passenden und leicht zu erfassenden Format, um den Interpretationskontext zu begünstigen.	darstellungsbezogen
9	<i>Redundanzfreiheit</i>	Das Kriterium <i>Redundanzfreiheit</i> liegt vor, wenn die Daten eindeutig sind, keine Duplikate aufweisen und ein Minimum an Redundanzen anstreben.	inhärent
10	<i>Sicherheit</i>	Das Kriterium <i>Sicherheit</i> wird erfüllt, wenn die Datensicherheit und der Datenschutz gewährleistet werden. Der <i>Datenschutz</i> beinhaltet den Schutz vor Unbefugten, d. h. die Daten müssen über Passwörter etc. geschützt werden, um zur Bearbeitung freigegeben werden zu können. Die <i>Datensicherheit</i> beschreibt den ausreichenden Schutz der Daten vor Datenverlusten, ausgelöst durch beispielweise Ausfälle oder Programmfehler, mithilfe von Backups usw.	systemunterstützt

11	<i>Vertrauenswürdigkeit</i>	Das Kriterium <i>Vertrauenswürdigkeit</i> beschreibt Daten als vertrauenswürdig, sobald die Daten aus vertrauenswürdigen und kompetenten Quellen stammen. Weiter muss der Datenanwender den Dateninhalt auf semantischer Ebene als vertrauenswürdig empfinden.	inhärent, systemunterstützt
12	<i>Antwortzeit</i>	Das Kriterium <i>Antwortzeit</i> definiert die Antwortzeit als die Bearbeitung der Daten (z. B. Datenabfrage und -verarbeitung), welche möglichst kurz sein muss.	systemunterstützt
13	<i>Zeitnahe und periodengerechte Verbuchung</i>	Das Kriterium <i>zeitnahe und periodengerechte Verbuchung</i> wird erfüllt, wenn die Daten im System zeitnah und periodengerecht verbucht werden.	systemunterstützt

### 3.3 Exemplarische Validierung

Nach dem die grundlegenden Kriterien festgelegt worden sind, erfolgt ihre Anwendbarkeit auf den vorliegenden Datensatz. Hierbei handelt es sich um einen Dummy-Datensatz mit dem Titel „Supply chain management for Car“ von der Datenanalyse und Machine Learning Plattform *Kaggle* (Kaggle, 2019). In dem vorliegenden Datensatz wird deutlich, welcher Endkunde welche Fahrzeugbestellung von welchem Hersteller beauftragt hat und welcher Lieferant für die Zustellung des Produkts, des Fahrzeugs, verantwortlich ist.

Tabelle 15: Untersuchte SC-Daten aus dem Datensatz "Supply chain management for Car"

<b>Supplier</b>	<b>Producer</b>	<b>Customer</b>	<b>Order &amp; Shipping information</b>	<b>Payment</b>
<i>Supplier ID</i>	<i>Product ID</i>	<i>Customer ID</i>	<i>Order Date</i>	<i>Sales</i>
<i>Supplier Address</i>	<i>Car Maker</i>	<i>Customer Name</i>	<i>Order ID</i>	<i>Quantity</i>
<i>Supplier Name</i>	<i>Car Model</i>	<i>Gender</i>	<i>Ship Date</i>	<i>Discount</i>
<i>Supplier Contact Details</i>	<i>Car Color</i>	<i>Job Title</i>	<i>Ship Mode</i>	<i>Credit Card Type</i>
	<i>Car Model Year</i>	<i>Phone Number</i>	<i>Shipping</i>	<i>Credit Card</i>
	<i>Car Price</i>	<i>Email Address</i>		
		<i>City</i>		
		<i>Country</i>		
		<i>Country Code</i>		
		<i>State</i>		
		<i>Customer Address</i>		
		<i>Customer Feedback</i>		

Neben den Lieferanteninformationen wie z. B. die Lieferanten-ID, die Lieferantenadresse sowie der Lieferantennamen etc., liegen Informationen bezüglich des Produkts, des Herstellers, des Kunden, des Auftrags und der Zahlung vor. Kundenspezifische Informationen sind beispielsweise die Kunden-ID, der Kundennamen und die Kundenadresse. Weiterführende Informationen befinden sich in der Tabelle 15.



Mithilfe des vorliegenden Datensatzes wird die Anwendbarkeit der, in 3.2 aufgestellten, Kriterien überprüft, um die DQ zu bewerten (siehe Tabelle 14). Während der Überprüfung der Kriterien bietet die vorherige Kategorisierung eine Orientierung, wonach genau untersucht wird. Die erste Kategorie darstellungsbezogen schließt die beiden Kriterien Format und Konsistenz ein und legt den Untersuchungsfokus auf die Darstellung. Auf diese Weise wird im ersten Schritt die Darstellung betrachtet und hinsichtlich des Kriteriums *Format* überprüft. Die tabellarische Darstellung des Datensatzes ist aufgrund der Übersichtlichkeit der hohen Datenmenge sowohl passend als auch leicht erfassbar ausgewählt. Diesbezüglich können die ersten vier Spalten, die Basisadressdaten des Lieferantenstammsatzes beobachtet werden. In einer brauchbaren und übersichtlichen Reihenfolge, Supplier ID, Supplier Address, Supplier Name und Supplier Contact Details, sind sie für den Datenanwender leicht zu erfassen.

Tabelle 16: Lieferantenstammsatz  
aus dem untersuchten Datensatz "Supply chain management for Car"

Supplier ID	Supplier Address	Supplier Name	Supplier Contact Details
1	542 Dayton Center	Bubbltube	871-57-6028
2	0674 Springview Circle	Tagopia	337-64-4060
3	70 Autumn Leaf Center	Zoomdog	218-19-1802
4	649 Corben Lane	Oozz	635-15-3112

Gleichzeitig erlaubt die Darstellung dem Datenanwender einen schnellen Überblick in die Thematik und gewährleistet somit die zeitsparende Einarbeitung in den Interpretationskontext. Erst die Erfüllung dieses Kriteriums ermöglicht die Bewertung anderer DQ-Kriterien für den Datenanwender, welche in diesem Fall vorliegt. Somit wird das Kriterium Format an dieser Stelle eingehalten.

Nachfolgend wird die *Konsistenz* bzw. spezifischer die Darstellungskonsistenz analysiert. Bei Betrachtung der Datensatoberfläche fällt auf, dass sie in gleicher Form vorliegt. Hierzu kann ein Beispiel genannt werden. Bei den Stammdaten Gender (Geschlecht) existieren in der Wertemenge außer female und male, keine zusätzlichen Einträge wie z.B. man oder woman, die den Datenanwender verwirren und der Einheitlichkeit widersprechen könnten. Auch die Bewegungsdaten wie beispielsweise die Daten, order date und ship date, sind einheitlich in der Form 16.12.18 aufgeführt. Eine

mögliche inkonsistente Darstellung wäre z. B. 16.12.18 *und* 16.Dezember 2018. (siehe Tabellen 17 und 18)

Tabelle 17: Stammdaten Gender  
aus dem untersuchten Datensatz "Supply chain management for Car"

Gender
Male
Male
Male
Male
Female
Male

Tabelle 18: Bewegungsdaten Order Date und Ship Date  
aus dem untersuchten Datensatz "Supply chain management for Car"

Order Date	Ship Date
16.12.18	14.03.19
13.02.19	06.03.19
16.08.18	20.01.19
22.02.19	16.03.19

Analog zum Format ergibt sich aus der konsistenten Darstellung, also die Einheitlichkeit der Datensatzoberfläche, eine einfachere Erkennung für den Datenanwender. Mit dieser Erkenntnis wird deutlich, dass erneut ein darstellungsbezogenes Kriterium erfüllt wird.

Die nächsten zu untersuchenden DQ-Kriterien stammen aus der Kategorie der inhärenten Dimensionen und beziehen sich dabei auf den Untersuchungsgegenstand, den Inhalt der Daten bzw. den Inhalt der Datenstruktur bei dem Kriterium Redundanzfreiheit. Dazu werden im Folgenden die DQ-Kriterien Konsistenz, Vertrauenswürdigkeit, Genauigkeit, Redundanzfreiheit, Verständlichkeit und Vollständigkeit zur näheren Untersuchung herangezogen.

Das erste Kriterium aus der inhärenten Kategorie ist die *Genauigkeit* der Dateninhalte. Hierbei wird primär untersucht, ob die Daten die Realität wiedergeben und fehlerfrei sind. Da die kunden-, produkt- und lieferantenspezifischen Informationen schwer ermittelbar sind, wird Bezug zu allgemeineren Informationen wie State und Postal Code

genommen. Stichprobenartig werden drei Datenreihen ausgewählt und auf die Genauigkeit der beiden Informationen geprüft. Die Übereinstimmung beider Daten bestätigt die Richtigkeit. Jedoch ist anzumerken, dass hier eine enorme Menge an Daten vorliegt, welche die Überprüfung für den ganzen Datensatz schwierig macht. Weiter werden exemplarisch erfasste Adressen beobachtet und festgestellt, dass die Aufführung aller Informationen wie bei der Adresse, die Daten genau darstellt. Bei ungenauen Daten könnten beispielsweise die Hausnummer oder die Postleitzahl fehlen, welches das Kriterium Genauigkeit ebenfalls nicht erfüllen würden.

Tabelle 19: Genauigkeitsüberprüfung des Postal Codes, State und City Bewegungsdaten Order aus dem untersuchten Datensatz "Supply chain management for Car"

Supplier ID	Postal Code	State	City	Genauigkeit
1	99522	Alaska	Anchorage	✓
765	43699	Ohio	Toledo	✓
661	36125	Alabama	Montgomery	✓

Im Anschluss folgt die Bewertung des Kriteriums *Vollständigkeit*. Der Datensatz zeigt, dass die Kundeninformationen vollständig dokumentiert sind. Es sind alle relevanten Informationen vorhanden. Jedoch sind auch einige Schwachstellen zu beobachten. Da die Lieferanten- sowie Herstelleradressen teilweise unvollständig sind. Es fehlen z. B. Informationen zu der Herstelleradresse, Lieferantenadresse (City, Country und Postal Code), dem Herstellungsdatum und dem Namen des jeweiligen Auftragsbeauftragten. Außerdem liegen zwar die Lieferantenkontaktinformationen vor, aber nicht der Name der dazugehörigen Ansprechperson (siehe Kaggle, 2019). Somit wird erkennbar, dass das Kriterium Vollständigkeit nicht ganzheitlich erfüllt wird und wichtige zusätzliche Informationen fehlen, die die DQ beeinträchtigen können.

Bei dem nachfolgenden Kriterium, der *Verständlichkeit*, wird der Datensatz mit den sich darin befindenden Daten hinsichtlich des Verständnisses bewertet. Hierzu werden als Beispiel die Produktinformationen betrachtet und festgestellt, dass der Datenanwender die Informationen direkt versteht und weiß, um welches Produkt es sich dabei handelt (siehe Tabelle 20). In dem Fall, dass sich nur die Product ID im Datensatz befinden würde, wäre das Verständnis, um welches Produkt es sich handelt nicht da und müsste daraus resultierend nachermittelt werden.

Tabelle 20: Produktinformationen  
aus dem untersuchten Datensatz "Supply chain management for Car"

Product ID	Car Maker	Car Model	Car Color	Car Model Year	Car Price
8893	Dodge	Ram 2500	Goldenrod	2017	521963.45
9444	Toyota	Tundra	Crimson	2010	672222.04
253	GMC	Savana 1500	Crimson	2011	504465.72
1283	Volkswagen	Cabriolet	Fuscia	1990	646077.11

Aus diesem Grund wird das Kriterium Verständlichkeit bestätigt.

Zur Überprüfung des nächsten Kriteriums, der *Redundanzfreiheit*, wird der Blickwinkel auf den Inhalt der Daten erweitert und die Datenstruktur ganzheitlich auf den Inhalt überprüft. Für diesen Zweck wird der Datensatz auf Excel auf mögliche Duplikate untersucht. Diese Überprüfung ist nicht notwendig bei der ersten Spalte, der Supplier ID, die ohnehin eine laufende Nummer darstellt. Das Ergebnis der ganzheitlichen Betrachtung zeigt, dass keine identischen Datenreihen nachgewiesen werden. Bei näherer Betrachtung der besonders wichtigen Spalten wie der ID (Product ID, Customer ID, Order ID) ist auffällig, dass die Vergabe von ID-Nummern doppelt auftritt und somit das Kriterium der Redundanzfreiheit verletzt wird, da die Daten eindeutig vorliegen müssen (siehe Tabelle 21 a. bis c.).

Tabelle 21: Redundanzen innerhalb der Datenstruktur  
aus dem untersuchten Datensatz "Supply chain management for Car"

a.

Supplier ID	Product ID	Car Maker	Car Model	Car Color
6	8877	Toyota	Land Cruiser	Crimson
511	8877	Mazda	MX-5	Puce

b.

Supplier ID	Customer ID	Customer Name
21	49693-1801	Barbee Cathel
587	49693-1801	Othello Whettleton

c.

Supplier ID	Order Date	Order ID	Ship Date
19	03.06.19	0363-0459	30.04.19
244	14.07.18	0363-0459	13.03.19

Um das nächste Kriterium, die *Konsistenz* der Dateninhalte, zu bestätigen, sind für den Nachweis dieses Kriteriums mehrere Datenbanken erforderlich. So kann festgestellt werden, ob der Inhalt der Daten mit denen aus anderen Datenbanken übereinstimmt. In der SC sind bekanntlich mehrere verteilte Datenbanken gängig, weshalb die Überprüfung dieses Kriteriums an einem Datensatz in diesem Fall nicht ausreicht.

Für die Feststellung des nächsten Kriteriums *Vertrauenswürdigkeit* sind vertrauenswürdige Dateninhalte für den Datenanwender erforderlich. Auf semantischer Ebene liegen logische Dateninhalte in Form von verständlichen Zeichenketten vor, die für den Datenanwender in erster Linie vertrauenswürdig erscheinen. Jedoch ist dieses Kriterium gemeinsam mit der systemunterstützten Kriterien-Einteilung zu betrachten, da primär der Datenursprung die Dateninhalte in ihrer Vertrauenswürdigkeit bewahrheitet und in diesem Fall fiktive Daten vorliegen.

Anschließend erfolgt die Untersuchung der DQ-Kriterien der zweckabhängigen Kategorie, welche den Fokus auf die Nutzung legt. Angefangen mit dem Kriterium *Relevanz*, wird überprüft, ob die Daten für den Datenanwender nützliche Informationen liefern. Auffällig ist, dass die meisten Daten relevant sind, da sie wichtige Kontaktinformationen, Zuordnungen mithilfe der ID-Nummern und Produktinformationen repräsentieren. Gleichzeitig erlauben die Versandinformationen eine gute Rückverfolgbarkeit des jeweiligen Produkts. Unter allen Daten sind keine irrelevanten Daten vorhanden, da jedes Datum für den spezifischen Nutzen informativ ist. Ein mögliches Beispiel ist das Geschlecht des Kunden, welches im ersten Moment unwichtig erscheint, aber für die Anrede beim Kontaktieren des Kunden unerlässlich ist. Somit ist zu sagen, dass das Kriterium Relevanz erfüllt wird.

Bei der Überprüfung des nächsten zweckabhängigen Kriteriums, der *Aktualität*, ist die Betrachtung der Nutzung hinsichtlich der Zeitdimension nicht abgebildet, da sich der vorliegende Datensatz nicht fortlaufend ändert und keine neuen Datenwerte generiert, bei der die Dateneigenschaften aktualisiert werden könnten. Beobachtbar ist jedoch, dass sich ausgehend von dem Jahr 2019 ältere Daten des Vorjahres (2018) befinden, die der zeitnahen Datenabbildung widersprechen und in diesem Fall veraltet sind (siehe Tabelle 18). Nichtsdestotrotz kann keine genaue Aussage bezüglich der Aktualität des vorliegenden Datensatzes getroffen werden.

Da zuvor die Anwendbarkeit der *Vollständigkeit* aus der inhärenten Kategorie geprüft wurde, wird sie nun im zweckabhängigen Kontext betrachtet. Aus Anwendersicht werden vollständige Daten für den eigenen Verwendungszweck benötigt, die beim ersten

Anblick der Datenfelder des Datensatzes ersichtlich sind. Alle Datenfelder weisen zwar vollständige Einträge und keine fehlenden Daten auf, sodass bei Bedarf der Datenanwender diese nutzen wird. Jedoch fehlen, wie auch in der inhärenten Kategorie angedeutet, ergänzende möglicherweise wichtige Informationen, die für den unmittelbaren Gebrauch nicht vorhanden sind.

Ein weiteres, sowohl in der inhärenten als auch in der zweckabhängigen Kategorie zuordenbares, Kriterium ist die *Verständlichkeit*, welche im Vorfeld innerhalb der inhärenten Kategorie bestätigt worden ist. Aufgrund der Verständlichkeit der Daten ergibt sich die Umsetzung der Datennutzung für den Datenanwender und die Einhaltung dieses Kriteriums.

Letztlich wird die Anwendbarkeit der Kriterien; *Vertrauenswürdigkeit*, *Zugänglichkeit*, *Sicherheit*, *Antwortzeit* und *zeitnahe und periodengerechte Verbuchung* aus der systemunterstützten Kategorie bewertet und als Untersuchungsgegenstand das Datensystem verwendet. Angesichts der Tatsache, dass an dieser Stelle kein Datensystem, sondern lediglich ein Datensatz vorliegt, wird die Überprüfung dieser DQ-Kriterien nicht erlaubt. Exemplarisch werden die Kriterien und ihre Nichtanwendbarkeit aufgezeigt.

Für die Anwendbarkeit des Kriteriums *Vertrauenswürdigkeit* wird das System auf die Generierung seiner Daten betrachtet, bei der die Quellen einen guten Ruf aufweisen müssen. In dem vorliegenden Fall existiert weder ein System noch Vorinformationen zur Entstehung der Daten, welches keine diesbezügliche Bewertung zulässt. Die beiden Kriterien *Zugänglichkeit* und *Sicherheit* werden ebenfalls nicht näher untersucht, da das System im Normalfall die Abrufbarkeit der Daten bzw. den Zugriff auf die Daten erlaubt und die Sicherheit der Daten mittels Passwörter, Backups etc. garantiert. Bei dem Kriterium *Aktualität* wurde vorher festgestellt, dass kein dynamischer Zustand besteht und keine Daten neu generiert oder aktiv geändert werden. Dies begründet, wieso das Kriterium, *zeitnahe und periodengerechte Verbuchung*, weder anwendbar noch überprüfbar ist.

Gleichzeitig gestattet das nicht-vorhandene System keine Datenverarbeitung, welches normalerweise die Überprüfung der Antwortzeit erlauben würde.

Anhand der exemplarischen Validierung wurde überprüft, welche SC-Kriterien inwieweit auf den SC-Datensatz anwendbar sind und wie sie sich bewerten lassen. Im nächsten Teil, 3.4 Diskussion und Fazit, werden die in 3.2 und 3.3 gewonnenen Ergebnisse gegenübergestellt und kritisch bewertet.

### 3.4 Diskussion und Fazit

Im abschließenden Teil werden die Ergebnisse zusammengefasst und näher interpretiert. Dann werden mögliche Implikationen der Theorie in der Praxis thematisiert und Limitationen innerhalb der Arbeit präsentiert und ihre Behebung durch künftige Forschungsarbeiten vorgestellt. Weiter werden Potentiale für weitere Forschungsarbeiten angesprochen.

Zu Beginn der Arbeit wurden in 2.1.3 sechs relevante Werke zu verschiedenen DQ-Kriterien behandelt, welche den Grundstein der Erstuntersuchung in 3.1 gebildet haben. Mit dem Ziel die bisherigen genannten Anforderungen an die SC-Daten zu bestätigen und zu erweitern, wurden im ersten Schritt alle Kriterien der Werke verglichen und ihre Häufigkeit bestimmt und im nächsten Schritt sowohl den SC-Daten als auch den Anforderungen an die Datenbanksysteme gegenübergestellt. Zur Bestätigung der Bedeutung der genannten Kriterien wurden auch DQ-Mängel überprüft. Schließlich wurden 13 wichtige Kriterien abgeleitet, welche in die vier Kategorien, inhärent, darstellungsbezogen, zweckabhängig und systemunterstützt unterteilt wurden. Die Einordnung war insbesondere für die exemplarische Validierung der Kriterien anhand eines SC-Datensatzes relevant, um die Untersuchungsgegenstände zu spezifizieren und auf diese Weise die Anwendbarkeit der Kriterien auf den Datensatz zu ermöglichen. Erste Untersuchungsergebnisse ergaben die finale Häufigkeitstabelle mit den am häufigsten genannten Kriterien, von denen die Mehrheit der ersten 13 Kriterien besonders Berücksichtigung findet (siehe Tabelle 13). Ergänzend durch die SC-Kriterien von Hildebrand (2021) werden 13 relevante Kriterien abgeleitet, die in jedem Fall erfüllt sein müssen, um die DQ innerhalb der SC sicherzustellen (siehe Tabelle 14).

Tabelle 22: Anwendbarkeit der Kriterien und ihr Einfluss auf die Datenqualität

Nr.	Kriterien	Kategorie	Anwendbarkeit	DQ-Einfluss
1	Format	darstellungsbezogen	ja	↑
2	Konsistenz	darstellungsbezogen	ja	↑
		inhärent	nein	-
3	Genauigkeit	inhärent	teilweise	-
4	Vollständigkeit	inhärent	ja	↓
		zweckabhängig	ja	↓
5	Verständlichkeit	inhärent	ja	↑
		zweckabhängig	ja	↑

6	Redundanzfreiheit	inhärent	ja	↓
7	Vertrauenswürdig- keit	inhärent	teilweise	-
		systemunterstützt	nein	-
8	Zugänglichkeit	systemunterstützt	nein	-
9	Sicherheit	systemunterstützt	nein	-
10	Zeitnahe und perio- dengerechte Verbu- chung	systemunterstützt	nein	-
11	Antwortzeit	systemunterstützt	nein	-
12	Relevanz	zweckabhängig	ja	↑
13	Aktualität	zweckabhängig	nein	-

Bei der Übertragung dieser Mindestkriterien auf den SC-Datensatz sind folgende Erkenntnisse auffällig. Die beiden darstellungsbezogenen Kriterien, das Format und die Darstellungskonsistenz, sind unmittelbar anwendbar. Aufgrund der Tatsache, dass sie erfüllt werden und daraus resultierend dem Datenanwender flexible Einblicke gewähren, erhöhen sie die DQ innerhalb des Datensatzes maßgeblich. Zudem sind sie unabhängig von der Art des Datensatzes anwendbar und liefern schnelle Aussagen der Qualität der Darstellung und somit der damit verbindlichen DQ. Gleichzeitig erfolgt die Überprüfbarkeit beider Kriterien innerhalb kurzer Zeit und erweist sich bei ihrer sofortigen Behebung als vorteilhaft, da spätere Korrekturen zeitaufwändig sind.

Weitere Untersuchungen innerhalb der Kriterien aus der inhärenten Kategorie zeigen folgende Ergebnisse. Bei der Überprüfbarkeit des Kriteriums der Genauigkeit liegt eine erste Limitation vor, da es sich bei dem Großteil der Daten um fiktive Daten handelt. Aber bei universell deklarierten Daten wie die Postleitzahl und die dazugehörige Stadt treten Übereinstimmungen auf. Zudem sind die fiktiven Adressendaten genau nach dem Schema Nummer-Straße definiert. Nichtsdestotrotz kann hier keine vollständige Überprüfung der Wiedergabe der Realität, welche für die Genauigkeit entscheidend ist, stattfinden und keine Entscheidung über die Wirksamkeit dieses Kriteriums auf die DQ getroffen werden. Bei dem nächsten inhärenten Kriterium, der Vollständigkeit, tritt eine Verletzung dieses Kriteriums auf, welche die DQ negativ beeinflusst. Unvollständige Daten erzeugen zeitaufwändige Nachermittlungen für den Datenanwender und zusätzliche Arbeit, durch welche die Daten an Ansehen verlieren. Einen positiven direkten Einfluss auf die DQ hat das Kriterium Verständlichkeit, da der Datenwender die Daten schnell versteht, daraus seine erforderlichen Informationen ableitet und darauf basierend handeln wird. Anders als beim vorherigen Kriterium resultieren negative



Erkenntnisse bei der Untersuchung des Redundanzfreiheit Kriteriums, da unerwünschte Duplikate vorliegen. Falls der Datenanwender nur eine der vorliegenden ID besitzt und die Informationen damit abrufen wird, werden ihm zwei Datenreihen erscheinen, die keine eindeutige Zuordenbarkeit gewähren. Aus diesem Grund werden mehrere Informationen benötigt, bei der die Beschaffung dieser wieder Zeit in Anspruch nimmt, obwohl sie im Vorfeld behoben werden könnte.

Um eine Verbindung dieser Erkenntnisse mit dem Stand der Technik herzustellen, wird Bezug zum Kriterium Konsistenz genommen, bei der die Dateninhalte auf die Konsistenz geprüft werden. An der SC sind mehrere Unternehmen beteiligt und somit mehrere verteilte Datenbanken vorhanden, bei der es keine Zentralisierung der Daten gibt. Dies begünstigt inkonsistente Daten besonders, da keine einheitliche Datenbank vorliegt, sondern mehrere, die sich unabhängig voneinander ändern können. Daher ist auch keine Anwendbarkeit dieses Kriteriums möglich, weil nicht mehrere Datenbanken gleichzeitig betrachtet werden, was wiederum eine erneute Limitation innerhalb dieser Arbeit darstellt. Während der Validierung wird erkennbar, dass einige Kriterienanwendungen erst dann sichere Ergebnisse aufzeigen, wenn die Erfüllung durch eine weitere Kategorie (z. B. Vertrauenswürdigkeit inhärent-systemunterstützt) bestätigt wird. Anknüpfend an die vorhin genannte Kategorie, werden nun die Ergebnisse der systemunterstützten Kriterien betrachtet. Aufgrund der Tatsache, dass hier kein Datensystem, sondern ein Datensatz vorliegt, ist eine Anwendung aller Kriterien aus dieser Kategorie nicht möglich, welche erneut keine Aussagen über den Einfluss auf die DQ schaffen. In diesem Fall liegt erneut eine Limitation vor, zu der im nachfolgenden Verlauf hinsichtlich ihrer Behebung Verbesserungsvorschläge gemacht werden. Allerdings werden im Vorhinein die Kriterien der letzten, zweckabhängigen Kategorie betrachtet und ihre Ergebnisse geschildert. Positiv sind die Erkenntnisse bezüglich der Relevanz und der Verständlichkeit. Infolge ihrer Anwendbarkeit auf den Datensatz zeichnet sich ein vorteilhafter Einfluss auf die DQ aus, da der Datenanwender die größtenteils nützlichen Informationen verstehen und sie zweckabhängig einsetzen wird. Daneben sind die Ergebnisse der anderen beiden Erkenntnisse von Nachteil, da analog zur inhärenten Vollständigkeit, die zweckabhängige ebenfalls nicht eingehalten wird und bei dem Kriterium Aktualität kein datenaktualisierendes System vorliegt, welches explizit untersucht wird.

Zusammenfassend ist zu sagen, dass ein Großteil der ein- oder zweifach kategorisierten Kriterien, gar nicht oder nur teilweise anwendbar sind und bei allen keine Aussage

über die Entwicklung der DQ getätigt werden kann. Mehrheitlich lassen sich die Kriterien nicht systembedingt untersuchen und wertvolle Erkenntnisse gehen dabei verloren. Die nicht mögliche Anwendbarkeit ist jedoch nicht mit der Qualität der aufgestellten Kriterien zu begründen, da innerhalb der Untersuchung auch positive Beispiele existieren. So sind bei den anwendbaren Kriterien bemerkbar, dass sich immer ein konkretes Ergebnis bezüglich der DQ-Einflüsse bemerkbar macht. Der Großteil der anwendbaren Kriterien wird zum Positiven bewertet (siehe Tabelle 22). Insbesondere die kategorische Überschaubarkeit der Kriterien sowie die damit einhergehenden Untersuchungsgegenstände erleichtern die Anwendbarkeit dieser in hohem Umfang und bringen Struktur in die Untersuchung ein. Zum Nachteil wird dies aber, wenn einige Untersuchungsgegenstände, wie zuvor bestimmt, nicht greifbar sind. Demgegenüber stehen aber die Vielzahl an Potentialen, die durch die Implikationen der Theorie in der Praxis entstehen. Um kurzerhand die DQ innerhalb der Datenbank zu überprüfen, besteht für den Datenanwender die Möglichkeit den Untersuchungsfokus auf die benannten Untersuchungsgegenstände sowie die darin enthaltenen Kriterien zu lenken. Zeit sparend und effizient können Fehler behoben werden, welche unter Umständen die Ursachen für künftige DQ-Mängel darstellen.

Schließlich wird eine Empfehlung gegeben, womit die zuvor beschriebenen Limitationen umgangen werden können. Basierend auf den in dieser Arbeit hergeleiteten Kriterien kann gemeinsam mit weiteren Kriterien, ohne den Fokus nur auf die abgeleiteten Kriterien zu lenken, ein Katalog erstellt werden, welcher den SC-Beteiligten zur Bewertung vorgelegt werden kann. Die Untersuchungsergebnisse dieser empirischen Arbeit können Rückschlüsse darüber geben, ob die mindestens angeforderten SC-Kriterien bestätigt werden oder Ergänzungen benötigen. Daraufhin kann die Untersuchung aller Untersuchungsgegenstände sichergestellt und somit die Kriterien an einem System mit echten Daten und bekanntem Zweck erneut angewandt werden.

## 4. Zusammenfassung und Ausblick

In der vorliegenden Arbeit wurden für die SC wichtige DQ-Kriterien erarbeitet, welche in den SC-Datenbanken besondere Beachtung verdienen und die Unternehmen für das Erreichen einer besseren DQ unterstützen.

Zu Beginn dieser Arbeit wurden DQ-Anforderungen von sämtlichen Autoren und DQ-Mängel vorgestellt und im nächsten Schritt miteinander detailliert verglichen. Der Vergleich hatte den Zweck, besonders häufig genannte und somit als wichtig eingestufte Kriterien zu identifizieren. Mithilfe dieser Identifikation und der Betrachtung der DQ-Mängel konnten die Kriterien mit den aus der Literatur bekannten SC-Mindestkriterien von Hildebrand (2021) bestätigt werden und zusätzlich diese SC-Kriterien erweitern. Nach der Ableitung der DQ-Kriterien fand die Kategorisierung der Kriterien statt, um dem jeweiligen Kriterium seinen Untersuchungsgegenstand zuzuordnen. Schließlich half die Kategorisierung bei der Untersuchung der Kriterien auf ihre Anwendbarkeit während der exemplarischen Validierung. Die exemplarische Validierung mithilfe eines SC-Datensatzes konnte die Anwendbarkeit der Kriterien aufzeigen und Rückschlüsse über die DQ ziehen.

Zusammenfassend lässt sich sagen, dass die abgeleiteten Kriterien eine übersichtliche Basis schaffen, um ein hohes Niveau an DQ zu erreichen. Insbesondere die unterschiedliche Kategorisierung half dabei das vorliegende Datensystem ganzheitlich zu betrachten und mehrere, im Hinblick auf die unterschiedlichen Untersuchungsgegenstände, für die DQ wichtigen Aspekte zu berücksichtigen. Das breite Spektrum an Kriterien innerhalb vieler Werke wurde mithilfe der Untersuchung anschaulich eingegrenzt, welche die Untersuchung der DQ maßgeblich erleichterte. Zudem wurde ersichtlich, dass bestimmte Voraussetzungen zur Untersuchung der DQ existieren. Hierzu zählte die einwandfreie Zugriffsmöglichkeit auf alle Untersuchungsgegenstände. Im Hinblick auf die Gegebenheiten in der Praxis ist dies aber keine unerfüllbare Voraussetzung. Im Großen und Ganzen präsentiert der in dieser Arbeit geleistete Beitrag die Bedeutung der besonders wichtigen DQ-Kriterien, deren Erfüllung in Datenbanken unabdingbar ist. Zusätzlich werden bei der Berücksichtigung der in dieser Arbeit ermittelten DQ-Kriterien Potentiale wie die Erhöhung der DQ geschaffen. Weiter erlauben die aufgestellten Kriterien den Grundstein für die Auslegung künftiger empirischer Arbeiten, da die relevanten Kriterien mithilfe dieser Arbeit im Vorfeld bestimmt worden sind und nur noch auf die praktische Anwendung warten.

## Literaturverzeichnis

- Ballou, D., Madnick, S. & Wang, R. (2004) Assuring information quality. *Journal of Management Information Systems*.
- Becker, T. & Geimer, H. (2001). *Supply Chain-Strategie. Supply Chain Management. Konzepte, Erfahrungsberichte und Strategien auf dem Weg zu digitalen Wertschöpfungsnetzen* (2. Aufl.). Springer Fachmedien Wiesbaden GmbH.
- Beckmann, H. (2004). *Supply Chain Management: Strategien und Entwicklungstendenzen in Spitzenunternehmen*. Springer-Verlag Berlin Heidelberg.
- Biedermann, L. & Kotzab, H. (2018). *Die strategische Bedeutung des Supply Chain Managements. Handbuch Produktions- und Logistikmanagement in Wertschöpfungsnetzwerken*. De Gruyter Oldenbourg.
- Bissantz N., Hagedorn J. & Mertens P. (1996). *Data-Mining als Komponente eines Data-Warehouse*. Gabler Verlag.
- Bodendorf, F. (2006). *Daten- und Wissensmanagement* (2. Aufl.), Springer-Verlag.
- Chen, P. P. (1976). The Entity-Relationship Model - Toward a Unified View of Data; *ACM Transactions on Database Systems*, 9-36.
- Corsten, H. & Gössinger, R. (2008). *Einführung in das Supply Chain Management*. (2. Aufl.). Oldenbourg Wissenschaftsverlag München.
- Dickersbach, J. T. (2009). *Supply Chain Management with SAP APO – Structures, Modelling Approaches and Implementation of SAP SCM 2008* (3. Aufl.). Springer.
- Dieker, S. & Güting, R. H. (2018). *Datenstrukturen und Algorithmen*. Springer Fachmedien Wiesbaden GmbH.
- DIN, Deutsches Institut für Normung e. V (1995). *Qualitätsmanagement, Statistik, Zertifizierung: Begriffe aus DIN-Normen* (2. Aufl.). Beuth Verlag GmbH.
- DIN, Deutsches Institut für Normung (1988). *Informationsverarbeitung Begriffe Teile 1-9*. Beuth Verlag GmbH.
- Ehrig, H., Mahr, B., Cornelius, F., Große-Rode, M., Zeitz, P. (2001): *Mathematisch-strukturelle Grundlagen der Informatik*. Springer-Verlag Berlin Heidelberg.
- English, L. P. (1999). *Improving Data Warehouse and Business Information Quality*. Wiley.
- Gebauer, M. & Windheuser, U. (2021). *Daten- und Informationsqualität. Die Grundlage der Digitalisierung* (5. Aufl.). Springer Fachmedien Wiesbaden GmbH.
- Geiger, W. & Kotte, W. (2008). *Handbuch Qualität. Grundlagen und Elemente des Qualitätsmanagements: Systeme – Perspektiven* (5. Aufl.). Vieweg und Teubner Verlag.
- Grimm, A. (2011). *Prozessorientierter Umgang mit Anforderungen für die kundenspezifische Auftragsabwicklung*. Springer Fachmedien Wiesbaden GmbH.
- Helfert, M. (2000). *Data-Warehousing-Strategie: Erfahrungen, Methoden*, Springer.

- Helfert, M. (2002). Messung und Planung der Datenqualität in Data-Warehouse-Systemen. Dissertation Universität St. Gallen. Difo-Druck GmbH.
- Hildebrand, K. (2004). Datenqualität im Supply Chain Management. Informatik 2004-Informatik verbindet-Band 1.
- Hildebrand, K. (2021). Daten- und Informationsqualität. Die Grundlage der Digitalisierung (5. Aufl.). Springer Fachmedien Wiesbaden GmbH.
- Hoppe, M. (2007). Absatz- und Bestandsplanung mit SAP APO. Galileo Press.
- Houlihan, J. B. (1985). International Supply Chain Management. International Journal of Physical Distribution and Materials Management
- Juran, J. M. (1999). Juran's Quality Handbook (5. Aufl.). McGraw Hill.
- Kamble, P. (2019). Datensatz Supply chain management for Car. verfügbar unter <https://www.kaggle.com/prashantk93/supply-chain-management-for-car> (Zugriff am 06.11.2021)
- Kahn, B. K., Strong, D. M. & Wang, R. Y. (2002): Information quality benchmarks: Product and service performance. Communications of the ACM.
- Köhler, J. & Oswald, A. (2009). Die Collective Mind Methode. Projekterfolg durch Soft Skills. Springer-Verlag Berlin Heidelberg.
- Kurbel, K. (2021). Enterprise Resource Planning und Supply Chain Management in der Industrie. De Gruyter Oldenbourg.
- Lange, O. & Stegemann, G. (1987). Datenstrukturen und Speichertechniken. Springer Fachmedien Wiesbaden GmbH.
- Lee, Y. W., Pipino, L. L. & Wang, R. Y. (2002): Data quality assessment. Communications of the ACM.
- Mehta, D. P. & Sahni, S. (2018). Handbook of data structures and applications (2. Aufl.). Chapman & Hall/CRC computer information science series. CRC Press Taylor & Francis Group.
- Mertens, P., Bodendorf, F. & König, W., Schumann, M., Hess, T. & Buxmann, P. (2017). Grundzüge der Wirtschaftsinformatik (12. Aufl.). Springer Gabler.
- Miller, H. (1996). The multiple dimensions of information quality. Information Systems Management, 13 (2), 79-82.
- Müller, J. (2000). Transformation operativer Daten zur Nutzung im Data Warehouse, Springer Fachmedien Wiesbaden GmbH.
- Otto, B. & Österle, H. (2016). Corporate Data Quality. Voraussetzung erfolgreicher Geschäftsmodelle. Springer Gabler.
- Piro, A. & Gebauer, M. (2021). Daten- und Informationsqualität. Die Grundlage der Digitalisierung (5. Aufl.). Springer Fachmedien Wiesbaden GmbH.
- Pollmeier, I. & Steven, M. (2008). Wertschöpfungsnetzwerke. Konzepte für das Netzwerkmanagement und Potenziale aktueller Informationstechnologien. Physica-Verlag.

- Redman, T. (1996): Data Quality for the information age, Artech House.
- Redman, T. (1998): The impact of poor data quality on the typical enterprise. Communications of the ACM.
- Rohweder, J. P., Kasten, G., Malzahn, D., Piro, A. & Schmid, J. (2021). Daten- und Informationsqualität (5. Aufl.). Springer Fachmedien Wiesbaden GmbH.
- Ryu, K. S., Park, J. H. & Park, J. S. (2006): A data quality management maturity model. ETRI Journal.
- Saake, G. & Sattler, K. U. (2021). Algorithmen und Datenstrukturen. Eine Einführung mit Java. dpunkt.verlag GmbH.
- Schäfer, G. (1989): Datenstrukturen und Datenbanken. Verlag Vieweg.
- Scheidler, A. A. (2017): Methode zur Erschließung von Wissen aus Datenmustern in Supply-Chain-Datenbanken. Cuvillier Verlag.
- Schuh, G. & Stich, V. (2012). Produktionsplanung und -steuerung 1. Grundlagen der PPS (4. Aufl.). Springer Vieweg.
- Schulte, C. (2016). Logistik. Wege zur Optimierung der Supply Chain (7. Aufl.). Verlag Franz Vahlen München.
- Schwemm, J. W. (2009). Zwischenbetriebliches Stammdatenmanagement: Lösung für die Datensynchronisation zwischen Handel und Konsumgüterindustrie. Springer Verlag.
- Simchi-Levi, D., Kaminsky, P. & Simchi-Levi, E. (2008). Designing and Managing the Supply Chain: Concepts, Strategies and Case Studies. McGraw-Hill Irwin.
- Stahlknecht, P. (1995). Einführung in die Wirtschaftsinformatik. Springer-Verlag Berlin Heidelberg.
- Steyer, R. (2019). Webanwendungen erstellen mit Vue.js. Springer Fachmedien Wiesbaden.
- Strong, D. M. & Wang, R. Y. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems, 12 (4), 5-33.
- Wannenwetsch, H. (2021). Integrierte Materialwirtschaft, Logistik, Beschaffung und Produktion. Supply Chain im Zeitalter der Digitalisierung (6. Aufl.), Springer Verlag GmbH Deutschland.
- Wittmann, W. (1959). Unternehmung und unvollkommene Information. Unternehmerische Voraussicht. Ungewissheit und Planung, Westdeutscher Verlag.
- Wolf, J. (2021). Daten- und Informationsqualität (5. Aufl.). Springer Fachmedien Wiesbaden GmbH.
- Wood, D. C. (2007). SAP SCM – Applications and Modeling for Supply Chain Management. Wiley.

# 5. Anhang

## Anhang A: Kriterien aller Werke

	Untersuchte Werke (1 - 6)					
	Wang und Strong (1996)	English (1999)	Redman (1996)	Müller (2000)	Miller (1996)	Rohweder et al. (2021)
<b>Kriterien</b>	Glaubwürdigkeit	Übereinstimmung mit der Datendef.	Relevanz	Technisch. Verfügbarkeit der Daten	Relevanz	Zugänglichkeit
	Genauigkeit	Vollständigkeit der Datenwerte	Zugriff	der Daten	Genauigkeit	Bearbeitbarkeit
	Objektivität	plausibel u. mit den Geschäftsregeln	Klarheit der Definitionen	Technisch. Verwendbarkeit der Daten	Aktualität	Hohes Ansehen
	Vertrauenswürdigkeit	übereinstimmend	Vollständigkeit	der Daten	Vollständigkeit	Fehlerfreiheit
	Zusatznutzen	Genauigkeit zu Referenzdaten	Wesentlich	Datensicherheit	Stimmigkeit	Objektivität
	Relevanz	Genauigkeit zur Realität	Attributgranularität	Einheitlichkeit von Formaten	Format	Glaubwürdigkeit
	Aktualität	Granularität der Datenwerte	Domaingenauigkeit	Präzision	Zugänglichkeit	eindeutige Auslegbarkeit
	Vollständigkeit	Eindeutigkeit (keine Duplikate)	Natürlich	Detailliertheit	Kompatibilität	einheitliche Darstellung
	Angemessenes Datenvolumen	Übereinstimmung von redundanten und verteilten Datenbeständen	Identifizierbar	Validität	Sicherheit	Übersichtlichkeit
	Interpretierbarkeit		Homogen	Quantifizierbarkeit	Gültigkeit	Verständlichkeit
	Verständlichkeit	Zeitliche Übereinstimmung von "	Minimum an Redundanz	Vertrauenswürdigkeit		Relevanz
	Konsistente Darstellung	Allgemeine Zugriffsfähigkeit	Semantische Konsistenz	Fehlerfreiheit		Angemessener Umfang
	Knappe Darstellung	Aktualität	Strukturelle Konsistenz	Konsistenz		Vollständigkeit
	Zugriffsmöglichkeit	Klarheit für den Datenanwender	Stabil (Reaktionen auf Veränderungen)	Zeitliche Eignung für den gegeb. Zweck		Wertschöpfung
	Zugriffssicherheit		Flexibel (")			Aktualität
		Übereinstimmung zwischen abgeleit. Daten zu den Ursprungsdaten	Korrekt	Sachliche Eignung für den gegeb. Zweck		
		Nützlichkeit	Vollständig (Entitäten/Attribute)	den gegeb. Zweck		
	Vollständigkeit bezogen auf den Informationsbedarf	Konsistenz	Vollständigkeit der Daten für den gegeb. Zweck			
	Pünktlichkeit	Aktualität	den gegeb. Zweck			
	Interpretierbarkeit für den Datenanwender	Angemessen	Prüfbarkeit			
		Formatgenauigkeit	Einheitlichkeit von Darstellungsformen			
		Effiziente Speichernutzung				
		Interpretierbarkeit				
		Formatflexibilität				
		Übertragbarkeit				
		Fähigkeit Nullwerte abzubilden				
		Darstellungskonsistenz				

Anhang B: Eliminierung gleicher Kriterien

Kriterien	Untersuche Werke (1 - 6)					
	Wang und Strong (1996)	English (1999)	Redman (1996)	Müller (2000)	Miller (1996)	Rohweder et al. (2021)
Glaubwürdigkeit	Übereinstimmung mit der Datendef.	Relevanz	Technisch. Verfügbarkeit	Relevanz	Zugänglichkeit	
Genauigkeit	Vollständigkeit der Datenwerte	Zugriff	der Daten	Genauigkeit	Bearbeitbarkeit	
Objektivität	plausibel u. mit den Geschäftsregeln	Klarheit der Definitionen	Technisch. Verwendbarkeit	Aktualität	Hohes Ansehen	
Vertrauenswürdigkeit	übereinstimmend	Vollständigkeit	der Daten	Vollständigkeit	Fehlerfreiheit	
Zusatznutzen	Genauigkeit zu Referenzdaten	Wesentlich	Datensicherheit	Konsistenz	Objektivität	
Relevanz	Genauigkeit zur Realität	Attributgranularität	Einheitlichkeit von Formaten	Format	Glaubwürdigkeit	
Aktualität	Granularität der Datenwerte	Domaingenauigkeit	Präzision	Zugänglichkeit	eindeutige Auslegbarkeit	
Vollständigkeit	Eindeutigkeit (keine Duplikate)	Natürlich	Detailliertheit	Kompatibilität	einheitliche Darstellung	
Angemessenes Datenvolumen	Übereinstimmung von redundanten	Identifizierbar	Validität	Sicherheit	Übersichtlichkeit	
Interpretierbarkeit	und verteilten Datenbeständen	Homogen	Quantifizierbarkeit	Gültigkeit	Verständlichkeit	
Verständlichkeit	Zeitliche Übereinstimmung von "	Minimum an Redundanz	Vertrauenswürdigkeit		Relevanz	
Konsistente Darstellung	Allgemeine Zugriffsfähigkeit	Semantische Konsistenz	Fehlerfreiheit		Angemessener Umfang	
Knappe Darstellung	Aktualität	Strukturelle Konsistenz	Konsistenz		Vollständigkeit	
Zugriffsmöglichkeit	Klarheit für den	Stabil (Reaktionen auf Veränderungen)	Zeitliche Eignung für		Wertschöpfung	
Zugriffssicherheit	Datenanwender	Flexibel (")	den gegeb. Zweck		Aktualität	
	Übereinstimmung zwischen abgeleit.	Korrekt	Sachliche Eignung für			
	Daten zu den Ursprungsdaten	Vollständig (Entitäten/Attribute)	den gegeb. Zweck			
	Nützlichkeit	Konsistenz	Vollständigkeit der Daten für			
	Vollständigkeit bezogen auf den In-	Aktualität	den gegeb. Zweck			
	formationsbedarf	Angemessen	Prüfbarkeit			
	Pünktlichkeit	Formatgenauigkeit	Einheitlichkeit von Darstellungs-			
	Interpretierbarkeit für den	Effiziente Speichernutzung	formen			
	Datenanwender	Interpretierbarkeit				
		Formatflexibilität				
		Übertragbarkeit				
		Fähigkeit Nullwerte abzubilden				
		Darstellungskonsistenz				



Anhang C: Häufigkeitstabelle ohne Synonyme (Teil 1)

Kriterium	Werke						Häufigkeit
	1	2	3	4	5	6	
Glaubwürdigkeit	1					1	2
Genauigkeit	1				1		2
Objektivität	1					1	2
Vertrauenswürdigkeit	1			1			2
Relevanz	1		1		1	1	4
Aktualität	1	1	1		1	1	5
Vollständigkeit	1		1		1	1	4
Interpretierbarkeit	1		1				2
Verständlichkeit	1					1	2
Konsistenz			1	1	1		3
Zugänglichkeit					1	1	2
Konsistente Darstellung	1						1
Knappe Darstellung	1						1
Zugriffsmöglichkeit	1						1
Zugriffssicherheit	1						1
Übereinstimmung mit der Datendef.		1					1
Vollständigkeit der Datenwerte		1					1
plausibel u. mit den Geschäftsregeln übereinstimmend		1					1
Genauigkeit zu Referenzdaten		1					1
Genauigkeit zur Realität		1					1
Granularität der Datenwerte		1					1
Eindeutigkeit (keine Duplikate)		1					1
Übereinstimmung von redundanten und verteilten Datenbeständen		1					1
Zeitliche Übereinstimmung von "		1					1
Allgemeine Zugriffsfähigkeit		1					1
Klarheit/Interpretierbarkeit für den Datenanwender		1					1
Übereinstimmung zwischen abgeleiteten Daten zu den Ursprungsdaten		1					1
Nützlichkeit		1					1
Vollständigkeit bezogen auf den In- formationsbedarf		1					1
Pünktlichkeit		1					1
Zugriff			1				1
Klarheit der Definitionen			1				1
Wesentlich			1				1
Attributgranularität			1				1
Domaingenauigkeit			1				1
Natürlich			1				1
Identifizierbar			1				1
Homogen			1				1

Anhang D: Häufigkeitstabelle ohne Synonyme (Teil 2)

Minimum an Redundanz			1				1
Semantische Konsistenz			1				1
Strukturelle Konsistenz			1				1
Stabil (Reaktionen auf Veränderungen)			1				1
Flexibel (")			1				1
Korrekt			1				1
Vollständig (Entitäten/Attribute)			1				1
Angemessen			1				1
Formatgenauigkeit			1				1
Effiziente Speichernutzung			1				1
Formatflexibilität			1				1
Übertragbarkeit			1				1
Fähigkeit Nullwerte abzubilden			1				1
Darstellungskonsistenz			1				1
Technisch. Verfügbarkeit der Daten				1			1
Technisch. Verwendbarkeit der Daten				1			1
Datensicherheit				1			1
Einheitlichkeit von Formaten und Darstellungsformen				1			1
Präzision				1			1
Detailliertheit				1			1
Validität				1			1
Quantifizierbarkeit				1			1
Fehlerfreiheit				1			1
Zeitliche Eignung für den gegeb. Zweck				1			1
Sachliche Eignung für den gegeb. Zweck				1			1
Vollständigkeit der Daten für den gegeb. Zweck				1			1
Prüfbarkeit				1			1
Format					1		1
Kompatibilität					1		1
Sicherheit					1		1
Gültigkeit					1		1
Bearbeitbarkeit						1	1
Hohes Ansehen						1	1
Fehlerfreiheit						1	1
eindeutige Auslegbarkeit						1	1
einheitliche Darstellung						1	1
Übersichtlichkeit						1	1
Angemessener Umfang						1	1
Wertschöpfung						1	1

Anhang E: Kriterien mit der gleichen Bedeutung

Kriterien	Untersuche Werke (1 - 6)					
	Wang und Strong (1996)	English (1999)	Redman (1996)	Müller (2000)	Miller (1996)	Rohweder et al. (2021)
Glaubwürdigkeit	Übereinstimmung mit der Datendef.	Relevanz	Technisch. Verfügbarkeit der Daten	Relevanz	Zugänglichkeit	
Genauigkeit	Vollständigkeit der Datenwerte	Zugriff	Technisch. Verwendbarkeit der Daten	Genauigkeit	Bearbeitbarkeit	
Objektivität	plausibel u. mit den Geschäftsregeln übereinstimmend	Klarheit der Definitionen	Technisch. Verwendbarkeit der Daten	Aktualität	Hohes Ansehen	
Vertrauenswürdigkeit	Genauigkeit zu Referenzdaten	Vollständigkeit	Datensicherheit	Vollständigkeit	Fehlerfreiheit	
Zusatznutzen	Genauigkeit zur Realität	Wesentlich	Einheitlichkeit von Formaten	Konsistenz	Objektivität	
Relevanz	Granularität der Datenwerte	Attributgranularität	Format	Format	Glaubwürdigkeit	
Aktualität	Eindeutigkeit (keine Duplikate)	Domaingenauigkeit	Präzision	Zugänglichkeit	eindeutige Auslegbarkeit	
Vollständigkeit	Eindeutigkeit (keine Duplikate)	Natürlich	Detailliertheit	Kompatibilität	einheitliche Darstellung	
Angemessenes Datenvolumen	Übereinstimmung von redundanten und verteilten Datenbeständen	Identifizierbar	Validität	Sicherheit	Übersichtlichkeit	
Interpretierbarkeit	Zeitliche Übereinstimmung von "	Homogen	Quantifizierbarkeit	Gültigkeit	Verständlichkeit	
Verständlichkeit	Allgemeine Zugriffsfähigkeit	Minimum an Redundanz	Vertrauenswürdigkeit		Relevanz	
Konsistente Darstellung	Aktualität	Semantische Konsistenz	Fehlerfreiheit		Angemessener Umfang	
Knappe Darstellung	Klarheit für den Datenanwender	Strukturelle Konsistenz	Konsistenz		Vollständigkeit	
Zugriffsmöglichkeit	Übereinstimmung zwischen abgeleit. Daten zu den Ursprungsdaten	Stabil (Reaktionen auf Veränderungen)	Zeitliche Eignung für den gegeb. Zweck		Wertschöpfung	
Zugriffssicherheit	Nützlichkeit	Flexibel (")	Sachliche Eignung für den gegeb. Zweck		Aktualität	
	Vollständigkeit bezogen auf den Informationsbedarf	Korrekt	Vollständig (Entitäten/Attribute)			
	Pünktlichkeit	Konsistenz	Vollständigkeit der Daten für den gegeb. Zweck			
	Interpretierbarkeit für den Datenanwender	Aktualität	Prüfbarkeit			
		Angemessen	Einheitlichkeit von Darstellungsformen			
		Formatgenauigkeit				
		Effiziente Speichernutzung				
		Interpretierbarkeit				
		Formalflexibilität				
		Übertragbarkeit				
		Fähigkeit Nullwerte abzubilden				
		Darstellungskonsistenz				

Anhang F: Häufigkeitstabelle gleichbedeutender DQ-Kriterien

Nr.	Kriterium	Werke						Häufigkeit
		1	2	3	4	5	6	
1	Genauigkeit	1	1	1	1	1	1	6
2	Vollständigkeit	1	1	1	1	1	1	6
3	Relevanz	1	1	1	1	1	1	6
4	Aktualität	1	1	1	1	1	1	6
5	Zugänglichkeit	1	1	1	1	1	1	6
6	Konsistenz	1	1	1	1	1	1	5
7	Verständlichkeit	1	1	1			1	4
8	Format			1	1	1	1	4
9	Zugriffssicherheit	1			1	1		3
10	Interpretierbarkeit	1	1	1				3
11	Angemessener Umfang	1		1			1	3
12	Einheitliche Darstellung			1	1		1	3
13	Vertrauenswürdigkeit	1			1		1	3
14	Glaubwürdigkeit	1					1	2
15	Minimum an Redundanz		1	1				2
16	Objektivität	1					1	2
17	Kompatibilität		1			1		2
18	Bearbeitbarkeit				1		1	2
19	Granularität (der Datenwerte)		1	1				2
20	Gültigkeit				1	1		2
21	Wertschöpfung						1	1
22	Eindeutige Auslegbarkeit						1	1
23	Quantifizierbarkeit				1			1
24	Domaingenauigkeit			1				1
25	Genauigkeit zu Referenzdaten		1					1
26	Zeitliche Übereinstimmung von redundant. Und verteilt. Datenbest.		1					1
27	Pünktlichkeit		1					1
28	Natürlich			1				1
29	Identifizierbar			1				1
30	Stabil			1				1
31	Flexibel			1				1
32	Effiziente Speichernutzung			1				1
33	Formatflexibilität			1				1
34	Übertragbarkeit			1				1
35	Fähigkeit Nullwerte abzubilden			1				1