

## **Bachelorarbeit**

Zur Erlangung des akademischen Grades

Bachelor of Science (B. SC.)

# **Data Driven - Untersuchung datengetriebener Prozesse im Umfeld von Produktion und Logistik**

Verfasst von

Lars Heimeshoff

Studiengang: Maschinenbau

Matrikel-Nr.: 134353

Erstbetreuerin: Dr.-Ing. Dipl.-Inform. Anne Antonia Scheidler

Zweitbetreuer: M. Sc. Sean Paul Christiansen-Lenger

Ausgegeben am: 23. Juni 2021

Eingereicht am: 14. September 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 EINSATZ VON DATEN IM UMFELD VON PRODUKTION UND LOGISTIK .....</b>	<b>3</b>
2.1 Daten im Kontext von Produktion und Logistik .....	3
2.2 Informationstechnische Prozesse in Produktion und Logistik .....	6
2.3 Einführung des Begriffes Industrie 4.0 .....	10
2.4 Einführung des Begriffes Big Data .....	13
2.5 Einführung des Begriffes Data Driven .....	15
2.6 Methoden der Literaturanalyse und Analysekriterien zur Begriffsfindung.....	17
<b>3 UNTERSUCHUNG DES BEGRIFFES DATA DRIVEN ZUR ENTWICKLUNG EINES DEFINITIONSVORSCHLAGES .....</b>	<b>21</b>
3.1 Literaturanalyse zu datengetriebenen Prozessen .....	21
3.1.1 Methodenauswahl und -anpassung der Literaturanalyse und Analysekriterien ..	22
3.1.2 Durchführung und Ergebnisdarstellung der Literaturanalyse zu datengetriebenen Prozessen .....	27
3.1.3 Kernaussagen zu datengetriebenen Prozessen aus der Literaturanalyse .....	40
3.2 Literaturanalysegestützte Ableitung von Begriffsmerkmalen datengetriebener Prozesse.....	47
3.3 Sichtweisen verschiedener Fachdisziplinen auf den Begriff Data Driven .....	51
3.4 Abgrenzung von Data Driven zu anderen Begriffen .....	52
3.4.1 Abgrenzung zu etablierten informationstechnischen Prozessen.....	52
3.4.2 Abgrenzung zu Industrie 4.0.....	54
3.4.3 Abgrenzung zu Big Data .....	55
3.5 Diskussion und Definitionsvorschlag für den Begriff Data Driven .....	57
<b>4 EVALUIERUNG DER BEGRIFFSKRITERIEN UND DES DEFINITIONSVORSCHLAGES .....</b>	<b>60</b>

---

4.1 Evaluierungskonzept .....	60
4.2 Einführung von Beispielprozessen .....	62
4.3 Durchführung der Evaluierung .....	63
4.4 Bewertung der Begriffskriterien und des Definitionsvorschlages.....	70
<b>5 ZUSAMMENFASSUNG UND AUSBLICK .....</b>	<b>73</b>
<b>LITERATURVERZEICHNIS .....</b>	<b>77</b>
<b>ANHANG .....</b>	<b>83</b>
Anhang 1: Liste des digitalen Anhangs .....	83
Anhang 2: Ergänzende Tabellen.....	83
<b>EIDESSTATTLICHE VERSICHERUNG .....</b>	<b>92</b>

## Abbildungsverzeichnis

Abbildung 1: 4 Stufen der Wissenstreppe .....	4
Abbildung 2: Klassifikation von Daten nach der Darstellungsform .....	5
Abbildung 3: Darstellung eines Einzelprozesses.....	7
Abbildung 4: Ebenen der betrieblichen IT-Systemlandschaft.....	8
Abbildung 5: Fraunhofer-Schalenmodell der Industrie 4.0-Wertschöpfung.....	11
Abbildung 6: Beziehungen von KI, ML und DL.....	12
Abbildung 7: Die 5 Vs von Big Data .....	14
Abbildung 8: Schematische Darstellung des Hirsch-Indexes.....	20
Abbildung 9: Word-Cloud Keywords .....	33
Abbildung 10: Word-Cloud Methoden und Technologien.....	34
Abbildung 11: Diagramm Datenerfassung .....	35
Abbildung 12: Diagramm Datenstruktur .....	36
Abbildung 13: Diagramm Aufgabenorientierung.....	38
Abbildung 14: Diagramm Analysesystem.....	39
Abbildung 15: Beispielmodell aktueller datengetriebener Prozesse .....	45
Abbildung 16: Evaluierungsmatrix .....	61

## Tabellenverzeichnis

Tabelle 1: Quellen der Literaturanalyse .....	28
Tabelle 2: die sechs relevantesten Quellen.....	31
Tabelle 3: Informale Kriterien Daten .....	32
Tabelle 4: Informale Kriterien Verwendung & Methoden und Technologien.....	32
Tabelle 5: Informale Kriterien Eingrenzung & Keywords .....	32
Tabelle 6: Kernaussagen der sechs relevantesten Quellen .....	40
Tabelle 7: Morphologischer Kasten aktueller datengetriebener Prozesse.....	48
Tabelle 8: Abgrenzung zu etablierten IT-Prozessen.....	52
Tabelle 9: Abgrenzung zu Industrie 4.0 .....	54
Tabelle 10: Abgrenzung zu Big Data .....	55
Tabelle 11: Morphologischer Kasten Beispielprozess A .....	63
Tabelle 12: Morphologischer Kasten Beispielprozess B.....	66
Tabelle 13: Morphologischer Kasten Beispielprozess C.....	68
Tabelle A. 1: Ranking aller Quellen.....	83
Tabelle A. 2: Definitionsvorschlag Herleitungstabelle .....	86

## Abkürzungsverzeichnis

AutoID	Automatische Identifizierung und Datenerfassung
BDA	Big Data Analytics
DDPD	Data Driven Process Design
DL	Deep Learning
DMS	Dokumentenmanagement System
DTM	Data Driven Transport Management
EDV	Elektronische Datenverarbeitung
ERP	Enterprise Resource Planning
ETL	Extraktion, Transformation, Laden
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
IT	Informationstechnik
JIF	Journal Impact Factor
KDD	Knowledge Discovery in Databases
KI	Künstliche Intelligenz
MDM	Master Data Management
ML	Machine Learning
OLAP	Online Analytical Processing
OLTP	Online Transaction Processing
RFID	Radio Frequency Identification
SC	Supply Chain (dt. Lieferkette)
SCM	Supply Chain Management
SPS	Speicherprogrammierbare Steuerung
TUL-Logistik	Transport-, Umschlag- und Lagerlogistik

# 1 Einleitung

Gegenstand dieser Arbeit ist eine Untersuchung informationstechnischer datengetriebener Prozesse im Hinblick auf die Verwendung des Begriffes Data Driven im Kontext verschiedener Fachdisziplinen. Trotz der durch die fortschreitende Digitalisierung bedingten Zunahme datengetriebener Prozesse, existiert bis in das Jahr 2021 keine eindeutige Definition für den Begriff Data Driven, beziehungsweise für den Begriff datengetrieben, welcher in der deutschsprachigen Literatur und daher auch im Zuge dieser Arbeit meist äquivalent verwendet wird. Aktuell wird der Begriff im Zusammenhang mit Prozessen aus unterschiedlichen Fachdisziplinen genannt. Um einen inhaltlichen Schwerpunkt zu setzen, beschränkt sich diese Arbeit hauptsächlich auf die Begriffsbestimmung im Kontext informationstechnischer Prozesse in Produktion und Logistik.

Die stetig wachsende Begriffswelt der Digitalisierung erschwert es zunehmend, neue Begriffe einzuordnen und von weiteren Begriffen zu differenzieren, welche ähnliche oder gleiche Ansätze verfolgen Aufgaben zu erfüllen. Für die Forschung ist eine möglichst exakte Formulierung von Begriffen jedoch elementar, um subjektive Aussagen zu vermeiden (Pietsch, 2021). Insbesondere im Bereich von Produktion und Logistik ist vor dem Hintergrund der interdisziplinären Zusammenarbeit eine einheitliche Fachkommunikation wichtig, um Missverständnissen vorzubeugen. Aus diesen Gründen ist es bedeutsam entweder eine allgemeingültige Definition für den Begriff Data Driven zu finden oder zumindest dessen Bedeutung in unterschiedlichen Fachdisziplinen klar voneinander abzugrenzen.

Diese Arbeit verfolgt das Ziel, Merkmale datengetriebener Prozesse zu erarbeiten, um diese möglichst genau von anderen Begriffen wie zum Beispiel Big Data oder Industrie 4.0 und von bereits etablierten informationstechnischen Lösungen (IT-Lösungen) in Produktion und Logistik abzugrenzen. Im Anschluss wird mit Hilfe der zuvor erarbeiteten Merkmale ein eigener Definitionsvorschlag für den Begriff Data Driven entwickelt. Zukünftig könnten die Merkmale und der Definitionsvorschlag genutzt werden, um datengetriebene Prozesse zu identifizieren oder um abweichende Prozesse von datengetriebenen Prozessen zu differenzieren. Des Weiteren könnte ein einheitlicher Definitionsvorschlag zu einem allgemeinen Begriffsverständnis beitragen, um eine einheitliche Kommunikation zu erleichtern.

Um dieses Ziel zu erreichen, teilt sich die Arbeit im Wesentlichen in zwei Abschnitte auf. Im ersten Teil werden in einer Literaturrecherche Aussagen zu datengetriebenen Prozessen und angrenzenden Themengebieten und deren Kernaussagen herausgearbeitet. Aufbauend darauf werden Merkmale datengetriebener Prozesse erarbeitet und auf dieser Grundlage der Begriff Data Driven zu anderen Begriffen abgegrenzt. Am Ende des ersten Teils steht der Versuch einen eigenen Definitionsvorschlag für den Begriff Data Driven zu entwickeln. Im zweiten Abschnitt der Arbeit werden die zuvor gewonnen Ergebnisse anhand von Beispiel-

prozessen evaluiert und abschließend hinsichtlich eventueller Einsatzmöglichkeiten und Verbesserungspotentiale bewertet.

Als Ausgangspunkt für die weitere Arbeit werden zunächst im zweiten Kapitel die Grundlagen auf dem Gebiet der Datennutzung in der Produktion und Logistik vermittelt, um aufbauend auf diesem theoretischen Wissen erste mögliche Nutzenpotentiale und Anforderungen datengetriebener Prozesse zu identifizieren. Zudem werden am Ende des zweiten Kapitels geeignete Methoden für die spätere Kriterienbildung für die Bestimmung des Begriffes Data Driven aus der Literatur eruiert (Abschnitt 2.6). Das dritte Kapitel verfolgt das Ziel Merkmale datengetriebener Prozesse zu erarbeiten und einen eigenen Definitionsvorschlag für den Begriff Data Driven zu entwickeln. Dazu werden, unter der Verwendung einer geeigneten Methode, in einer Literaturanalyse Aussagen verschiedener Autor\*innen zu datengetriebenen Prozessen in der Literatur gesichtet. In einem weiteren Schritt werden die Kernaussagen zum Thema datengetriebene Prozesse zusammengefasst und tabellarisch dargestellt (Abschnitt 3.1). Mit Hilfe der Tabelle und den in der Literaturanalyse erarbeiteten Analyse Kriterien werden mögliche Begriffsmerkmale datengestützter Prozesse identifiziert (Abschnitt 3.2). Im Folgenden werden unterschiedliche Perspektiven verschiedener Fachdisziplinen auf den Begriff Data Driven diskutiert (Abschnitt 3.3). Daraufhin werden die Ergebnisse der Literaturanalyse und der Erarbeitung der Merkmale genutzt, um auf Grundlage dieser den Begriff Data Driven von etablierten informationstechnischen Prozessen in Produktion und Logistik und von den Begriffen Industrie 4.0 und Big Data abzugrenzen (Abschnitt 3.4). Im Anschluss steht eine Diskussion der zuvor gewonnenen Erkenntnisse, welche letzten Endes in dem Versuch mündet einen eigenen Definitionsvorschlag für den Begriff Data Driven zu entwickeln (Abschnitt 3.5). Um die Ergebnisse der Arbeit zu bewerten, werden diese im vierten Kapitel evaluiert. Hierzu wird zunächst ein Evaluierungskonzept erarbeitet und das genaue Vorgehen der Evaluierung erläutert (Abschnitt 4.1). Danach werden Beispielprozesse vorgestellt (Abschnitt 4.2) mit deren Hilfe die Evaluierung, der in Kapitel drei abgeleiteten Kriterien durchgeführt wird (Abschnitt 4.3). Zum Schluss werden die Ergebnisse der Arbeit auf Grundlage der vorangegangenen Evaluierung bewertet (Abschnitt 4.4). Das letzte Kapitel (Abschnitt 5) beinhaltet eine Zusammenfassung der in der Arbeit gewonnenen Erkenntnisse und schließt mit einem Ausblick.



## 2 Einsatz von Daten im Umfeld von Produktion und Logistik

Gerade in wettbewerbsintensiven und stark fluktuierenden Märkten sind Daten unverzichtbar. Sogenannte Tech-Unternehmen wie Facebook, Amazon oder Google sammeln große Datenmengen, um beispielsweise das Kundenverhalten genau zu analysieren. Auch im Bereich von Produktion und Logistik bietet die Verwendung von Daten große Potentiale und kann beispielsweise genutzt werden, um Materialflüsse zu optimieren oder auch um neue Geschäftsmodelle zu erschließen. Um ein besseres Verständnis für datengetriebene Prozesse im Umfeld von Produktion und Logistik zu entwickeln und darüber hinaus unterschiedliche Auffassungen des Begriffes Data Driven näher betrachten zu können, wird am Anfang dieses Kapitels grundlegendes zum Datenbegriff erläutert (Abschnitt 2.1). Infolgedessen werden die Begriffe Daten, Informationen und Wissen klar voneinander abgegrenzt und eine Definition von Daten wird für den weiteren Verlauf der Arbeit festgelegt. Im Anschluss werden unterschiedliche Einsatzmöglichkeiten der zuvor definierten Daten im Umfeld von Produktion und Logistik vermittelt. Zunächst werden etablierte informationstechnische Prozesse in Produktion und Logistik beleuchtet (Abschnitt 2.2). Zu diesem Zweck wird eingangs erläutert, was informationstechnische Prozesse ausmacht, um nach einer kurzen Einführung der Bereiche Produktion und Logistik die häufigste Nutzung von Daten in diesen Bereichen zu spezifizieren. Um im späteren Verlauf der Arbeit Data Driven von anderen Begriffen abgrenzen zu können und um weitere Potentiale der Datennutzung abseits der etablierten Prozesse darzulegen, folgt im Abschnitt 2.3 eine Erläuterung zur Industrie 4.0. Unter anderem widmet sich die Industrie 4.0 der Erfassung und der Nutzung großer Datenmengen. Der Begriff, welcher in diesem Bereich am häufigsten genannt wird, ist Big Data, welcher im Abschnitt 2.4 eingeführt wird. Auch datengetriebene Prozesse verwenden größere Mengen an Daten. Um einen Eindruck über den Einsatz datengetriebener Prozesse zu bekommen und um Ansätze für die spätere Analyse aufzustellen, werden folgend diese Prozesse kurz betrachtet (Abschnitt 2.5). Im letzten Teil dieses Kapitels werden Methoden aus der Literatur eruiert, welche bei der literaturanalytischen Untersuchung des Begriffes Data Driven herangezogen werden (Abschnitt 2.6).

### 2.1 Daten im Kontext von Produktion und Logistik

Bei den Prozessen, welche in dieser Arbeit beschrieben werden, handelt es sich ausschließlich um informationstechnische Prozesse. Die Basis solcher Prozesse bilden Daten. Um die Wirkungsweise informationstechnischer Prozesse besser zu verstehen, eventuelle Unterschiede in den Auffassungen datengetriebener Prozesse nachvollziehen zu können und um für den weiteren Verlauf der Arbeit einen gemeinsamen Begriffsapparat zu schaffen, wird im Folgenden der Begriff Daten näher erläutert. Hierzu wird zuerst der Datenbegriff von den verwandten Begriffen Information und Wissen abgegrenzt. Danach wird kurz auf die Problematik einer

einheitlichen Formulierung des Datenbegriffes eingegangen und für den weiteren Verlauf der Arbeit eine Definition für den Datenbegriff festgelegt.

Häufig werden im alltäglichen Sprachgebrauch die Begriffe Daten, Informationen und Wissen synonym verwendet. Diese stehen zwar zueinander in Beziehung, sind jedoch klar voneinander zu differenzieren, um eine unmissverständliche Kommunikation zu ermöglichen. Meist werden die Begriffe einer hierarchischen oder linearen Gliederung unterstellt um deren Zusammenhänge und Unterschiede genauer zu verdeutlichen (Notté, 2013). Solch eine Darstellung in hierarchischer Form wird unter anderem durch die Wissenstreppe nach North abgebildet. Die ersten 4 Stufen der Wissenstreppe (Abbildung 1) veranschaulichen das Zusammenspiel und die Unterschiede von Daten, Informationen und Wissen.

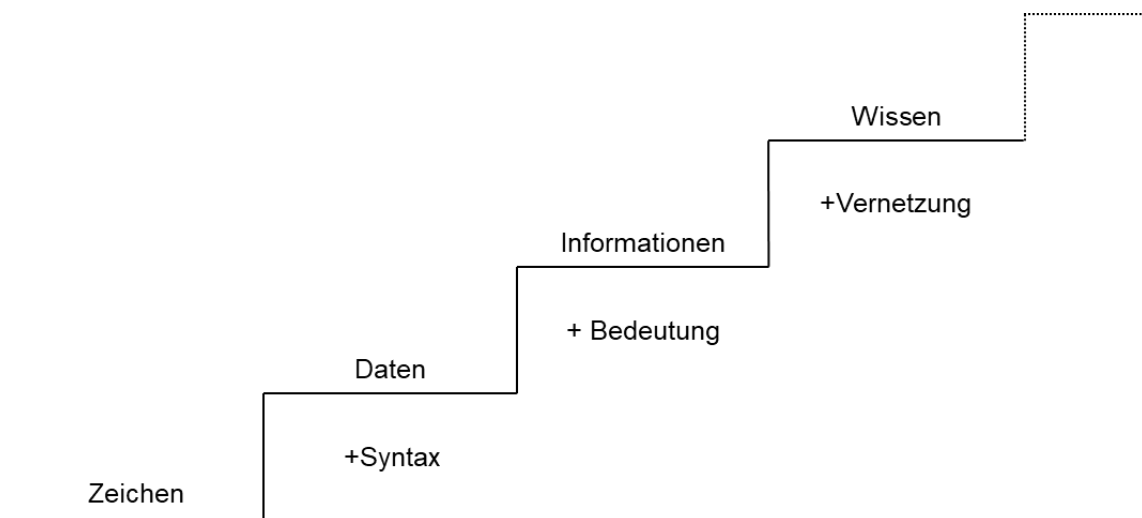


Abbildung 1: 4 Stufen der Wissenstreppe (North, 2021, S. 37)

**Daten** beschreibt North als vereinzelte oder eine Anreihung von Zeichen, welche bestimmten Ordnungsregeln (einer Syntax) unterworfen sind. Die Erscheinungsform der Zeichen ist dabei nicht von Bedeutung und kann Zahlen, Buchstaben, Symbole und andere beinhalten (North, 2021). **Informationen** sind Daten, denen eine bestimmte Bedeutung zugesprochen wird (North, 2021). Somit bezeichnen Informationen Inhalte, welche Sachverhalte und Vorgänge beschreiben. Diese Inhalte können sowohl strategisch als auch operativ bedeutsam sein (Lassmann, W. et al. 2006). **Wissen** entsteht aus der Vernetzung unterschiedlicher Informationen und dient der Erfüllung einer bestimmten Absicht. Somit ist Wissen das Ergebnis eines Lernvorganges, welches aus dem Rohstoff Information gewonnen aber auch in dieser Form gespeichert und kommuniziert wird (North, 2021). Prozesse welche nicht nur Daten und Informationen verarbeiten, sondern in der Lage sind durch die Vernetzung von Informationen Wissen zu generieren bieten somit einen klaren Vorteil, um beispielsweise schnell Handlungen auf Grundlage des erworbenen Wissens einzuleiten.

Aber nicht nur bei der Differenzierung von Daten, Informationen und Wissen herrschen im alltäglichen Gebrauch Unschärfen. Auch im Kontext verschiedener Fachdisziplinen wird der Datenbegriff oft unterschiedlich gehandhabt (Dornseif und Schumann, 2002). So werden zum Beispiel alleine im Duden vier unterschiedliche Bedeutungen von Daten gelistet (Duden online, 2021). Diese Unschärfe kann gerade in jenen Bereichen zu Problemen führen, in denen eine interdisziplinäre Fachkommunikation gefragt ist wie beispielsweise in den Rechtswissenschaften (Dornseif und Schumann, 2002) oder im Bereich des Industrial Data Science. Aus diesem Grund wird im Folgenden eine Definition des Begriffes Daten für den weiteren Verlauf der Arbeit festgelegt.

Da es sich bei datengetriebenen Prozessen in Produktion und Logistik hauptsächlich um informationstechnische Prozesse handelt, wird in dieser Arbeit der **Datenbegriff der Informatik** und der elektronischen Datenverarbeitung (EDV) genutzt. The International Organization for Standardization (ISO) und the International Electrotechnical Commission (IEC) definieren Daten als „*reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing*” - eine wieder interpretierbare Darstellung von Informationen in formalisierter Art, geeignet zur Kommunikation, Interpretation oder Verarbeitung (ISO/IEC 2382). Diese Definition von Daten entspricht auch der Beschreibung von North, welche durch seine Wissenstreppe (Abbildung 1) veranschaulicht wird. Demnach sind Daten einer bestimmten Syntax (einem Code) folgende Anreihung von Zeichen, welche in ihrer vereinbarten und interpretierbaren Form Informationen darstellen (Lassmann, W. et al. 2006). Des Weiteren erwähnt die ISO und die IEC in einer Randnotiz zur Datendefinition: „*Data can be processed by humans or by automatic means.*” - Daten können von Menschen oder automatisiert verarbeitet werden (ISO/IEC 2382). In diesem Zusatz wird die menschliche oder maschinelle Lesbarkeit und Bearbeitbarkeit von Daten beschrieben. Ob Daten nur von Maschinen oder auch von Menschen interpretiert werden können, hängt unter anderem auch von ihrer Darstellungsform ab, welche eine Möglichkeit bietet Daten zu klassifizieren (Lassmann, W. et al. 2006). Dies wird anhand von Abbildung 2 veranschaulicht.

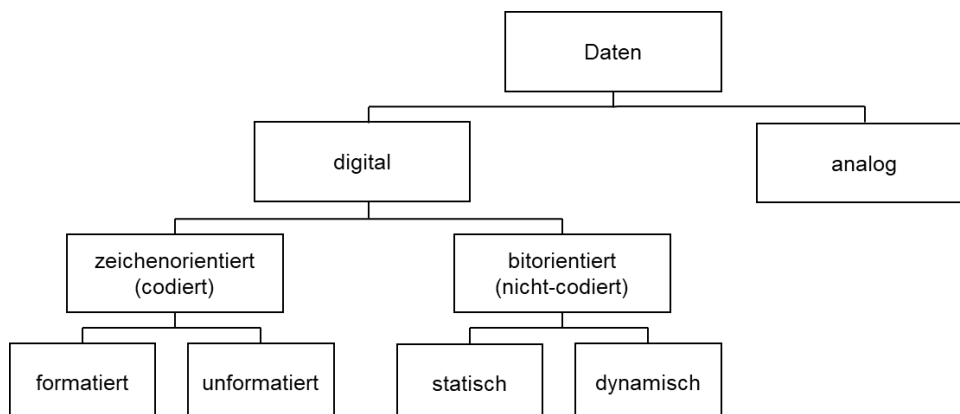


Abbildung 2: Klassifikation von Daten nach der Darstellungsform (Lassmann, W. et al. 2006, S. 216)

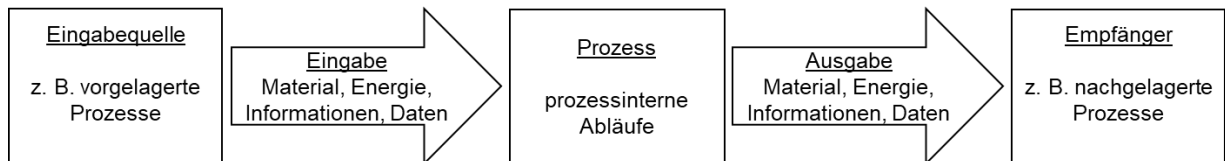
Die vorliegende Arbeit beschäftigt sich mit digitalen Inhalten, da mittlerweile alle kommerziell genutzten Rechner digital arbeiten (Lassmann, W. et al. 2006) und somit auch datengetriebene Prozesse in den Bereichen Produktion und Logistik digitale Daten verwenden. Falls analoge Daten zum Zweck der Verarbeitung von informationstechnischen Prozessen aufgenommen werden, werden diese mit Hilfe eines Analog Digital Wandlers in digitale Daten transformiert. Um später analysieren zu können welche Daten von datengetriebenen Prozessen verarbeitet werden können, folgt eine kurze Beschreibung der Klassifikation digitaler Daten. **Bitorientierte Daten** bestehen aus einer Abfolge von Einsen und Nullen (Lassmann, W. et al. 2006) und lassen sich darüber hinaus in ihrer Erscheinungsform unterscheiden, wie beispielsweise akustisch (z. B. Audioaufzeichnungen) oder visuell (z. B. Bilder und Filme) (Mertens et al., 2017). Abhängig von ihrer Erscheinungsform können bitorientierte Daten in statische Daten (z. B. Bilder und Grafiken) und in dynamische Daten (z. B. Videos und Audioaufnahmen) unterteilt werden (Lassmann, W. et al. 2006). **Zeichenorientierte Daten** setzen sich aus Zeichen bestimmter Datentypen zusammen, welche wiederum aus einer Abfolge von Einsen und Nullen codiert werden. Die Datentypen können in numerische (Ziffern), alphabetische (Buchstaben) und alphanummerische Daten (Ziffern, Buchstaben und Sonderzeichen) unterschieden werden (Mertens et al., 2017). Des Weiteren können Daten in der Art ihrer Strukturierung unterteilt werden. Unterschieden wird zwischen unstrukturierten (unformatierten), semistrukturierten und strukturierten (formatierten) Daten. Falls keine Länge oder Struktur der Daten vorgegeben ist, werden diese als unstrukturiert bezeichnet. Unter unstrukturierten Daten fallen demnach z. B. Bilder, Videos und Sprachaufzeichnungen also bitorientierte Daten aber auch z. B. Texte also zeichenorientierte Daten. Semistrukturierte Daten sind Daten, die an sich unstrukturiert sind, denen aber von außen eine Struktur auferlegt wird (z. B. Webseiten). Strukturierte (formatierte) Daten haben eine feste Reihenfolge innerhalb ihres Datensatzes, die Attribute sind definiert und die Datentypen sind festgelegt, dies sind zum Beispiel Datenbanktabellen (Cleve und Lämmel, 2014). Daten lassen sich darüber hinaus auch nach ihrem Verwendungszweck unterscheiden. Um verschiedene Einsatzmöglichkeiten von Daten im Bereich produktionstechnischer- und logistischer Prozesse zu beleuchten, werden im nächsten Abschnitt etablierte informationstechnische Prozesse eingeführt, welche in vielen Systemen in der Produktion und Logistik zum Einsatz kommen.

## 2.2 Informationstechnische Prozesse in Produktion und Logistik

Da sich diese Arbeit mit informationstechnischen datengetriebenen Prozessen im Umfeld von Produktion und Logistik beschäftigt, werden folgend der Begriff informationstechnischer Prozess (IT-Prozess) und die Bereiche Produktion und Logistik näher erläutert, um im nächsten Schritt einen kurzen Überblick über etablierte IT-Prozess-Lösungen in den Bereichen zu schaffen. Dies ist zum einen wichtig, um den vorrangigen Wirkungsbereich der in dieser Ar-

beit behandelten Prozesse einzuordnen und zum anderen um auf Grundlage dieses Kapitels eine adäquate Differenzierung von datengetriebenen Prozessen und bereits etablierten IT-Lösungen vornehmen zu können (Abschnitt 3.4.1).

Ein **Prozess** ist laut (DIN EN ISO 9000) ein spezielles Vorgehen, welches durch zusammenhängende Tätigkeiten Eingaben in Ausgaben transformiert, welche den Zielen und funktionsübergreifenden Barrieren einer Organisation entsprechen. Prozesse können, müssen aber nicht entscheidend sein.



**Abbildung 3: Darstellung eines Einzelprozesses (i. A. a. DIN IEC 60050-351, S. 32.)**

Wie in Abbildung 3 dargestellt, finden innerhalb eines Prozesses technologische und organisatorische Abläufe statt, welche Material, Energie oder Informationen einer Eingabequelle nutzen, um eine Ausgabe zu generieren. Sowohl Eingabequelle als auch Empfänger können ebenfalls Prozesse sein. **IT-Prozesse** sind in Bezug auf dieser Definition spezielle Prozesse, welche Computer nutzen, um aus einer Eingabe (Input) eine Ausgabe (Output) zu generieren. Die prozessinternen Abläufe verarbeiten demnach hauptsächlich digitale Daten (vgl. Abschnitt 2.1). Somit arbeiten IT-Prozesse nach dem EVA-Prinzip, welches das Grundprinzip der EDV darstellt und für Eingabe, Verarbeitung und Ausgabe steht. Die Ein- und Ausgabe Werte müssen nicht gezwungenermaßen aus Daten bestehen. Bei der Eingabe können beispielsweise Informationen mit Sensoren aufgenommen und digitalisiert werden. Bei der Ausgabe von IT-Prozessen können zum Beispiel verarbeitete Daten mittels Aktoren in physische Energie umgesetzt werden. Um ein besseres Verständnis für die Einsatzmöglichkeiten und Aufgaben von IT-Prozessen in der Produktion und Logistik zu entwickeln, werden die Einsatzbereiche kurz beschrieben, um abschließend konkret auf etablierte informationstechnische Systeme im Umfeld von Produktion und Logistik einzugehen.

Unter der **Produktion** wird ein Funktionsbereich einer funktional gegliederten Unternehmensorganisation verstanden, welcher sich zwischen den Bereichen Beschaffung und Absatz befindet. Das Ziel der Produktion ist die Erstellung von Produkten. Um eine Wertschöpfung zu erzielen, werden Produktionsfaktoren (menschliche Arbeit, Werkstoffe und Betriebsmittel) nach bestimmten Verfahren kombiniert und transformiert (Corsten, H. et al. 2018). Die Aufgaben der Produktion umfassen:

- Strategische (langfristige): z. B. Produktprogramm- und Produktionssystemplanung
- Taktische (mittelfristige): z. B. Mengenplanung, zeitliche- und räumliche Verteilung
- Operative (kurzfristige): z. B. Steuerung der Produktionsabläufe

Die **Logistik** im Allgemeinen befasst sich mit „[...] der Planung, Steuerung und Optimierung der Material-, Personen-, Energie- und Informationsflüsse in Systemen, Netzen und Prozessen“ (ten\_Hompel und Heidenblut, 2011). Die Logistik innerhalb eines Unternehmens befasst sich mit der Planung, Organisation, Abwicklung, Überwachung und Steuerung der anfallenden Waren- und Informationsflüsse entlang der gesamten Lieferkette, auch Supply Chain (SC) genannt. Die SC reicht von der Beschaffung, über die Produktion und Distribution bis hin zum Kunden (Mathar und Scheuring, 2012). Die Hauptfunktionen der Logistik beschäftigen sich hierbei mit Aufgaben des Transports, des Umschlags und der Lagerung (TUL-Logistik). Aber auch außerhalb der TUL-Logistik sind logistische Aufgabenbereiche vielfältig und beinhalten beispielsweise logistische Mehrwertdienste oder der Datenerhaltung und dem Informationsmanagement (ten\_Hompel und Heidenblut, 2011).

**Informationstechnische Prozesse in der Produktion und Logistik** unterstützen die zuvor beschriebenen Aufgaben. Diese lassen sich in Anlehnung an DIN EN 62264-1, S. 16–19. in fünf Ebenen unterteilen, wie in Abbildung 4 dargestellt. Die Ebenen beinhalten die Geschäftsplanung und Logistik, das Betriebsmanagement sowie die Chargensteuerung, kontinuierliche oder diskrete Steuerung. Jede Ebene erfüllt andere Aufgaben und Funktionen in unterschiedlichen zeitlichen Rahmen.

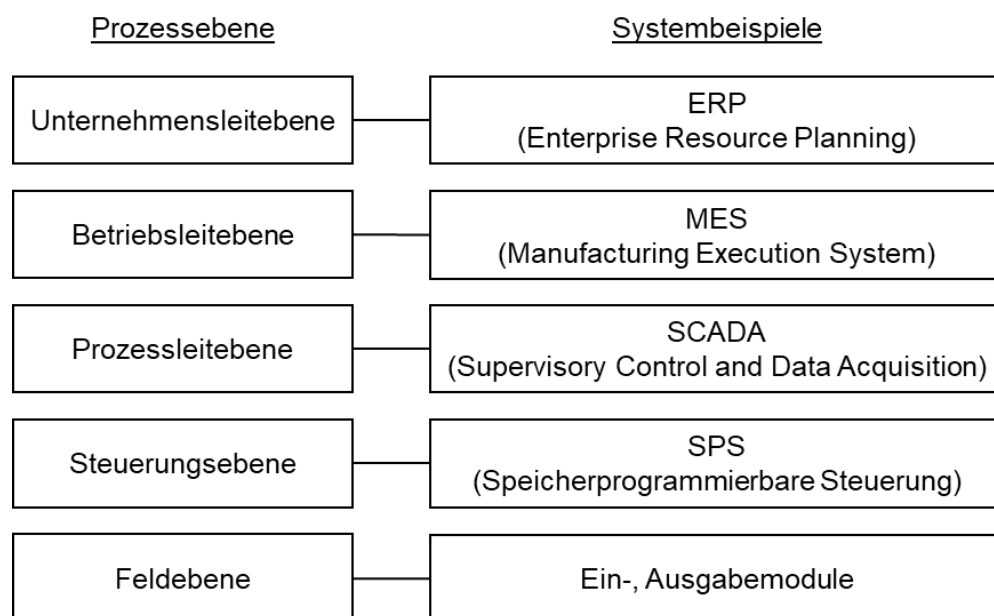


Abbildung 4: Ebenen der betrieblichen IT-Systemlandschaft (i. A. a. DIN EN 62264-1, S. 16–19.)

Die Feldebene auf der untersten Stufe repräsentiert die Schnittstelle zwischen den physisch technischen Systemen und den informationstechnischen Systemen. Die Steuerungsebene befasst sich mit der rechnergestützten Steuerung und Regelung von Vorgängen. Die Prozessleitebene ist für die Überwachung und Führung sämtlicher untergeordneter Prozesse zuständig. Die Betriebsleitebene regelt die Arbeitsabläufe, pflegt die Daten und optimiert zum Beispiel

die Produktionsprozesse. Die Unternehmensebene repräsentiert die höchste Ebene und ist für alle strategischen und taktischen Planungsaufgaben für alle Bereiche eines Unternehmens zuständig. Die Kommunikation der einzelnen Systeme geschieht sowohl innerhalb einer Prozessebene, wie auch übergreifend mit angrenzenden Ebenen (Arnold und Furmans, 2019). Auf jeder dieser Ebenen können unterschiedliche IT-Systeme eingesetzt werden. Eine genaue Darstellung der Funktionsweisen sämtlicher etablierten IT-Prozesse in Produktion und Logistik würde an dieser Stelle zu weit führen und wäre im Sinne dieser Arbeit nicht zielführend. Jedoch werden im Folgenden die Vorgänge der Datenauswertung „klassischer“ IT-Prozesse näher erläutert, um deren Ziele, Datenstruktur und -quellen und Technologien aufzuzeigen. Die Prozesse lassen sich meist in die folgenden vier Systemarten eingliedern:

- **Online Transaction Processing Systeme (OLTP):** Sind klassische Transaktionssysteme, wie zum Beispiel Enterprise Resource Planning (ERP) oder Supply Chain Management (SCM) Systeme. Diese Systeme erzeugen und verarbeiten strukturierte Daten (Stamm-, Bewegungs- und Bestandsdaten). OLTP-Systeme bilden komplette operative Geschäftsprozesse ab, sind aber ungeeignet, um auf ihrer Datengrundlage fundierte betriebswirtschaftliche Entscheidungen zu treffen.
- **Online Analytical Processing Systeme (OLAP):** Sind Analysesysteme zur Verdichtung von Daten und zeitraumbezogenen Analysen (z. B. Data Warehouse). Es werden Daten mit historischem, aktuellem oder zukünftigem Bezug einbezogen, um auf deren Basis fundierte Entscheidungen zur Steuerung des Unternehmens treffen zu können. Die benötigten Daten werden aus verschiedenen Quellsystemen (z. B. ERP-Systeme), mittels ETL-Prozess (Extraktion, Transformation, Laden) extrahiert und harmonisiert.
- **Master Data Management Systeme (MDM):** Übernimmt die Klammernung redundanter Daten in den Datenquellsystemen (z. B. ERP-Systeme), um widerspruchsfreie Daten zu gewährleisten. Das MDM bereinigt redundante Daten innerhalb einzelner Systeme und bei Systemübergreifenden Abfragen.
- **Dokumentenmanagement-Systeme (DMS):** Zum Speichern und Auffinden von Dokumenten. Sie sorgen dafür, dass Daten zentral gesichert und zur Verfügung gestellt werden.

Die übergeordneten Ziele der „klassischen“ IT-Systeme in Produktion und Logistik sind eine Basis zu schaffen, um fundierte Entscheidungen treffen zu können und das bestehende Geschäft zu optimieren und auszuweiten. Diese Systeme, welche auf physikalischen und statistischen Modellen aufbauen, arbeiten hauptsächlich mit strukturierten Daten, welche für gewöhnlich in relationalen Datenbanken (Datenbanken mit tabellarischer Struktur und klaren Schemata zur Organisation der Datenstruktur) gespeichert werden. Die Datenquellen beschränken sich dabei meist auf unternehmensinterne Bereiche, in denen unterschiedlichste Daten erhoben werden beispielsweise Stammdaten, Bewegungsdaten, Plandaten und Aktualdaten (Appelfeller und Feldmann, 2018). Diese Daten werden unter anderem von Sensoren der jeweiligen Systeme erfasst. Vor allem der vermehrte Einsatz von Systemen zur automati-

schen Identifizierung und Datenerfassung (AutoID), hat die erfassten Datenmengen im Umfeld von Produktion und Logistik steigen lassen und ermöglicht darüber hinaus eine sehr genaue und schnelle Datenerfassung materialflussbezogener Daten (ten\_Hompel et al., 2008). Unter anderem diese Vernetzung des Materialflusses, der alle physischen Vorgänge verknüpft welche Objekte durchlaufen und des Informationsflusses, welcher den Materialfluss digital abbildet und beispielsweise Informationen zur Auftragsabwicklung umfasst, macht eine Vielzahl neuer Prozesse und Anwendungen (z. B. Cyber-Physische-Systeme) möglich, welche unter dem Begriff Industrie 4.0 zusammengefasst werden.

### 2.3 Einführung des Begriffes Industrie 4.0

Aufgrund der fortschreitenden Digitalisierung und der Vernetzung von Menschen, Maschinen, Produkten und Unternehmen untereinander, befindet sich die Industrie derzeit in einer vierten Industriellen Revolution. Analog zu der Benennung neuer Softwareversionen wird dieser Wandel auch als Industrie 4.0 bezeichnet, welche für innovative und transformative Prozesse der industriellen Wertschöpfung steht (Bundesministerium für Wirtschaft und Energie). Nachfolgend wird der Begriff Industrie 4.0 näher beschrieben, um später die Beziehung von datengetriebenen Prozessen und Industrie 4.0 zu untersuchen.

Das Bundesministerium für Wirtschaft und Energie beschreibt **Industrie 4.0** auf ihrer Plattform Industrie 4.0 als „[...] *die intelligente Vernetzung von Maschinen und Abläufen in der Industrie mit Hilfe von Informations- und Kommunikationstechnologie*“ (Bundesministerium für Wirtschaft und Energie). Diese Beschreibung greift aktuelle Themenfelder der Industrie 4.0 im Bereich von Produktion und Logistik auf. Diese sind vorrangig die Vernetzung und Integration, die Datenerfassung und -verarbeitung, die Dezentralisierung, Assistenzsysteme und Selbstorganisation und autonomes Handeln (ten\_Hompel et al., 2020). Vor allem die Nutzung von Künstlicher Intelligenz (KI) spielt in vielen Industrie 4.0 Anwendungen eine größere Rolle (ten\_Hompel et al., 2020). Unter Industrie 4.0 wird keine spezifische Einzelanwendung verstanden, sondern Technologien und Themen, welche auf Produkt und Markt ausgerichtet sind. Um einen Überblick über die große Themenvielfalt zu gewährleisten und Unternehmen die Entwicklung einer Industrie 4.0 Unternehmensstrategie zu erleichtern, wurde 2016 das Fraunhofer-Schalenmodell der Industrie 4.0-Wertschöpfung entwickelt.



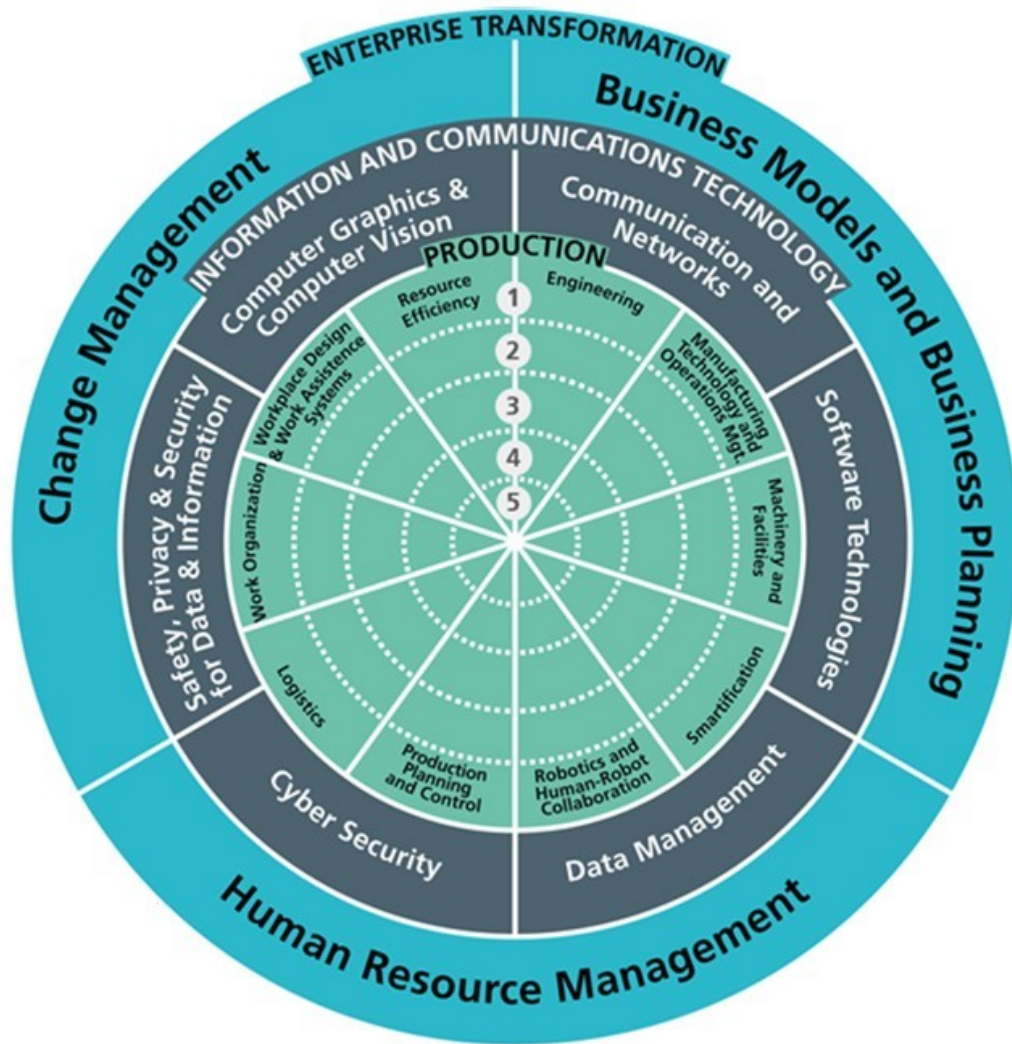


Abbildung 5: Fraunhofer-Schalenmodell der Industrie 4.0-Wertschöpfung (Fraunhofer IOSB, 2017, S. 15)

Abbildung 5 teilt die Themen von Industrie 4.0 in die drei Bereiche Produktion, Informations- und Kommunikationstechnik und Unternehmenstransformation auf. Untergeordnet befinden sich 19 Themenbereiche, unter denen sich eine Vielzahl von Einzelthemen befinden, welche extern in einem „Themenprisma“ beschrieben werden. Drei derzeit bedeutungsvolle Themen sind die Verwendung von KI, die Datenerfassung durch AutoID und Anwendungen auf dem Gebiet Big Data, welche auch bei datengetriebenen Prozessen von Belangen sein können und daher näher betrachtet werden müssen.

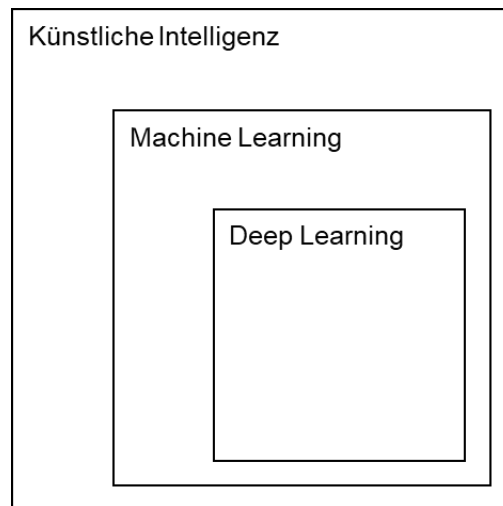


Abbildung 6: Beziehungen von KI, ML und DL (i. A. a.ten\_Hompel et al., 2020, S. 232)

Wie bereits erwähnt spielt die Verwendung von **künstlicher Intelligenz** in vielen Bereichen der Industrie 4.0 eine große Rolle. Diese ermöglicht zum einen die Forschung an der Erschaffung von Intelligenz und zum anderen, durch die Erstellung und Verwendung komplexer Lernalgorithmen, das intelligente Agieren von Maschinen. Zur Verarbeitung von Informationen werden diese aus aufgenommenen Daten entnommen, um Schlussfolgerungen zu ziehen und gegeben falls daraus Handlungen abzuleiten (ten\_Hompel et al., 2020). Häufig werden die Begriffe Machine Learning (ML) und Deep Learning (DL) als Synonyme für KI verwendet. Aus Sichtweise der Logistik stellen diese aber nicht zusammenfallende Teilbereiche von KI dar, deren Beziehung zueinander in Abbildung 6 verdeutlicht wird. Abbildung 6 muss kritisch betrachtet werden, denn sie beschreibt lediglich eine disziplinbegründete Auffassung der Logistik, welche aufgrund der Eingrenzung der Arbeit auf die Bereiche Produktion und Logistik übernommen wird. Laut Mitchell, 1997 ist ML die Verhaltensänderung eines Computers zwischen zwei definierten Zeitpunkten, welche durch ein bestimmtes Ereignis induziert wurde. Diese Verhaltensänderung stellt eine messbare Verbesserung zum Ursprungszustand dar. Je nach Einsatzgebiet wie zum Beispiel der automatischen Erkennung von bestimmten Bauteilgeometrien, existieren eine Vielzahl an Algorithmen zur Erfüllung der Aufgaben. Eine spezielle Gruppe dieser Lernverfahren ist unter DL vereint. DL nutzt künstliche neuronale Netze, um am Computer die Lernverfahren des menschlichen Hirns zu imitieren. Künstliche Neuronen werden dabei in Schichten organisiert und diese Schichten untereinander über die schichtübergreifende Verbindung von Neuronen vernetzt. Durch die Aktivierung von Neuronen im Netz und deren Gewichtung werden Informationen verarbeitet. DL Anwendungen arbeiten mit sehr vielen komplexen Schichten. Dies ermöglicht es auch sehr komplexe Entscheidungen zu finden, die nicht wie in der traditionellen Programmierung durch das händische Programmieren von Regeln, sondern durch das Trainieren des neuronalen Netzes erschlossen werden (Bhattacharyya, S. 2020). Um solche künstlichen neuronalen Netze zu trainieren, benötigen DL Anwendungen jedoch eine sehr große Anzahl an Trai-

ningsdaten. Wie zuvor im Abschnitt 2.2 erwähnt, haben AutoID Systeme dazu beigetragen, dass sehr viele Daten im Bereich von Produktion und Logistik erfasst werden, sodass diese beispielsweise zum Training der DL Anwendungen genutzt werden können.

**AutoID** ist ein übergeordneter Begriff für Technologien, welche der automatischen Identifizierung von Objekten in Materialflüssen dienen. Diese tasten mit Sensoren Informationen aus dem Materialfluss ab, codieren und übermitteln diese zum Zwecke der Weiterverarbeitung. Vor allem die vermehrte Nutzung von sogenannten RFID Systemen führt zu einer erhöhten Aufnahme von Daten. RFID steht für Radio Frequency Identification und ist ein elektromagnetisches beziehungsweise induktives AutoID System, welches aufgrund seiner Genauigkeit, Schnelligkeit und vielseitigen Einsatzmöglichkeiten zunehmend an Bedeutung gewinnt. RFID Systeme zeichnen sich insbesondere dadurch aus, dass für Aktionen kein direkter Sichtkontakt benötigt wird und ein mobiler Datenspeicher direkt am Objekt mitgeführt werden kann (ten\_Hompel et al., 2008). Auch diese Möglichkeit Daten am Objekt mitzuführen und zu speichern, aber auch zusätzliche Wege der Datenbeschaffung zum Beispiel über das Internet, führen zu einer stetigen Zunahme an Datenmengen. Ein großer Bereich der Industrie 4.0, welcher sich mit der Erfassung und der Verarbeitung solch großer Datenmengen beschäftigt, wird unter dem Namen Big Data zusammengefasst. Dieser bietet im Gegensatz zu den im Abschnitt 2.2 behandelten IT-Prozessen Möglichkeiten die großen Datenmengen aus unterschiedlichsten Datenquellen nutzbar zu machen.

## 2.4 Einführung des Begriffes Big Data

Aufgrund ihrer Gemeinsamkeiten werden die Begriffe Data Driven und Big Data häufig nicht genau differenziert. Im folgenden Kapitel wird der Begriff Big Data etabliert und die Charakteristiken von Big Data Anwendungen näher beleuchtet, um einen grundlegenden Überblick zu schaffen und um später den Begriff Data Driven von der Begrifflichkeit Big Data abgrenzen zu können.

Obwohl der Begriff **Big Data** seit einigen Jahren allgegenwärtig ist, hat sich bis in das Jahr 2021 noch keine allgemeine Definition für den Begriff Big Data durchgesetzt (Pietsch, 2021). Wörtlich bedeutet der Begriff große Datenmengen. Auch wenn die Nutzung großer Datenmengen einen zentralen Punkt von Big Data darstellt, ist dies keine Beschreibung, die diesem Begriff hinreichend genau widerspiegelt. Eine in der Literatur oft zitierte Begriffsformulierung ist die des Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V. (BITKOM). Der BITKOM beschreibt Big Data als „[...] Einsatz großer Datenmengen aus vielfältigen Quellen mit einer hohen Verarbeitungsgeschwindigkeit zur Erzeugung wirtschaftlichen Nutzens [...]“ (BITKOM, 2014, S. 12). In dieser Beschreibung wird Big Data unter anderem dadurch gekennzeichnet, dass beim Einsatz die Erzeugung eines wirtschaftlichen Nutzens gefordert wird, um einen Mehrwert für das Unternehmen zu generieren.

Darüber hinaus werden von der BITKOM drei weitere Facetten von Big Data angeführt (große Datenmengen, vielfältige Quellen und hohe Verarbeitungsgeschwindigkeit). Diese vier Teilaspekte werden im Folgenden durch eine weitere wichtige Größe der Datenqualität zu den in der Literatur häufig benannten 5 Vs von Big Data ergänzt (Bachmann et al., 2014).

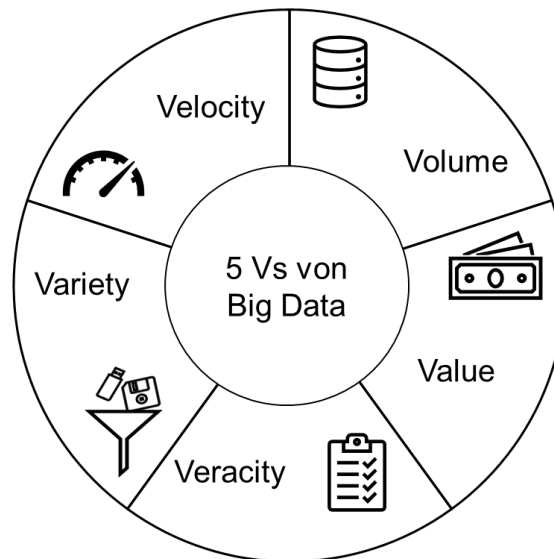


Abbildung 7: Die 5 Vs von Big Data

In Abbildung 7 sind die 5 Vs von Big Data bildlich dargestellt. Diese werden im nächsten Abschnitt kurz erläutert. In der Literatur werden manchmal weitere Aspekte von Big Data angeführt, welche in dieser Arbeit jedoch nicht näher erörtert werden (Sun, 2018).

- **Volume:** Große Datenmengen, welche bei Big Data Anwendungen zur Verfügung stehen und herangezogen werden. Die Bestandsmenge der Daten übersteigt dabei meist die Menge, die von Unternehmen tatsächlich in einer bestimmten Zeit verarbeitet werden kann und lag bereits 2016 im Zettabyte-Bereich (Fasel und Meier, 2016). Die Datenmenge hat bis heute (Jahr 2021) weiter zugenommen und wird wahrscheinlich auch in den nächsten Jahren steigen.
- **Velocity:** Geschwindigkeit der Datenübermittlung, Datengenerierung und Datenauswertung. In vielen Anwendungsfällen sind die Datenmengen erst nutzbar, wenn diese nahezu in Echtzeit erfasst, verarbeitet und übermittelt werden können (Gadatsch und Landrock, 2017).
- **Variety:** Vielfalt an Datentypen, -strukturen und -klassen, Datenquellen aber auch an Verfahren zur Datenanalyse. Big Data Verfahren verarbeiten jede Datenstruktur sowohl strukturierte-, semistrukturierte- wie auch unstrukturierte Daten (vgl. Abschnitt 2.1) (Gadatsch und Landrock, 2017). Auch die Technologien auf dem Gebiet Big Data sind vielfältig und neben spezifischen Methoden notwendig, um die Datenmengen zu verarbeiten (z. B. Hadoop oder Apaches Spark) (Kollmann, 2020). Die Vielfalt an Da-

tenquellen, welche auch zum rasanten Zuwachs an Daten beitragen, reichen von Unternehmensdaten, Social Media, Smartphones, Selbstbeobachtungsdaten, Open Data bis hin zu Sensordaten (Holland, 2021).

- **Value:** Der Wert von Big Data Anwendungen für ein Unternehmen. Durch das Zusammenspiel der anderen Facetten von Big Data werden wichtige Erkenntnisse und somit ein Nutzen generiert, welcher einen zentralen Aspekt von Big Data darstellt. Dieser Mehrwert wird generiert, indem in gesammelten Datenbeständen nach Zusammenhängen und Erkenntnissen gesucht wird. Dieser Nutzen ist im Vorfeld eines Big Data Projektes meist nicht klar erkennbar (Gadatsch und Landrock, 2017). Auf Grundlage dieser Erkenntnisse können künftige Ereignisse mit einer gewissen Wahrscheinlichkeit vorhergesagt werden und daraus potenzielle Handlungen bewertet werden. Zudem ermöglicht der Einsatz von Big Data die Erschließung neuer Geschäftsmodelle. Nach Holland ermöglicht somit erst Big Data echte datengestützte und datengetriebene Unternehmen (Holland, 2020).
- **Veracity:** Oft auch Validity genannt, bezeichnet die Anforderung einer hohen Datenqualität. Vor allem durch den Punkt der hohen Vielfalt an Datenquellen, -strukturen und -klassen wird der Validierung der Daten eine hohe Bedeutung zugesprochen (Bachmann et al., 2014).

Durch die 5 Vs von Big Data, werden vor allem die Unterschiede zu etablierten IT-Systemen in der Produktion und Logistik deutlich. Diese liegen nicht nur in der Nutzung großer Datenmengen, sondern auch in der Erschließung neuer Datenquellen vor allem auch durch die Verwendbarkeit von unstrukturierten Daten. Der Nutzen von Big Data lässt sich zwar im Voraus nicht genau abschätzen, dafür besteht aber die Möglichkeit präskriptiv Handlungen zu bewerten.

## 2.5 Einführung des Begriffes Data Driven

Datengetriebene Prozesse scheinen eine weitere Möglichkeit zu bieten aus großen Datenmengen einen Nutzen zu ziehen. Aber wie zuvor angeführt, sind datengetriebene Prozesse im Umfeld von Produktion und Logistik noch nicht hinlänglich beschrieben worden, weshalb bisher kein direkter Vergleich zu ähnlichen Begriffen wie Big Data oder Industrie 4.0 möglich ist. Um einen Vergleich zu anderen Begriffen zu ermöglichen und um die Datennutzung und weitere wichtige Aspekte datengetriebener Prozesse zu analysieren, muss der Begriff Data Driven näher betrachtet werden.

Datengetriebene Prozesse werden mittlerweile in vielen Anwendungsbereichen verwendet beziehungsweise für diese entwickelt. So bietet beispielsweise Data Driven Learning einen neuen auf Daten basierenden Lernansatz, der unter anderem das Lernen neuer Fremdsprachen

effektiver gestaltet (Men, 2020) oder Data Driven Journalism Prozesse, welche große Datenmengen nach Informationen über spezifische Nachrichteninhalte analysieren (Hammond, 2017). Auch im Bereich von Produktion und Logistik existieren datengetriebene Prozesse für eine Vielzahl von Anwendungsfällen, wie etwa das Data Driven Transport Management (DTM), das das Ziel einer gänzlich digitalisierten Transportkette verfolgt (Heinbach et al., 2021). Abseits dieser konkreten Anwendungsfälle wird der Begriff Data Driven jedoch nicht näher spezifiziert, weshalb dieser bisher (Jahr 2021) für den Bereich Produktion und Logistik nicht eindeutig definiert ist. Falls für den Begriff Data Driven im Umfeld von Produktion und Logistik eine Definitionsbestimmung möglich ist, existieren auch Überschneidungen in den Beschreibungen datengetriebener Prozesse, welche als Merkmale zur Begriffsbestimmung herangezogen werden können. Die einzige Aussage, die ohne genauere Betrachtung konkreter Prozesse getätigt werden kann, ist dass es sich bei datengetriebenen Prozessen um informationstechnische Prozesse handelt, welche eine größere Menge an Daten (genaue Datenmenge nicht spezifiziert) nutzen, um einen Mehrwert zu generieren. Somit ähneln datengetriebene Prozesse auf dem ersten Blick sehr Big Data, was erneut die Relevanz einer klaren Trennung dieser Begriffe verdeutlicht.

Welchen genauen Mehrwert datengetriebene Prozesse schaffen, ob diese beispielsweise eher Anwendungs- oder Analyseorientiert sind, ist eine Frage, die es für eine Begriffsbestimmung zu beantworten gilt. Weitere Fragen an den Begriff Data Driven ergeben sich aus den vorangegangenen Beschreibungen über die Einsatzmöglichkeiten von Daten in der Produktion und Logistik. IT-Prozesse zeichnen sich beispielsweise darüber aus, welche Datentypen, -struktur, -qualität und -menge diese nutzen und aus welchen Quellen sie diese beziehen. Ein weiteres wichtiges Kriterium ist die Verwendung solcher Systeme. Neben den verschiedenen Anwendungsbereichen unterscheiden sich z. B. etablierte IT-Prozesse und Big Data Anwendungen darin, ob diese nur zur Optimierung alter- oder auch zur Erschließung neuer Geschäftsmodelle eingesetzt werden können. Zudem ist es entscheidend, ob ein Analysesystem nur vergangene Geschehnisse analysiert, oder es ermöglicht Voraussagen über zukünftige Ereignisse zu treffen. Um für eine Begriffsbestimmung datengetriebener Prozesse möglichst viele dieser Fragen und eventuell neu auftauchende Fragen zu klären, wird im weiteren Verlauf der Arbeit eine Literaturanalyse zu datengetriebenen Prozessen durchgeführt. Zur Gewährleistung eines standardisierten Durchführungsprozesses dieser Analyse werden im nächsten Abschnitt Methoden der Literaturanalyse vorgestellt.

## 2.6 Methoden der Literaturanalyse und Analysekriterien zur Begriffsfindung

Im Folgenden wird eine Auswahl an Methoden eruiert, welche mögliche Vorgehensweisen der anschließenden Literaturrecherche und der Definitionsbestimmung, insbesondere bei der Kriterienbildung darstellen. Bei den Vorgehensweisen handelt es sich vorwiegend um Forschungstechniken der empirisch arbeitenden Kommunikationswissenschaft, welche hauptsächlich die Berichterstattung von Ereignissen und Themen analysiert (Springer et al., 2015). Die Techniken lassen sich unterteilen in die Beobachtung, die Befragung und die Inhaltsanalyse. Jede dieser Methoden kann entweder die Gesamtheit oder nur eine Stichprobe einer Menge untersuchen. Da in dieser Arbeit ausnahmslos Inhalte aus der Literatur und aus Internetquellen untersucht werden, beschränken sich die folgenden Erläuterungen auf die Methode der Inhaltsanalyse, da bei textuellen Analysen weder die Beobachtung noch die Befragung Verwendung finden.

Die empirische Methode der Inhaltsanalyse kann sowohl Inhalte textueller Form (z. B. Bücher, Internetseiten), sowie auch nicht textuelle Inhalte wie beispielsweise Bilder erheben und analysieren. Wichtig ist zum einen, dass alle Schritte intersubjektiv nachvollziehbar gestaltet sind, also verständlich und vollständig dokumentiert werden, um eine prinzipielle Wiederholbarkeit der Analyse zu gewährleisten. Zum anderen wird eine systematische Vorgehensweise gefordert, bei der alle Daten unter gleichen Bedingungen erhoben und analysiert werden (Früh, 2017). Bei der Inhaltsanalyse wird grundsätzlich zwischen zwei Methoden unterschieden, der quantitativen und der qualitativen Inhaltsanalyse. Bei der **quantitativen Inhaltsanalyse** werden die zu untersuchenden Inhalte auf zuvor klar definierte Merkmalausprägungen hin quantitativ untersucht. In der Regel wird geprüft ob oder in welchem Umfang bestimmte Merkmale auftreten. Im Anschluss werden diese Merkmale bestimmten Kategorien zugeordnet. Die **qualitative Inhaltsanalyse** wurde vom deutschen Psychologen Philipp Mayring entwickelt und kann Auskunft darüber geben, welche Aussagen und Zusammenhänge einzelne Kommunikate wiedergeben. Aufgrund dessen kann sie auch zur empiriegeleiteten Kategorienbildung eingesetzt werden. Das heißt, dass mit Hilfe dieser Methode auch Kategorien aus einer Analyse heraus hergeleitet werden können (Springer et al., 2015). Nach Mayring, 2015 können Inhalte durch die quantitative Inhaltsanalyse sowohl zusammengefasst als auch expliziert und strukturiert werden. Nicht immer wird zwischen quantitativer- und qualitativer Analyse trennscharf unterschieden. Je nach Aufgabenstellung kann daher eine Mischform beider Vorgehensweisen die sinnvollste Methode darstellen, bei der zunächst qualitativ eine Fragestellung und erste Kategorien erarbeitet werden, die Kategorien quantitativ untersucht und gegeben falls ergänzt werden und zum Abschluss die Ergebnisse quantitativ vor dem Hintergrund der Fragestellung interpretiert werden (Mayring, 2015). Die erfassten und zu analysierenden Kategorien können Beispielsweise in einem Codebuch festgehalten werden in der auch formelle Identifikationskennzahlen (z. B. zur Identifizierung des\*der Au-

tors\*in oder des Erscheinungsjahres) festgehalten werden. Dieses Codebuch dient nicht nur dazu eine Übersicht der zu analysierenden Inhalte zu schaffen, sondern auch um ein systematisches Vorgehen zu gewährleisten, sodass die Ergebnisse der Analyse jederzeit reproduzierbar sind. Der genaue Aufbau des Codebuches ist jedoch dem\*der Forscher\*in selbst überlassen und von der konkreten Aufgabenstellung abhängig. Üblicherweise setzt sich ein Codebuch aus der Erläuterung der Forschungsfrage, den Definitionen der wichtigsten Begriffe und der Analysekriterien und dem eigentlichen Kategoriensystem inklusiver tabellarischer Übersicht der Kategorien zusammen (Rössler, 2017). Die Kategorien setzen sich meist zusammen aus inhaltlichen Kategorien (zur Untersuchung der eigentlichen Fragestellung) und formalen Kategorien (z. B. zur Messung der Relevanz der analysierten Quellen) (Früh, 2017).

Um ein Kategoriensystem für die Inhalte aus einer Inhaltsanalyse oder für diese zu entwickeln, bieten sich zwei Herangehensweisen an, nämlich die theoriegeleitete und die empiriegeleitete Kategorienbildung. Häufig werden diese auch in Verbindung miteinander angewandt. Bei der **theoriegeleiteten Kriterienbildung** werden Hypothesen direkt aus den interessierenden Bedeutungselementen des Forschungsthemas formuliert. Diese Hypothesen repräsentieren im weiteren Verlauf die Hauptanalysekriterien der Gliederung. Durch die erarbeiteten Kategorien erhält die Forschungsfrage eine Gliederung, mithilfe der eine Identifizierung oder Klassifizierung in Bezug auf die Textinhalte vorgenommen werden kann. Die Klassifizierung wird auf der Grundlage zuvor erarbeiteter Definitionen erarbeitet. Bei der **empiriegeleiteten Kategorienbildung** werden die zu untersuchenden Texte genutzt und mögliche gemeinsame Aussagen über den Untersuchungsgegenstand zu identifizieren. Aus den gemeinsamen Aussagen lassen sich mögliche Analysekategorien bilden, indem Aussagen erfasst werden und zusammen mit anderen Aussagen zu einer übergeordneten Kategorie zusammengefasst werden. Zudem kann die empiriegeleitete Kategorienbildung genutzt werden, um bestimmte Definitionen von Analysekategorien operational zu ergänzen. Bei einer gemischten Vorgehensweise werden zuerst theoriegeleitet die Hauptanalysekriterien aufgestellt, welche durch empiriegeleitete Kategorienbildung gebildete Unterkriterien ergänzt werden (Früh, 2017). Das Kategoriensystem der informellen Inhalte stellt eine Liste dar, anhand dessen ein Begriff kategorisch untersucht wird. Somit hilft dieses System die Analyse zu strukturieren. Die Ergebnisse der Inhaltsanalyse füllen das Kategoriensystem mit Merkmalen und erstellen somit ein essenzielles Abbild des zu untersuchenden Begriffes (Appelfeller und Feldmann, 2018).

Die **formalen Kategorien** dienen in einer Inhaltsanalyse der Erfassung manifester Textinformationen, wie zum Beispiel den Namen des\*der Autors\*in, das Jahr der Publikation oder den Seitenumfang des Analysetextes. Diese Informationen sind zwar nicht direkt relevant für die Analyse einzelner Textinhalte, sie erfüllen aber instrumentelle Funktionen, die für eine übergeordnete Auswertung der Analyse wichtig sind. Zum einen dienen formale Kategorien der Differenzierung einzelner Texte, zum anderen können diese verwendet werden, um die Relevanz von Quellen und Autoren\*innen zu betrachten und deren Aussagen bei der Auswer-



tung der Kategorien unterschiedlich zu gewichten (Rössler, 2017). Um die Relevanz von wissenschaftlichen Quellen und Autoren\*innen messbar zu machen, existieren verschiedene Ansätze. Alleine die Formen der Publikationen unterscheiden sich in ihrem Grad der Wissenschaftlichkeit (Voss, 2020). So werden Journale überwiegend für den Transfer von mehr oder weniger gesicherten Wissen publiziert, um dieses einem breiten Publikum zur Verfügung zu stellen. Im Gegensatz dazu haben Conference Paper eine höhere wissenschaftliche Relevanz, da diese primär neue wissenschaftliche Erkenntnisse sichern und daher einen hohen Anspruch an die wissenschaftliche Korrektheit stellen, welche durch Herausgeber und gutachtende Wissenschaftler\*innen geprüft wird (Kollmann et al., 2016).

Um die Relevanz der Quellen messbar zu machen, existieren eine Vielzahl von Indizes, welche durch die Verwendung verschiedener Kennzahlen wie zum Beispiel der Anzahl der Leser\*innen oder die Anzahl der Zitationen berechnet werden. Ein in den Naturwissenschaften gebräuchlicher Index ist der **Journal Impact Factor (JIF)**. Der JIF ist eine berechnete Maßzahl, welche sich nicht direkt an der Qualität einer Veröffentlichung orientiert, sondern daran, wie häufig die Aufsätze eines Journals innerhalb der letzten Jahre in anderen wissenschaftlichen Veröffentlichungen zitiert wurden (Voss, 2020). Der JIF wird aus den Datenbanken des sogenannten Web of Science des Unternehmens Clarivate ermittelt. Der JIF der letzten N Jahre für ein bestimmtes Jahr Y, berechnet sich aus der Anzahl der Zitate im Jahr Y in Bezug auf die Publikationen der vergangenen N Jahre, geteilt durch die Summe aller Veröffentlichungen der vergangenen N Jahre (Craig et al., 2014):

$$JIF_N \text{ des Jahres } Y = \frac{\sum_{i=1}^N \text{Zitat}(Y, Y - i)}{\sum_{i=1}^N \text{Publikation}(Y - i)}$$

Am gebräuchlichsten ist die Berechnung des JIF der vergangenen zwei Jahre. Die Anzahl an Zitationen als Maß für die wissenschaftliche Qualität von Journalen zu nutzen, wird oft kritisch gesehen, da beispielsweise die Anzahl der Zitationen stark vom Fach und dem Bekanntheitsgrad von Zeitschriften und Autor\*innen abhängt und darüber hinaus leicht zu manipulieren sind. Nichts desto trotz stellt der JIF eine berechnete und gebräuchliche Metrik dar (Craig et al., 2014). Eine Methode, welche zur Berechnung der Relevanz von Autor\*innen in der Wissenschaft Verwendung findet und ebenfalls eine Anzahl an Zitationen einbezieht, ist der nach Jorge E. Hirsch benannte **Hirsch-Index** oder auch h-Index genannt. Hirsch definiert den h-Index als die Anzahl der Veröffentlichungen mit einer Anzahl an Zitationen  $\geq h$ . Ein/Eine Wissenschaftler\*in hat einen Index von h, falls h seiner oder ihrer Veröffentlichungen über n Jahre jeweils mindestens h mal zitiert wurden und die restlichen Veröffentlichungen in diesem Zeitraum eine Anzahl an Zitationen  $\leq h$  aufweisen (Hirsch, 2005).

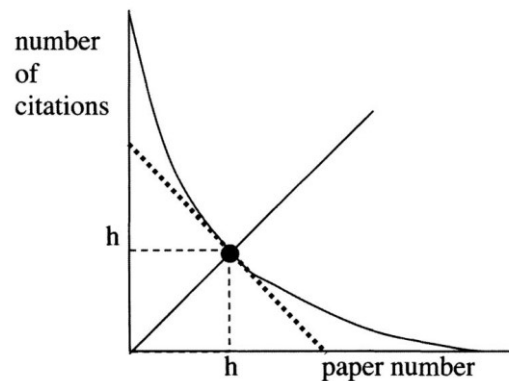


Abbildung 8: Schematische Darstellung des Hirsch-Indexes (Hirsch, 2005, S. 16570)

Aus dieser Definition ergibt sich eine grafische Darstellung des h-Indexes wie in Abbildung 8. Diese stellt den h-Index als Schnittpunkt einer Geraden mit einer Steigung von  $45^\circ$  und einer Kurve dar, welche sich aus der Anzahl der Veröffentlichungen aufgetragen über die Anzahl ihrer Zitationen ergibt. Mithilfe dieser Abbildung wird ersichtlich, dass der h-Index über die Zeit nicht sinken kann, aber auch nicht durch eine reine Zunahme an Veröffentlichungen steigt. Der Index kann aber andererseits auch nur maximal so groß werden, wie die Anzahl an Veröffentlichungen was dazu führt, dass in Fachdisziplinen mit einer höheren Anzahl an Veröffentlichungen auch höhere Indexe erreicht werden was interdisziplinäre Vergleiche schwierig gestaltet. Zudem beeinflussen einzelne Veröffentlichungen, selbst wenn sie herausragende wissenschaftliche Leistungen darstellen, den h-Index nicht nennenswert. Eine Gefahr die sowohl beim h-Index als auch beim JIF besteht ist, dass Autor\*innen die Indexe durch beispielsweise Selbstzitate oder durch Absprachen mit anderen Autor\*innen manipulieren (Craig et al., 2014). Onlinedatenbanken wie zum Beispiel Scopus oder CiteSeerX bieten die Möglichkeit neben der Anzeige der Gesamtanzahl an Zitationen auch unterschiedliche Zitationsmetriken wie den JIF oder den h-Index zu berechnen (Craig et al., 2014). Neben den bereits genannten Methoden zum Vergleich von Literatur und Autor\*innen existieren eine Vielzahl weiterer Metriken. Darunter beispielsweise der SNIP, der versucht einen fächerübergreifenden Vergleich von Fachzeitschriften zu gewährleisten, sich aber alleine auf die Datenbank Scopus beschränkt oder die Halbwertszeit der Zitationen, welche sich auf das Durchschnittsalter der im laufenden Jahr zitierten Arbeiten bezieht (Craig et al., 2014).

Zur Beschreibung von Begriffen dienen bestimmte **Begriffsmerkmale**. Laut DIN 55350 ist ein Merkmal eine kennzeichnende Eigenschaft eines Begriffes, welche entweder inhärent oder diesem zugeordnet sein kann. Unterscheiden lassen sich Merkmale unter anderem in quantitative Merkmale, welchen bestimmte Werte skalenhaft zugeordnet sind und qualitative Merkmale, die keiner skalenhaften Ordnung unterworfen sind. Durch eine konkrete Ausprägung eines Merkmals wird ein Merkmalswert eines Merkmals festgelegt. Erfasste Merkmale (z. B. Farbe) inklusive ihrer Ausprägungen (z. B. rot), dienen der Beschreibung einzelner Begriffe.

### **3 Untersuchung des Begriffes Data Driven zur Entwicklung eines Definitionsvorschlages**

Wie Kapitel 2 bereits verdeutlicht hat, bietet die Verwendung großer Datenmengen einen hohen potenziellen Nutzen für logistische- und produktionstechnische Prozesse. Datengestützte Prozesse bieten scheinbar eine Möglichkeit einen Mehrwert aus Datenmengen zu generieren, sind jedoch noch unerforscht weshalb es schwierig ist mögliche Einsatzgebiete zu identifizieren und über den Einsatz solcher Prozesse disziplinübergreifend zu kommunizieren. Um den Begriff Data Driven und dessen Bedeutung im Bereich von Produktion und Logistik im Kontext verschiedener Fachdisziplinen untersuchen zu können, werden auf einer Literaturanalyse gestützte Merkmale (Abschnitt 2.6) datengetriebener Prozesse erarbeitet. Die Aufgabe dieses Kapitels ist es, auf Grundlage der Kernaussagen verschiedener Autor\*innen und den auf dieser Basis erarbeiteten Begriffsmerkmalen eine Diskussion über den Begriff Data Driven zu ermöglichen, diesen in die Begriffswelt der Digitalisierung im Umfeld von Produktion und Logistik einzuordnen, um letztendlich einen eigenen Vorschlag einer Begriffsdefinition für Data Driven zu erarbeiten.

#### **3.1 Literaturanalyse zu datengetriebenen Prozessen**

Um die zuvor beschriebenen Ziele zu erreichen, folgt im kommenden Abschnitt die Literaturanalyse, welche den Kern dieser Arbeit bildet. Zur Vorbereitung der Analyse und um ein systematisches Vorgehen zu gewährleisten, wird in einem ersten Schritt eine geeignete Methode für die Analyse ausgewählt und diese gegebenenfalls an die Anforderungen dieser Arbeit angepasst. Im Zuge der Methodenanpassung und der nachfolgenden Analyse werden Analyse Kriterien erarbeitet, welche zum einen helfen die Kernaussagen verschiedener Autor\*innen besser zusammen zu fassen und zu bewerten und zum anderen eine Grundlage bieten für die spätere Erarbeitung von Merkmalen datengetriebener Prozesse. Unter Anwendung der ausgewählten Methode werden Aussagen verschiedener Autor\*innen zu datengetriebenen Prozessen und angrenzenden Themen aus der Literatur herausgearbeitet. Im Folgenden werden die Kernaussagen der Autor\*innen erfasst und tabellarisch übersichtlich dargestellt. Aufbauend auf den dargestellten Aussagen zum Begriff Data Driven und den entworfenen Analyse Kriterien können im weiteren Verlauf der Arbeit Begriffsmerkmale erarbeitet werden, welche eine Grundlage bieten den Begriff Data Driven näher zu untersuchen.

### 3.1.1 Methodenauswahl und -anpassung der Literaturanalyse und Analysekriterien

Um die Kernaussagen verschiedener Autor\*innen zum Thema Data Driven zu erfassen, wird eine passende Forschungsmethodik benötigt. Im Abschnitt 2.6 wurde die empirische Methode der Inhaltsanalyse bereits näher erläutert. Diese Methode ist bewährt, da sie gerade bei der Analyse textueller Inhalte eine systematische Vorgehensweise bietet, um vorhandenes Wissen zu spezifischen Themen sinnvoll zusammenzufassen und zu strukturieren (Abschnitt 2.6). Vor allem bei der Begriffsfindung kann die Inhaltsanalyse vorteilhaft sein, da diese mit Analyse-kriterien arbeitet, welche im Nachgang bei der Bestimmung der Begriffsmerkmale datenge-triebener Prozesse herangezogen werden können. Zuvor wurden die Methoden der quantitati-ven und der qualitativen Inhaltsanalyse kurz erläutert. Für die Durchführung der Analyse wird eine Mischform beider Forschungsmethoden genutzt, um zum einen qualitative Aussagen über datengetriebene Prozesse aus der Literatur zu eruieren und darüber hinaus Gemeinsam-keiten in den erfassten Aussagen messbar zu machen. Auch bei der Herleitung der Analyse-kriterien wird ein integrativer Ansatz aus theorie- und empiriegeleiteten Vorgehen verwendet. Wie in Abschnitt 2.6 angeführt können zum einen mit der theoriegeleiteten Kriterienbildung die zuvor als wichtig erachteten Fragestellungen untersucht werden und zum anderen mit der empiriegeleiteten Kriterienbildung während der Analyse Kriterien aufgenommen werden, welche zuvor noch keine Beachtung fanden. Die Kriterien werden in Anlehnung an ein von Kuckartz, 2012 angeführtes Ablaufschema für die Praxis qualitativer Inhaltsanalysen aufge-stellt. In seinem Ablaufschema teilt Kuckartz die Inhaltsanalyse in fünf Phasen auf, welche nicht zwangsläufig stufenweise abgearbeitet werden müssen. In der ersten Planungsphase wird die Problemstellung vorgestellt und daraus eine Kernhypothese formuliert. Die zweite Phase ist die Entwicklungsphase, in der die formalen und die ersten informalen Analysekre-terien entwickelt werden, um daraus eine Art Codebuch zu erarbeiten. Die Kriterien werden dabei unter Berücksichtigung der Kernhypothese und des erarbeiteten Vorwissens gebildet, sortiert und nach übergeordneten Kategorien strukturiert. Während der Durchführung der Analyse werden eventuell weitere Kriterien in das Codebuch aufgenommen. Nachdem das Codebuch verfasst worden ist, folgt eine kurze Testphase, um dieses beispielsweise anzupas-sen oder weiterführend zu ergänzen. Im Anschluss folgt die Anwendungs- oder auch Codier-phase in welcher Literatur erfasst, aufbereitet und in Hinblick auf die Fragestellung ausgewer-tet wird. Sowohl die Anwendungsphase als auch der letzte Schritt die Ergebnisdarstellung mit der Interpretation und der Berichterstattung der gewonnenen Erkenntnisse werden im nach-folgenden Abschnitt 3.1.2 behandelt.

**Planungsphase:** Ziel der Inhaltsanalyse ist es Begriffsmerkmale datengetriebener Prozes-se anhand von Analyse-kriterien zu erarbeiten und die Kernaussagen verschiedener Auto-ren\*innen zu datengetriebenen Prozessen zu erfassen und zu gewichten. Als Kernhypothese der Inhaltsanalyse wird daraus folgende Frage formuliert: Welche Erkenntnisse über datenge-

triebene Prozesse im Umfeld von Produktion und Logistik lassen sich aus verschiedenen literarischen Quellen entnehmen?

**Entwicklungsphase:** Zur Durchführung der Analyse werden formale und informale Analysekriterien erarbeitet, um ein Codebuch in tabellarischer Form zu entwickeln, welches der Strukturierung der Inhaltsanalyse dient. Um eine Vergleichbarkeit der Aussagen verschiedener Autor\*innen gewährleisten zu können, werden in Folge der Analyse formale Identifikationskennzahlen erfasst. Um die Verfasser\*innen zu identifizieren, werden ihre Namen erfasst. Darüber hinaus werden falls möglich die Fachdisziplinen der Verfasser\*innen notiert, um eventuelle Unterschiede in den Auffassungen verschiedener Fachdisziplinen vom Begriff Data Driven identifizieren zu können. Um die Quellen zu identifizieren, wird der Titel der Texte verwendet. Eine Kennzahl, die herangezogen werden kann, um die Verwendung des Begriffes Data Driven in verschiedenen Ländern abzugrenzen, ist das Land der Veröffentlichung. Um die Aktualität der Quellen zu bewerten, wird das Jahr der Veröffentlichung erfasst. Der Betrachtungszeitraum der zu analysierenden Quellen wird auf Veröffentlichungen ab dem 01.01.2018 festgelegt, damit der gegenwärtige Gebrauch des Begriffes Data Driven abgebildet werden kann. In Abschnitt 2.6 wurde beschrieben, wie die Relevanz von Quellen gemessen wird. Um in der Inhaltsanalyse die Relevanz zu messen, werden zum einen der Quellentyp erfasst (handelt es sich zum Beispiel um ein Journal oder ein Conference Paper) und zum anderen wird durch Zuhilfenahme von Literaturdatenbanken die Anzahl der Zitationen vermerkt. Um die Relevanz der analysierten Journale zu messen, wird für diese der JIF von 2020, bezogen auf die letzten zwei Jahre verwendet (Abschnitt 2.6), welcher vom Unternehmen Clarivate in ihren Journal Citation Reports ermittelt wurde (Clarivate, 2021). Da es sich bei datengetriebenen Prozessen in Produktion und Logistik um ein Thema handelt, welches aufgrund seiner Interdisziplinarität von Autoren\*innen verschiedener Fachdisziplinen behandelt wird und wie in Abschnitt 2.6 beschrieben es schwierig ist, Wissenschaftler\*innen unterschiedlicher Fachrichtungen mithilfe des h-Indexes zu vergleichen, wird dieser zum Ranken der Analysen kritisch betrachtet. Dennoch werden die jeweils höchsten h-Indizes der Autoren\*innen der einzelnen Analysetexte ermittelt. Für den Fall, dass die zuvor beschriebenen Bedenken nicht sehr in das Gewicht fallen, bietet die Verwendung des Indexes eine Chance, erfasste Conference Paper nicht nur auf Grundlage der Zitationsanzahl zu ranken, was ohnehin aufgrund der Aktualität vieler Dokumente differenziert betrachtet werden muss.

Aus den Überlegungen ergeben sich folgende formale Kennzahlen:

- Autorenkennzahlen:
  - a) Name
  - b) Fachdisziplin:
    - 01 Informatik
    - 02 Maschinenbau
    - 03 Wirtschaftswissenschaften
    - 04 Wirtschaftsinformatik

- 05 Andere
  - c) h-Index
  - Quellenkennzahlen:
    - a) Titel
    - b) Erscheinungsjahr: 2018-2021
    - c) Land der Veröffentlichung: Verwendung internationaler Ländercodes nach ISO 3166-1
    - d) Typ:
      - 01 Journal
      - 02 Lehrbuch
      - 03 Conference Paper
      - 04 Andere
    - e) Journal Impact Factor
    - f) Anzahl der Zitationen

Aus der verfassten Leitfrage der Planungsphase, den ersten theoretischen Überlegungen zum Begriff Data Driven aus Abschnitt 2.5 und auf Grundlage der im Kapitel 2 behandelten Nutzung von Daten innerhalb von Produktion und Logistik lassen sich präzisere prüfbare Fragen, durch eine theoriegeleitete Kriterienbildung (Abschnitt 2.6), formulieren, welche in der Inhaltsanalyse erfasst werden können. Um die Liste der Analysekriterien zu komplettieren, werden eventuell aus Textstichproben weitere gemeinsame Fragestellungen zum Thema Data Driven empiriegeleitet extrahiert, gebündelt und zu gemeinsamen Kriterien zusammengefasst, welche die zuvor erstellte Liste ergänzen. Im Folgenden werden die ersten Analysekriterien erarbeitet.

Wie in Abschnitt 2.2 beschrieben ist die Aufgabe von informationstechnischen Prozessen die Erfassung, Verarbeitung und Aus- oder Weitergabe von Daten und Informationen. Big Data Prozesse grenzen sich in den einzelnen Aufgabenbereichen unter anderem durch die 5Vs von etablierten IT-Prozessen ab, in denen vorrangig die Datennutzung im Vordergrund stehen (Abschnitt 2.4). Da sich datengetriebene Prozesse und Big Data, wie zuvor in Abschnitt 2.5 festgestellt, auf den ersten Blick sehr ähneln, erscheint es sinnvoll die Datennutzung datengetriebener Prozesse unter Berücksichtigung der 5Vs von Big Data zu analysieren, damit in diesem Punkt eventuelle Unterschiede festgestellt werden können. Als erstes Analysekriterium lässt sich somit die Datenmenge aufstellen. Bei der Analyse kann zum einen untersucht werden, ob in Texten explizit auf beispielsweise die Nutzung großer Datenmengen verwiesen wird und zum anderen kann, falls beschrieben, eine Größenordnung der genutzten Daten erfasst werden. Ein weiterer Punkt ist die Vielfalt der Datennutzung. Diese lässt sich in die Punkte Datentyp, -struktur und -quelle unterteilen. Mit dem Datentyp wird nicht der Wertebereich, sondern die Art der Objekte erfasst, welche mithilfe der Daten dargestellt werden, wie zum Beispiel Stamm- und Bewegungsdaten. Die Datenstruktur wurde bereits im Abschnitt

2.1 erläutert. Diese unterscheidet strukturierte, semistrukturierte und unstrukturierte Daten nach deren Nutzbarkeit datengetriebene Prozesse untersucht werden können. Datenquellen für digitale Daten sind vielfältig. Zur Unterscheidung verschiedener Datenquellen liefern Appelfeller und Feldmann, 2018 einen passenden Ansatz. Dieser unterteilt Datenquellen in vier Bereiche. Diese sind als erstes direkt eingegebene Daten, welche durch Mitarbeiter\*innen manuell in das System eingespeist werden. Als zweites werden übernommene Daten angeführt, die entweder aus dem eigenen oder einem fremden IT-System über eine passende Schnittstelle eingespeist werden. An dritter Stelle kommen analoge Daten, wie zum Beispiel Papierberichte oder Bilder, welche nachträglich digitalisiert wurden. Als letztes werden extrahierte Daten angeführt, welche auch häufig in Big Data Anwendungen eingesetzt werden und beispielsweise Maschinendaten, Smartdevices oder Internetplattformen umfassen. Neben der eventuellen Zuordnung eines Quellenbereiches kann es hilfreich sein, konkret genannte Quellen zu identifizieren, um daraus in der Analyse eine präferierte Quelle datengetriebener Prozesse hervorzuheben. Als letztes wird die Datenqualität in den Datenkriterien aufgenommen, damit untersucht werden kann, ob datengetriebene Prozesse besondere Anforderungen an die Qualität der zu verarbeiteten Daten stellen.

Die Verwendungszwecke und das Nutzen datengetriebener Prozesse sind auch ein Analysekriterium, welches von Interesse ist. In Kapitel 2.5 wurde unter anderem erläutert, dass datengetriebene Prozesse viele unterschiedliche Anwendungsbereiche haben, daher sollten diese in der Analyse erfasst werden. Da Big Data Anwendungen die Erschließung neuer Geschäftsmodelle ermöglichen, während etablierte IT-Prozesse alte Geschäftsmodelle optimieren (Abschnitte 2.2 und 2.4), wird zudem analysiert was von beiden ein datengetriebener Prozess ermöglicht. Ein weiteres entscheidendes Kriterium wird aus den Prozessebenen etablierter IT-Prozesse und der Unterteilung der Aufgabenbereiche in der Produktion und Logistik gewonnen (Abschnitt 2.2). Nämlich ob ein Prozess operativ an der unmittelbaren Durchführung von Prozessen beteiligt ist oder ob dieser eher analyseorientiert taktische oder strategische Aufgaben unterstützt. Ein weiteres Analysekriterium zur allgemeinen Verwendung datengetriebener Prozesse wird in Anlehnung an Appelfeller und Feldmann, 2018 als Art der Datenanalyse identifiziert. Diese lässt sich in vier Analysearten unterteilen: Die deskriptive Analyse behandelt was in der Vergangenheit passiert ist, dabei fasst sie Daten zusammen und bereitet diese übersichtlich und verständlich auf. Die diagnostische Analyse fragt danach, warum etwas geschehen ist, und verwendet dabei ebenfalls Daten aus der Vergangenheit, um Begründungen für bestimmte Ereignisse zu finden. Die prädikative Analyse analysiert was in der Zukunft geschehen wird. Beispielsweise durch die Zuhilfenahme von Datenmodellen und statistischen Methoden können so mit einer gewissen Wahrscheinlichkeit Aussagen über zukünftige Ereignisse getroffen werden. Die letzte Analyseart ist die präskriptive Analyse, welche untersucht wie auf Grundlage der Ergebnisse gehandelt werden sollte. Die präskriptive Analyse leitet den optimalen Lösungsweg aus den Analysen her.

Da die Abgrenzung des Begriffes Data Driven von den Begriffen Industrie 4.0, Big Data sowie von bereits etablierten IT-Prozessen für die Begriffsbestimmung von Bedeutung ist, werden auch explizit in den Analysetexten genannte Unterschiede in den Kriterien erfasst.

Aus den Überlegungen ergeben sich folgende erste Analyse Kriterien:

- 10 Daten allgemein
  - 11 Datenquelle (direkt eingegebene Daten, übernommene Daten, analoge Daten, extrahierte Daten)
  - 12 Datenstruktur (strukturierte Daten, semistrukturierte Daten, unstrukturierte Daten)
  - 13 Datentyp
  - 14 Datenmenge (Erwähnung große Datenmengen, Größe der Datenmenge)
  - 15 Datenqualität
- 20 Verwendung allgemein
  - 21 Anwendungs- oder Analyseorientiert (operativ, taktisch, strategisch)
  - 22 Analysesystem (deskriptiv, diagnostisch, prädikativ, präskriptiv)
  - 23 Optimierung alter- oder Erschließung neuer Geschäftsmodelle
  - 24 Anwendungsbereich
- 30 Eingrenzung allgemein
  - 31 Abgrenzung gegenüber etablierten IT-Prozessen
  - 32 Abgrenzung gegenüber Industrie 4.0
  - 33 Abgrenzung gegenüber Big Data Anwendungen

**Testphase:** In einer ersten Testphase wurden mit Hilfe, der zuvor erarbeiteten formalen und informalen Analyse Kriterien die ersten sechs Quellen der Literaturanalyse analysiert (die Auswahl der Quellen wird im Abschnitt 3.1.2 erläutert). Nach den ersten Einschätzungen sind die Kriterien für die Durchführung der Inhaltsanalyse geeignet, da in allen sechs Quellen Informationen zu fast jeden der Analyse Kriterien gewonnen werden konnten. Jedoch wird im Zusammenhang mit datengetriebenen Prozessen wenig über die Anforderungen an die Datenqualität geschrieben. Dennoch wird dieses Kriterium nicht aus der Analyseliste verworfen, da eventuelle Qualitätsansprüche an die Daten als wichtig erachtet werden, um die Datennutzung datengetriebener Prozesse genauer zu analysieren und um ein eventuell wichtiges Kriterium für eine spätere Abgrenzung zu anderen Begriffen nicht vorschnell zu verwerfen. Des Weiteren ist in der Testphase aufgefallen, dass datengetriebene Prozesse Methoden und Technologien (z. B. KI und AutoID) verwenden, welche auch von Big Data und der Industrie 4.0 genutzt werden. Um dies genauer analysieren zu können, wird unter Punkt 40 Methoden und Technologien ein weiteres Analyse Kriterium aufgenommen, welches Methoden und Technologien erfasst, die von den analysierten datengetriebenen Prozessen genutzt werden.



### 3.1.2 Durchführung und Ergebnisdarstellung der Literaturanalyse zu datengetriebenen Prozessen

Im Abschnitt 3.1.1 wurde zuvor das Vorgehen und das Ziel der Inhaltsanalyse erläutert und ein Kategoriensystem aufgebaut, welches bei der Analysedurchführung genutzt wurde, um die Inhalte nach den gleichen Aspekten hin zu untersuchen. Im Folgenden wird die Durchführung der Analyse beschrieben, um in Anschluss die Ergebnisse zu präsentieren.

**Anwendungsphase:** Bei der Durchführung der Inhaltsanalyse wurden mögliche Aussagen über die Analysekriterien erfasst, um daraus einheitliche Aussagen über datengetriebene Prozesse abzuleiten. Dabei wurde teilweise gezielt geprüft, ob Autoren\*innen Aussagen über bestimmte Kriterien treffen, um mögliche Übereinstimmungen zu finden. Zudem wurden weitere Angaben zu Kriterien Stichpunktartig aus der Literatur eruiert. Unter Zuhilfenahme der Kriterien und Randnotizen wurden abschließend die Kernaussagen der Autoren\*innen zum Begriff Data Driven in eigenen Worten zusammengefasst. Wie in Abschnitt 3.1.1 bereits erwähnt wurden ausschließlich Quellen erfasst, welche nach dem 01.01.2018 veröffentlicht wurden, um eine möglichst aktuelle Verwendung des Begriffes Data Driven zu untersuchen. Dies hat jedoch eine Rückwärtssuche weiterer relevanter Quellen erschwert, weil sich diese oft außerhalb des Analysezeitintervalles befanden. Da allein die Onlineliteraturdatenbank Scopus, für den festgelegten Zeitraum unter dem Suchbegriff Data Driven 11.198 Dokumente und begrenzt auf den Bereich der Ingenieurwissenschaften immer noch 5.335 Dokumente führte, konnte die Analyse lediglich an einer Stichprobe durchgeführt werden. Vor dem Hintergrund des Ziels qualitative Aussagen über den Begriff Data Driven zu erfassen und der damit einhergehenden tiefergehenden Auseinandersetzung der textuellen Inhalte, wurde die Anzahl der Dokumente nach einem ersten flüchtigen Lesen der Zusammenfassungen ohnehin auf einige wenige als interessant erachtete Dokumente reduziert. Diese wurden ausgewählt, weil sie überwiegend auf datengetriebene Anwendungen in der Produktion und Logistik eingehen und darüber hinaus den Begriff Data Driven nicht nur im Titel verwenden, sondern auch innerhalb des Textes auf diesen eingehen. Die Stichprobe hat einen Umfang von 28 Inhalten, welche sich aus Artikeln wissenschaftlicher Fachzeitschriften (17 Inhalte) und Publikationen aus Bänden wissenschaftlicher Konferenzen (11 Inhalte) zusammensetzt. Zur Suche geeigneter Literatur wurden zum einen der Katalog plus der Universitätsbibliothek der technischen Universität Dortmund und zum anderen weitere wissenschaftliche Datenbanken, darunter Scopus, Google Scholar und das Web of Science herangezogen. Um wichtige englischsprachige Literatur zu identifizieren, wurde diese in den genannten Datenbanken unter Verwendung der Suchbegriffe “Data Driven“, “Data Driven Manufacturing“, “Data Driven Production“, “Data Driven Supply Chain” und “Data Driven Logistics” ermittelt. In Tabelle 1 ist die gesamte ermittelte Literatur inklusive Quellenangaben getrennt nach Artikeln aus Konferenzbänden (CP) und Fachzeitschriften (J) sortiert nach den Zitationsanzahlen und dem JIF der Fachzeitschriften, in denen diese Artikel erschienen sind, aufgelistet.

Tabelle 1: Quellen der Literaturanalyse

Titel	Jahr	Autor	Typ	Quelle
Local Koopman Operators for Data Driven Control of Robotic Systems	2019	Mamakoukas, Giorgos; Castano, Maria; Tan, Xiaobo; et al.	CP	(Mamakoukas et al., 2019)
Data Driven management in Industry 4.0: a method to measure Data Productivity	2018	Miragliotta, Giovanni; Sianesi, Andrea; Convertini, Elisa; et al.	CP	(Miragliotta, G. et al. 2018)
A data- and knowledge-driven framework for digital twin manufacturing cell	2019	Zhang, Chao; Zhou, Guanghui; He, Jun; et al.	CP	(Zhang et al., 2019a)
Digital Twin for Machining Tool Condition Prediction	2019	Qiao, Qianzhe; Wang, Jinjiang; Ye, Lunkuan; et al.	CP	(Qiao, Q. et al. 2019)
The Usefulness of Data Driven, Intelligent Agent-Based Modelling for Transport Infrastructure Management	2018	Olusola. T. Faboya; Graziela P. Figueredo; Brendan Ryan; et al.	CP	(Olusola. T. Faboya et al. 2018)
RFID Data Driven Performance Evaluation in Production Systems	2019	Zhong, Ray Y.	CP	(Zhong, R. 2019)
Data Analytics for Manufacturing Systems – A Data Driven Approach for Process Optimization	2019	Ungermann, Florian; Kuhnle, Andreas; Stricker, Nicole; et al.	CP	(Ungermann et al., 2019)
A Data Driven Approach for Identification and Detection of Intermittent Faults	2021	Na, Yuhong; Peng, Shichen	CP	(Na und Peng, 2021)
A Pattern-moving-Based Data Driven Control Method for a Kind of Industrial Production Processes	2020	Li, Xiangquan; Xu, Zhengguang; Wang, Mushu; et al.	CP	(Li et al., 2020)
Data Driven Manufacturing Simulation: Towards a CPS-Based Approach	2020	Jeong, Yongkuk; Singh, Amita; Zafarza-deh, Masoud; et al.	CP	(Jeong et al., 2020)
Data Driven detection of moving bottlenecks in multi-variant production lines	2018	Roh, P.; Kunz, A.; Netland, T.	CP	(Roh, P. et al. 2018)
Manufacturing as a Data Driven Practice: Methodologies, Technologies, and Tools	2021	Cerquitelli, Tania; Pagliari, Daniele Jahier; Calimera, Andrea; et al.	J	(Cerquitelli et al., 2021)
Data Driven sustainable intelligent manufacturing based on demand response	2020	Ma, Shuaiyin; Zhang, Yingfeng; Liu, Yang; et al.	J	(Ma et al., 2020)
Data Driven smart manufacturing	2018	Tao, Fei; Qi, Qinglin; Liu, Ang; et al.	J	(Tao et al., 2018b)
From data to big data in production research: the past and future trends	2019	Kuo, Yong-Hong; Kusiak, Andrew	J	(Kuo und Kusiak, 2019)

Big Data Driven supply chain performance measurement system: a review and framework for implementation	2020	Kamble, Sachin S.; Gunasekaran, Angappa	J	(Kamble und Gunasekaran, 2020)
Data Driven operations and supply chain management: established research clusters from 2000 to early 2020	2021	Nguyen, Duy Tan; Adulyasak, Yossiri; Cordeau, Jean-François; et al.	J	(Nguyen et al., 2021)
A New Convolutional Neural Network-Based Data Driven Fault Diagnosis Method	2018	Wen, Long; Li, Xinyu; Gao, Liang; et al.	J	(Wen et al., 2018)
A digital supply chain twin for managing the disruption risks and resilience in the era of Industry 4.0	2021	Ivanov, Dmitry; Dolgui, Alexandre	J	(Ivanov und Dolgui, 2021)
Deep Transfer Network with Joint Distribution Adaptation: A New Intelligent Fault Diagnosis Framework for Industry Application	2020	Han, Te; Liu, Chao; Yang, Wenguang; et al.	J	(Han et al., 2020)
A Data Driven algorithm to predict throughput bottlenecks in a production system based on active periods of the machines	2018	Subramaniyan, Mukund; Skoogh, Anders; Salomonsson, Hans; et al.	J	(Subramaniyan et al., 2018)
Data Driven multi-scale multi-physics models to derive process-structure-property relationships for additive manufacturing	2018	Yan, Wentao; Lin, Stephen; Kafka, Orion L.; et al.	J	(Yan et al., 2018)
Data Driven Monitoring and Safety Control of Industrial Cyber-Physical Systems: Basics and Beyond	2018	Jiang, Yuchen; Yin, Shen; Kaynak, Okyay	J	(Jiang et al., 2018)
Digital twin-driven product design, manufacturing and service with big data	2018	Tao, Fei; Cheng, Jiangfeng; Qi, Qinglin; et al.	J	(Tao et al., 2018a)
Data Driven smart production line and its common factors	2019	Zhang, Yongping; Cheng, Ying; Wang, Xi Vincent; et al.	J	(Zhang et al., 2019b)
Data Driven Additive Manufacturing Constraints for Topology Optimization	2021	Weiss, Benjamin M.; Hamel, Joshua M.; Ganter, Mark A.; et al.	J	(Weiss et al., 2021)
Advanced Data Collection and Analysis in Data Driven Manufacturing Process	2020	Xu, Ke; Li, Yingguang; Liu, Changqing; et al.	J	(Xu et al., 2020)
Freight-Logistics-as-a-Service – Innovative Geschäftsmodelle für ein datengetriebenes Transportmanagement	2021	Heinbach, Christoph; Hagen, Simon; Thomas, Oliver	J	(Heinbach et al., 2021)

Die aufgeführten Artikel wurden insgesamt von Autoren\*innen aus zwölf verschiedenen Ländern verfasst, wobei der Großteil der Verfasser\*innen aus China kommt, gefolgt von den USA und Deutschland. Unter den Wissenschaftlern\*innen befinden sich Vertreter\*innen un-

terschiedlicher Fachdisziplinen, darunter unter anderen Maschinenbau- und Industrieingenieure\*innen, Logistiker\*innen, Informatiker\*innen und Regelungstechniker\*innen. Eine genaue Auflistung aller formaler Kennzahlen kann der Tabelle: Gesamtübersicht der Inhaltsanalyse entnommen werden, welche sich im digitalen Anhang der Arbeit befindet. Darunter befinden sich auch die Angaben zu den Zitationsanzahlen, den JIFs und den h-Indizes.

Wie zuvor im Abschnitt 3.1.1 beschrieben, wurden zum **Ranken** der Literatur die Anzahl der Zitationen und bei den Fachzeitschriften zusätzlich der JIF vom Jahr 2020 bezogen auf die letzten zwei Jahre ermittelt. Der JIF soll gewährleisten, besonders bei der späteren qualitativen Analyse der Kernaussagen (Abschnitt 3.1.3), Artikel aus möglichst qualitativ hochwertigen Journalen zu berücksichtigen. Eine Berechnung des JIF für Konferenzbände wäre möglich, jedoch sind diese in Bezug auf Häufigkeit und Inhalt unregelmäßig, weshalb dies nicht vorgenommen wird. Darüber hinaus haben Konferenzbände ohnehin eine höhere Relevanz (Abschnitt 2.6). Aufgrund einer sehr ähnlichen Verteilung der h-Indizes innerhalb der unterschiedlichen Fachdisziplinen der Autoren\*innen, wird dieser ebenfalls zur Messung der Relevanz herangezogen. Die JIFs der Fachzeitschriften, in denen die Artikel erschienen sind, umfasst eine Spannweite von 1,963 bis hin zu einem Faktor von 10,961. Dies entspricht in etwa der Spannweite der gesamten JIF vom Jahr 2020, bezogen auf die Kategorien Engineering, Industrial and Engineering, Mechanical, welche von 0,080 bis zu 10,998 reicht. Zur Berücksichtigung möglichst qualitativer Zeitschriften werden vor allem in der späteren Auswertung der Kernaussagen überwiegend diejenigen betrachtet, deren JIF höher als 8,000 ist, was in etwa dem obersten Quartil des gesamten JIF, der betrachteten Kategorien, entspricht. Dies trifft auf sieben der 17 Zeitschriften zu. Zur Berechnung der vier relevantesten Artikel aus Fachzeitschriften und Konferenzbänden, wurde folgende Formel zur Berechnung erstellt:

$$\text{Rangwert}_q = \left( \frac{JIF_q}{\frac{1}{n} \sum_{i=1}^n JIF_i} \right) + \frac{Zitate_q}{\frac{1}{n} \sum_{i=1}^n Zitate_i} + \frac{h_q}{\frac{1}{n} \sum_{i=1}^n h_i}$$

Die Formel berechnet den Rangwert einzelner Quellen (q). Je höher dieser Wert ausfällt, umso relevanter wird eine Quelle eingeschätzt. Die Formel beachtet zwar nicht, dass Artikel aus Konferenzbänden keinen JIF zugewiesen haben, jedoch werden diese aufgrund der allgemeinen höheren wissenschaftlichen Qualität (vgl. Abschnitt 2.6) gesondert betrachtet, weshalb die Formel in diesem Fall geeignet ist. Zur Berechnung werden der JIF (*JIF*), die Zitationsanzahl (*Zitate*) und der h-Index (*h*) einer spezifischen Quelle (q) jeweils durch das arithmetische Mittel der gleichen Referenzfaktoren aller erfassten Quellen (n) geteilt und deren Summe gebildet. Bei der Berechnung der Conference Paper entfällt der erste Summand, da diesen kein JIF zugerechnet wird. Aus diesem Grund und weil sie unabhängig voneinander berechnet werden, sind die Rangwerte der Journale und der Conference Paper nicht direkt miteinander vergleichbar. Die getrennten Berechnungen der Rangwerte für beide Quellentypen wurde un-

ter anderem auch deshalb als wichtig erachtet, da sich die Fachzeitschriftenartikel in der durchschnittlichen Zitationsanzahl signifikant von der der Konferenzbänden unterschied. Der Vorteil der Formel ist, dass beim Ranken die Werte aller Referenzfaktoren berücksichtigt werden. Zu beachten ist jedoch, dass im Falle von größeren Ausreißern innerhalb eines Faktors, diese in der Berechnung stärker gewichtet werden. Im Falle der in der Inhaltsanalyse behandelten Quellen (Tabelle 1) trifft dies auf die Anzahl der Zitationen zu. Nach dem Ausschluss aller Quellen mit einem JIF unter 8,000 und der Berechnung der Rangwerte der verbleibenden Fachzeitschriften und der anschließenden Berechnung der Rangwerte der Artikel aus Konferenzen, werden folgende sechs Quellen als am relevantesten identifiziert. Eine Liste sämtlicher Rangwerte befindet sich im Anhang (Tabelle A. 1: Ranking aller Quellen).

**Tabelle 2: die sechs relevantesten Quellen**

Titel	Name	Typ	JIF	Zitationen	h-Index	Rangwert
A New Convolutional Neural Network-Based Data Driven Fault Diagnosis Method	Wen, Long; Li, Xinyu; Gao, Liang; et al.	J	8,236	588	56	5,486
Data Driven smart manufacturing	Tao, Fei; Qi, Qinglin; Liu, Ang; et al.	J	8,633	443	65	4,808
Big Data Driven supply chain performance measurement system: a review and framework for implementation	Kamble, Sachin S.; Gunasekaran, Angappa	J	8,568	35	85	2,673
Local Koopman Operators for Data Driven Control of Robotic Systems	Mamakoukas, Giorgos; Castano, Maria; Tan, Xiaobo; et al.	CP	-	25	36	5,297
Digital Twin for Machining Tool Condition Prediction	Qiao, Qianzhe; Wang, Jinjiang; Ye, Lunkuan; et al.	CP	-	21	45	5,211
A data- and knowledge-driven framework for digital twin manufacturing cell	Zhang, Chao; Zhou, Guanghui; He, Jun; et al.	CP	-	13	16	2,614

Zur Analyse der Textinhalte wurden im Kapitel 3.1.1 informale Analyse Kriterien erarbeitet. Während der Analyse der Dokumente wurden zwei weitere informale Analyse Kriterien identifiziert, welche in die Untersuchung übernommen wurden. Zum ersten wurde unter 16 Datengeschwindigkeit geprüft, ob in der Literatur im Zusammenhang mit datengetriebenen Prozessen Aussagen über eine Echtzeiterfassung, -übertragung oder -verarbeitung von Daten getroffen wurden. Als zweites wurde unter 25 Rückkopplung analysiert, ob die Ergebnisse datengetriebener Prozesse zu den Prozessen deren Daten sie nutzen, rückgekoppelt werden, um somit eine automatische Prozessoptimierung oder eine Prozesssteuerung zu realisieren. Darüber hinaus wurden in den Inhaltsanalysen die Keywords der Texte erfasst mit der Intention häufig genannte Begriffe im Zusammenhang mit dem Begriff Data Driven zu identifizieren. In den folgenden Tabellen (Tabelle 3, Tabelle 4, Tabelle 5) sind die informale Analyse Kriterien dargestellt, wie sie bei der Durchführung der Literaturanalyse verwendet wurden:

**Tabelle 3: Informale Kriterien Daten**

Daten								
Datenquelle				Datenstruktur	Datentyp	Datenmenge	Datenqualität	Datengeschwindigkeit
eingegeben	übernommen	analog	extrahiert	strukturiert	semistrukturiert	unstrukturiert	Quellen	Echtzeiterfassung und -verarbeitung
				strukturiert	semistrukturiert	unstrukturiert	große Mengen	
							Größenordnung	

**Tabelle 4: Informale Kriterien Verwendung & Methoden und Technologien**

Verwendung							Methoden und Technologien
Anwendungsbereich	Aufgabenorientierung			Analysesystem			
	operativ	taktisch	strategisch	deskriptiv	diagnostisch	prädiktiv	präskriptiv

**Tabelle 5: Informale Kriterien Eingrenzung & Keywords**

Eingrenzung			Keywords
Abgrenzung zu etablierten IT-Prozessen	Abgrenzung zu Industrie 4.0	Abgrenzung zu Big Data	

Die Kriterien teilen sich in die Hauptkategorien Daten (Tabelle 3), Verwendung, Methoden und Technologien (Tabelle 4), Eingrenzung und Keywords (Tabelle 5) auf. Die senkrechten Kategorien wurden in der Analyse quantitativ erfasst, während zu den übrigen Kriterien stichpunktartig Quellenaussagen festgehalten wurden. Über die Verwendung der Kriterien hinaus wurden die Kernaussagen der Dokumente zu dem Begriff Data Driven zusammengefasst. Mit der Auswertung der Kernaussagen beschäftigt sich der kommende Abschnitt 3.1.3. Auch die Kriterien zur Eingrenzung des Begriffes Data Driven aus Tabelle 5, werden später in Abschnitt 3.4 Abgrenzung von Data Driven zu anderen Begriffen behandelt, da diesen eine größere Bedeutung bei der späteren Begriffsbestimmung datengetriebener Prozesse zugesprochen wird.

**Ergebnisdarstellung:** Um einen ersten Eindruck über datengetriebene Prozesse zu vermitteln, dienen im Folgenden die aus den einzelnen Texten ermittelten **Keywords** (Tabelle 5), welche stichpunktartig die Thematiken der Dokumente wiedergeben und die zusammengetragenen **Methoden und Technologien** (Tabelle 4), die laut Autoren\*innen im Zusammenhang mit datengetriebenen Prozessen verwendet werden. Zur Informationsvisualisierung wurde aus den eruierten Begriffen zwei Word-Clouds erstellt, welche Begriffe umso größer darstellt je häufiger sie in der Tabelle erfasst wurden. Die verschiedenen Farben dienen lediglich dazu Begriffe mit der gleichen Anzahl an Nennungen schneller zu identifizieren. Die Darstellungen sollen nur zum Einstieg dienen, eine genaue tabellarische Auflistung befindet sich in der Tabelle: Gesamtübersicht der Inhaltsanalyse im digitalen Anhang.



Abbildung 9: Word-Cloud Keywords

Abbildung 9 veranschaulicht die Keywords der Dokumente. Besonders auffällig ist, dass der Begriff Big Data, nach Data Driven am häufigsten benannt ist. Daraus lässt sich schlussfolgern, dass beide Begriffe sich nicht nur sehr ähneln (vgl. Abschnitt 2.5), sondern auch wahr-

scheinlich miteinander in Beziehung stehen. Ein weiterer Begriff ist Data Analytics, welcher eine Reihe überwiegend statistische Methoden in der Informationstechnik darstellt, die eine kontrollierte Suche nach bedeutsamen Inhalten in Daten ermöglichen (Kuo und Kusiak, 2019). Smart Manufacturing ist ein Begriff der Industrie 4.0, in dem datengetriebene Prozesse auf den ersten Blick auch einen Nutzen haben. Fertigungssysteme und das Supply Chain Management sind die Wirkungsbereiche, die in den Analysen hauptsächlich untersucht wurden. Maschinelles Lernen und digitale Zwillinge sind zwei Schlagwörter, welche ebenso häufig in den Methoden und Technologien fallen (Abbildung 10). Die Nutzung von Methoden des maschinellen Lernens, insbesondere Deep Learning, haben laut Analysen eine größere Bedeutung für datengetriebene Prozesse.

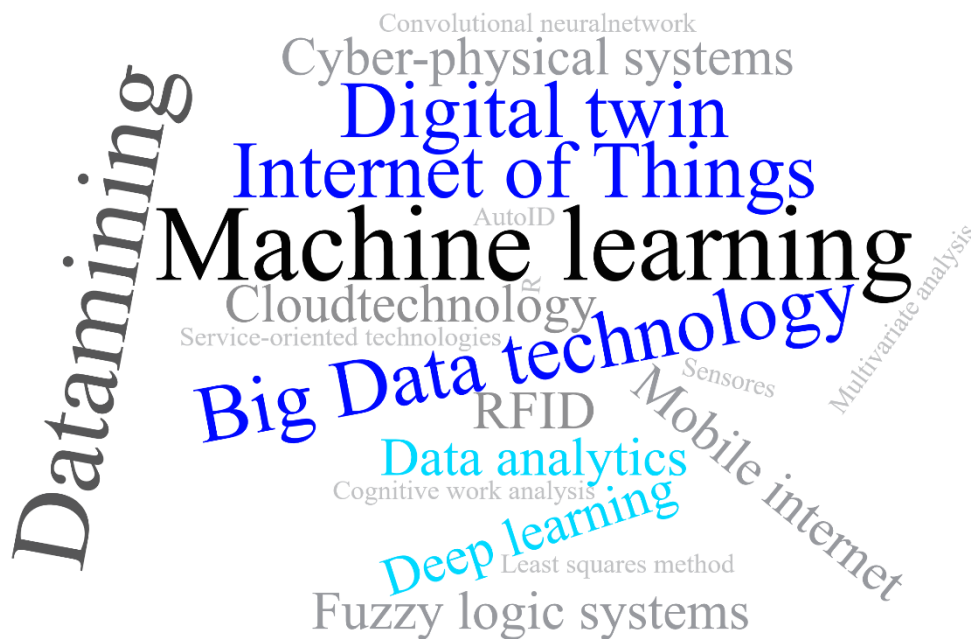


Abbildung 10: Word-Cloud Methoden und Technologien

Häufig werden in der Literatur auch die Wörter Digital Twin Driven (Tao et al., 2018a) oder auch Big Data Driven (Kamble und Gunasekaran, 2020) verwendet. Dies zeigt, wie sehr diese Begriffe miteinander verknüpft sind und lässt die Vermutung nahe, dass datengetriebene Ansätze eine Bedeutung in den Bereichen digitaler Zwilling und Big Data haben. Datamining ist eine Anwendung aus dem Bereich Data Analytics, welche sich vor allem mit der Analyse großer Datenmengen beschäftigt. Auch Technologien des Internet of Things (IoT), vor allem AutoID Systeme (Abschnitt 2.3) und insbesondere RFID, werden scheinbar häufiger in datengetriebenen Systemen eingesetzt. Weitere in Abbildung 10 enthaltene Technologien zur Datenerfassung sind mobiles Internet oder auch Cloudtechnologien, welche IT-Architekturen bereitstellen, die ein dezentrales Bevorraten, Analysieren und Abrufen von Daten ermöglichen (Hilber, 2014). Zusammenfassend lässt sich eine erste Vermutung zu datengetriebenen Prozessen im Bereich Produktion und Logistik aufstellen nämlich, dass sie größere Mengen an Daten mithilfe von Sensoren (RFID), mobilen Internet oder Cloudtechnologien erfassen



und diese mit Methoden und Technologien aus den Bereichen Data Analytics, ML, Big Data und dem Cloudcomputing analysieren, um einen Nutzen zu generieren. Um den Begriff Data Driven genauer zu untersuchen und um zu prüfen, inwieweit sich die erste Vermutung bestätigt, wurden Aussagen zu datengetriebenen Prozessen insbesondere in den Bereichen Daten und Verwendung analysiert.

Zur Analyse des Kriteriums **Daten** diene Tabelle 3. Als erstes wurde geprüft ob und welche Aussagen zu den Datenquellen datengetriebener Prozesse getroffen wurden. Dazu wurde quantitativ gemessen, welche Arten von Datenerfassung genannt wurden und zusätzlich wurden explizit genannte Datenquellen der Literatur entnommen. Zur Veranschaulichung der Messung dient das folgende Diagramm (Abbildung 11):

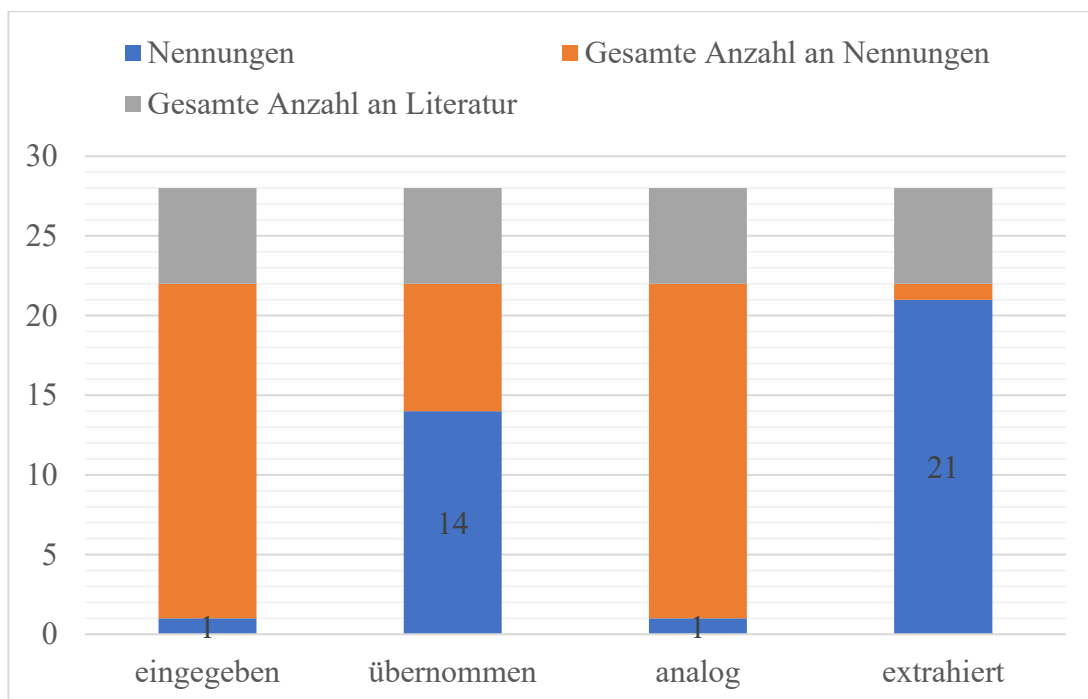
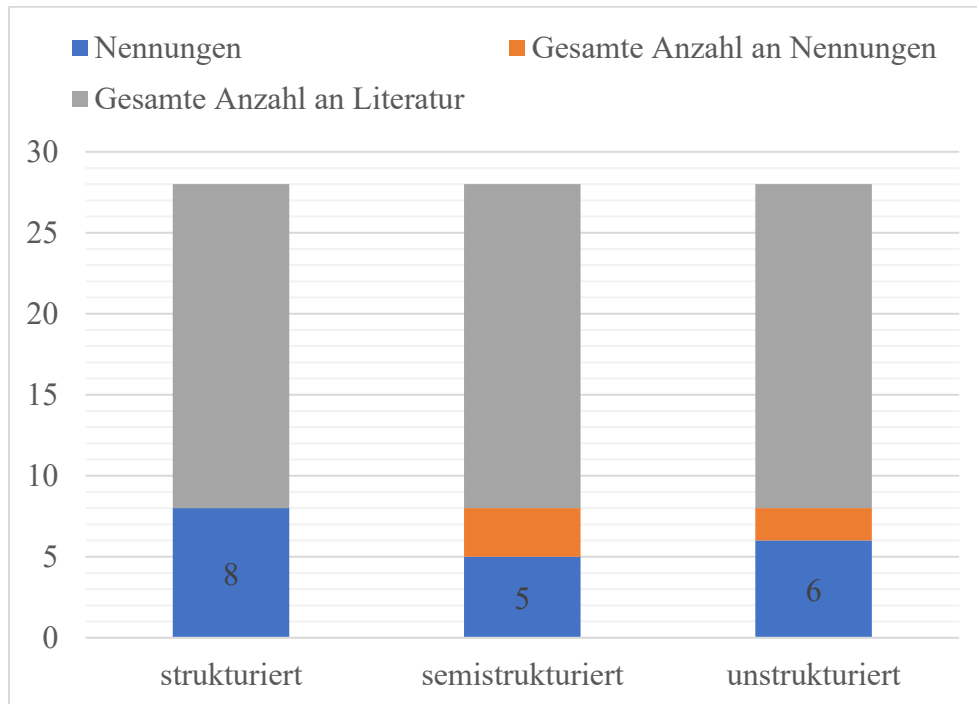


Abbildung 11: Diagramm Datenerfassung

Insgesamt wurden in 22 von den 28 analysierten Texten Angaben zur Art der Datenerfassung datengetriebener Prozesse gemacht. Wie aus Abbildung 11 ersichtlich, spielt laut Quellen die manuelle Eingabe von Daten und die Digitalisierung analoger Daten eine eher untergeordnete Rolle. Vor allem die Aufnahme von Daten durch Extraktion aus physischen Prozessen (21 von 22 Nennungen), scheint primär in datengetriebenen Prozessen relevant zu sein. Aber auch übernommene Daten aus internen oder auch externen IT-Systemen (14 von 22 Nennungen) werden, zumindest in Verbindung mit extrahierten Daten, genannt. Die meistgenannten Datenquellen sind Sensoren, welche auch von fünf der höchstgerankten Quellen (Tabelle 2) angeführt werden. Auch RFID Systeme werden zur Datenextraktion häufig angegeben. Als Quellen zur Datenübernahme werden meist eigene IT-Systeme (MES, ERP) zum Nutzen historischer Systemdaten, aber auch Clouddaten unter anderem zum Anlernen von künstlichen

Intelligenzen genannt. Zu der Verarbeitung verschiedener Datenstrukturen werden insgesamt weniger Angaben gemacht, nämlich lediglich von sieben Quellen. Im folgenden Diagramm (Abbildung 12) werden die Messungen zu den Aussagen über die Nutzung unterschiedlicher Datenstrukturen von datengetriebenen Prozessen dargestellt:

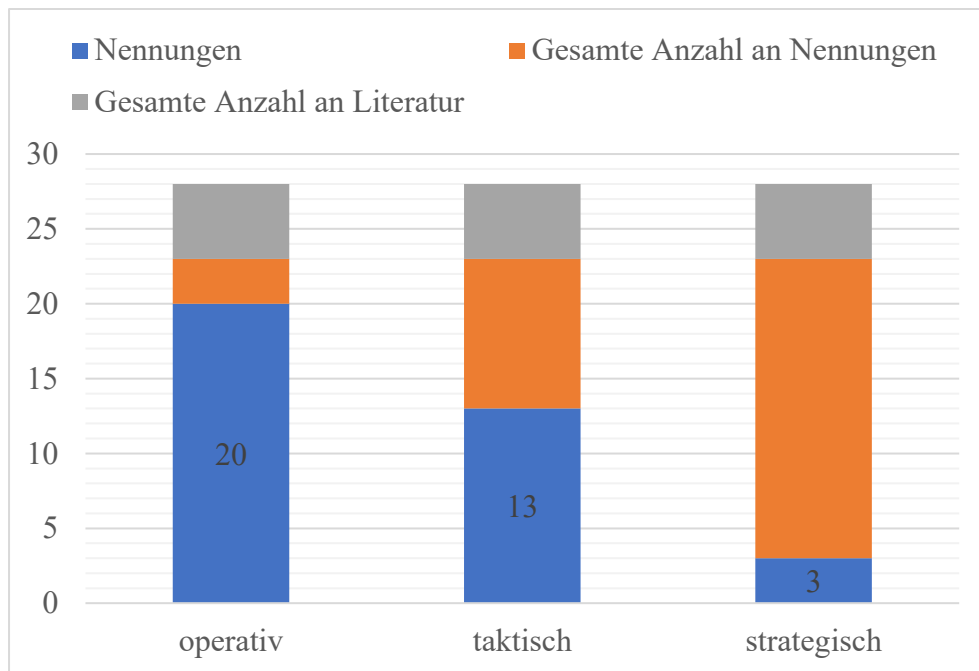


**Abbildung 12: Diagramm Datenstruktur**

Auffallend ist, dass sechs von acht Angaben zu den von datengetriebenen Prozessen nutzbaren Datenstrukturen, die Nutzung von unstrukturierten Daten als eine Eigenschaft beschreiben. Darunter befinden sich auch zwei der relevantesten Quellen. Obwohl datengetriebene Prozesse in der Lage zu sein scheinen semistrukturierter und unstrukturierter Daten zu verarbeiten, wird dies aufgrund der wenigen Aussagen eventuell als eher hintergründig betrachtet. Um genauere Aussagen zu der Verarbeitung unterschiedlicher Datenstrukturen treffen zu können, müssten weitere Daten erhoben werden. Im Bereich der Produktion werden, im Zusammenhang mit dem Begriff Data Driven, überwiegend die Aufnahme von Rohdaten aus Prozessen (z. B. Maschinendaten) und die Übernahme von historischen Daten der Prozesse thematisiert. Für regelungstechnische Aufgaben bieten datengetriebene Ansätze darüber hinaus die Möglichkeit nichtlineare dynamische Daten zu verarbeiten (Mamakoukas et al., 2019). Im Bereich des SCM werden unterschiedliche Daten verarbeitet (z. B. Bewegungsdaten, Produktdaten), welche Daten von datengetriebenen Prozessen im SCM verarbeitet werden hängt aber von der spezifischen Aufgabenstellung ab. Im Zusammenhang mit der Datenaufnahme scheint besonders der Erfassung von Daten in nahezu Echtzeit eine Bedeutung zugesprochen zu werden. Immerhin 18 der 28 Quellen (darunter fünf der sechs am höchsten gerankten) machen explizite Angaben dazu, dass datengetriebene Prozesse Daten in nahezu Echtzeit erfassen.

sen und gegeben falls auch verarbeiten. Ob die Möglichkeit große Datenmengen zu verarbeiten als ein Merkmal datengetriebener Prozesse gilt, muss kritischer betrachtet werden. Zwar machen 16 der 28 Quellen Angaben zur Verarbeitung großer Datenmengen, jedoch kann dies, neben den vermehrten Einsatz von AutoID Systemen, auch auf die Verbindung datengetriebener Prozesse und Technologien aus dem Bereich Big Data zurückzuführen sein. In diesem Zusammenhang wird auch häufiger der Begriff Big Data Driven erwähnt, zum Beispiel von Kamble und Gunasekaran, 2020. In der Literatur werden auch keine genauen Angaben zur Größenordnung aufnehmbarer und verarbeitbarer Daten gemacht. Selbst wenn datengetriebene Ansätze große Datenmengen verarbeiten können, scheinen sich diese nicht primär darüber zu definieren, wie im Fall von Big Data (Abschnitt 2.4). Die Anforderungen an die Datenqualität der zu verarbeitenden Daten hat sich scheinbar durch den Einsatz neuer Technologien auch gewandelt. So schreiben Wen et al., 2018, S. 5990-5992: „*However, the traditional Data Driven fault diagnosis methods rely on the features extracted by experts. The feature extraction process is an exhausted work and greatly impacts the final result. [...] In traditional Data Driven fault diagnosis methods, the data preprocessing method is vital since most of the Data Driven methods cannot handle the raw signals directly.*“ So scheint die Verwendung neuer Methoden im Bereich der datengetriebenen Prozesse auch die Anforderungen an die Datenvorverarbeitung von Rohdaten zu beeinflussen, welche nicht mehr mühsam durch Experten, sondern anderweitig durchgeführt wird. Die Vorverarbeitung von Daten bleibt jedoch weiterhin wichtig, um überflüssige, irreführende, doppelte und widersprüchliche Informationen zu entfernen (Tao et al., 2018b).

Zur Beantwortung der Frage zu welchem Zweck diese Daten in datengetriebenen Prozessen verwendet werden, wurden die Analysetexte unter Verwendung der informalen Analyse-kriterien aus Tabelle 4 auf ihre allgemeine **Verwendung** hin untersucht. Wie bei der Analyse der Daten wurden hierzu sowohl quantitativ als auch qualitativ Daten erhoben. Die Anwendungsbereiche datengetriebener Prozessen sind vielfältig, so werden diese beispielsweise beim Produktdesign, der Prozessoptimierung und der Prozessüberwachung eingesetzt (Kuo und Kusiak, 2019). Im Bereich der Produktion wird in der Literatur oft über den Einsatz datengetriebener Prozesse und Modelle zur Fehlerdiagnose und -prognose, zur Prozesssteuerung und -überwachung und zur Engpassüberwachung geschrieben. Unter anderem sind auch die Rolle bei der Realisierung intelligenter Fabriken und die Nutzung datengetriebener Prozesse zur Realisierung digitaler Zwillinge zwei zukunftssträchtige Felder, welche von einigen Quellen thematisiert werden. Im Bereich der Logistik werden überwiegend operative und taktische Aufgaben im SCM erwähnt. Auch bei regelungstechnischen Aufgaben scheinen datengetriebene Prozesse, vor allem in dynamischen nichtlinearen Systemen, Vorteile zu bieten. Um Prozesse zu optimieren oder auch zu steuern, können datengetriebene Prozesse laut zehn Quellen (darunter drei der relevantesten Quellen) ihre Ausgabewerte zum eigentlichen Prozess rückkoppeln. Allgemein scheint sich die Aufgabenorientierung datengetriebener Prozesse vornehmlich auf operative Anwendungsbereiche zu konzentrieren.



**Abbildung 13: Diagramm Aufgabenorientierung**

Aus Abbildung 13 wird ersichtlich, dass datengetriebene Prozesse überwiegend für ausführungorientierte Aufgaben eingesetzt werden (20 von 23 Nennungen) und somit einen unmittelbaren Einfluss auf beispielsweise Produktionsprozesse oder Transporte ausüben. Darüber hinaus werden sie auch zur Unterstützung mittelfristiger Planungsaufgaben genutzt (13 von 23 Nennungen), was anscheinend aber nicht das Haupteinsatzgebiet datengetriebener Prozesse ist. Strategische beziehungsweise langfristige Planungsaufgaben können eventuell datengetrieben unterstützt werden, allerdings wird dies in der Literatur nicht gründlich diskutiert (3 von 23 Nennungen). Welchen Nutzen datengetriebene Prozesse in den einzelnen Aufgabenbereichen bieten wurde durch die Betrachtung ihres Analysesystems erforscht. Dabei wurde zwischen deskriptiven, diagnostischen, prädiktiven und präskriptiven Analysen unterschieden.

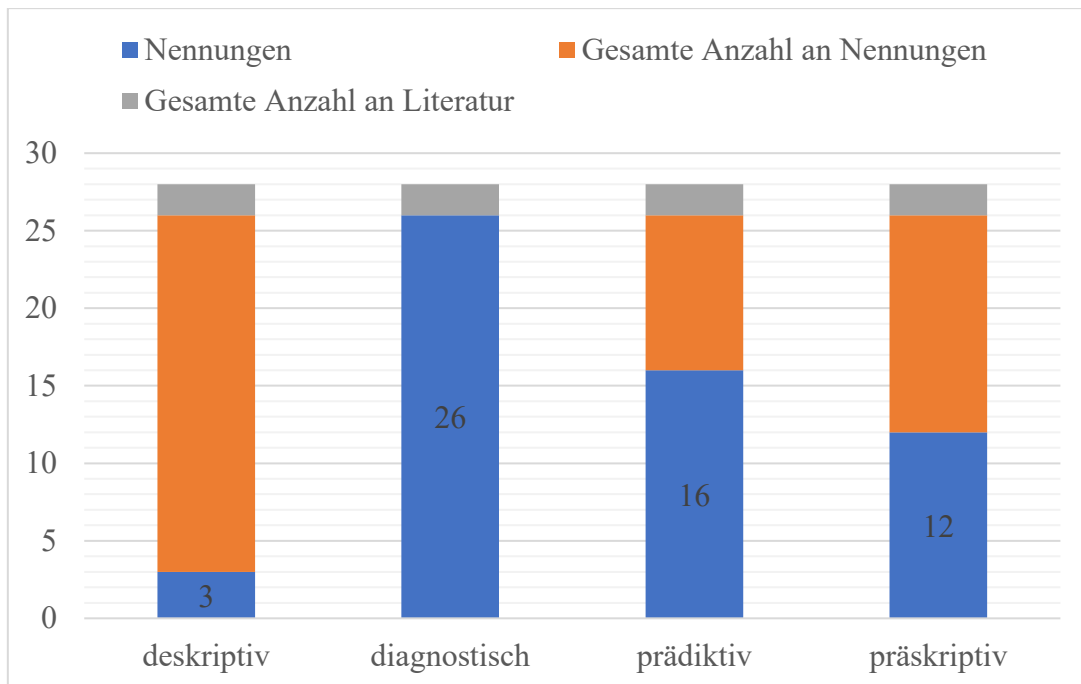


Abbildung 14: Diagramm Analysesystem

Durch Auswertung der Inhaltsanalyse wird ersichtlich, dass datengetriebene Prozesse nicht deskriptiv (3 von 26 Nennungen), sondern vor allem zur Diagnose eingesetzt werden (26 von 26 Nennungen). Dabei werden Daten analysiert, um Informationen zu gewinnen, welche zur Bearbeitung bestimmter Problemstellungen verwendet werden können. Das heißt, dass datengetriebene Prozesse überwiegend Daten aus physischen Prozessen aufnehmen, diese analysieren und als Ergebnis wiederum Daten ausgeben, welche zum Beispiel zur Überwachung oder zur Verbesserung der physischen Prozesse genutzt werden können. Neben einer primären Verwendung zur Prozessdiagnose, scheinen datengetriebene Prozesse ebenfalls zu prädiktiven Analysen (16 von 26 Nennungen) fähig zu sein, um mit einer gewissen Wahrscheinlichkeit Aussagen über zukünftige Ereignisse zu treffen. Zu präskriptiven Analysen machen weniger als 50% der Autoren\*innen Aussagen. Diese werden vorrangig im Zusammenhang mit Closed-Loop Methoden erwähnt, welche zum einen zur Lösung von Problemen und zum anderen zur Steuerung von Prozessen eingesetzt werden kann (Zhang et al., 2019b). Für einige Autoren\*innen sind diese Analysen allerdings nicht unbedingt ein fester Bestandteil datengetriebener Prozesse, sondern eigenständige Module, welche durch die Ergebnisse datengetriebener Prozesse angetrieben werden (Tao et al., 2018b). Genauere Aussagen zum Begriff Data Driven, welche aus den Quelltexten eruiert wurden, werden im folgenden Abschnitt thematisiert.

### 3.1.3 Kernaussagen zu datengetriebenen Prozessen aus der Literaturanalyse

Im Zuge der Inhaltsanalysen wurden die Texte aus Tabelle 1 nicht nur hinblickend auf die in Kapitel 3.1.1 erarbeiteten formalen und informalen Analyse Kriterien untersucht. Während der Durchführung der Analysen wurden mehrere Aussagen der Autoren\*innen zum Begriff Data Driven in den Texten erfasst, um in Anschluss deren Kernaussagen in eigenen Worten zusammenzufassen. Im Folgenden werden die wichtigsten Kernaussagen der analysierten Literatur vorgestellt. Die Aussagen werden zunächst tabellarisch gelistet, damit diese anschließend näher betrachtet werden können. Primär die sechs höchsten gerankten Texte aus Tabelle 2 werden genauer diskutiert. Im Anschluss werden mögliche Gemeinsamkeiten und Unterschiede in den Kernaussagen identifiziert, um am Ende des Abschnitts in einem Zwischenfazit die Ergebnisse in eigenen Worten zusammenzufassen. Die Zusammenfassung der Ergebnisse soll am Ende des dritten Kapitels als Grundlage zur Entwicklung eines Definitionsvorschlags für den Begriff Data Driven verwendet werden. Im Folgenden wird die Liste mit den Kernaussagen der drei relevantesten Artikel aus Konferenzbänden und der drei relevantesten Artikel aus Fachzeitschriften (Tabelle 2) vorgestellt. Eine Auflistung und detaillierte Beschreibung der Kernaussagen sämtlicher Autoren\*innen würde an dieser Stelle zu weit führen und wird vor dem Hintergrund, dass sich viele der Aussagen gleichen, zur Zielerreichung nicht als nötig erachtet. Diese können dennoch unterstützend herangezogen werden. Um einen vollständigen Überblick über die Aussagen aller analysierten Quellen zu erhalten, können sämtliche Kernaussagen der Tabelle: Gesamtübersicht der Inhaltsanalyse entnommen werden, welche dem digitalen Anhang der Arbeit beigelegt wird.

**Tabelle 6: Kernaussagen der sechs relevantesten Quellen**

Quelle	Kernaussage
Wen et al., 2018	Seit den Entwicklungen im Bereich Smart Factory sind datengetriebene Fehlerdiagnosen wieder interessant geworden, da durch die Möglichkeiten schnellerer Datenerfassung und -verarbeitung die Nutzenpotentiale dieser Diagnosen gestiegen sind. Datengetriebene Methoden werden auch knowledge-based Methoden genannt und sie benötigen eine große Menge historischer Daten, um auch ohne Modelle oder Signalmuster Fehler zu ermitteln. Aufgrund dessen sind sie gut geeignet für komplexe Systeme, bei denen eine Modellerstellung zu kompliziert ist. Traditionelle datengetriebene Methoden zur Fehlerdiagnose können Rohsignale nicht direkt verarbeiten und benötigen eine Methode zur Datenvorverarbeitung, um Merkmale zu extrahieren (oft händisch durch Experten*innen). Neue datengetriebene Prozesse nutzen statistische und ML Methoden bei der datengetriebenen Fehlerdiagnose, um zu lernen Merkmale automatisch zu identifizieren.

Tao et al., 2018b	<p>Die datengetriebene Fertigung kann als notwendige Voraussetzung für Smart Manufacturing angesehen werden. In einer datengetriebenen Fertigung helfen Fertigungsdaten Entscheidungsträgern*innen, Änderungen in kürzester Zeit zu verstehen, genaue Urteile über diese zu fällen und schnelle Reaktionsmaßnahmen zur Behebung von Problemen zu entwickeln. Infolgedessen können Produktionspläne, Fertigungsaktivitäten und Ressourcen eng aufeinander abgestimmt werden, um eine intelligente Fertigung zu fördern. Die Charakteristiken einer intelligenten datengetriebener Fertigung bestehen aus einer kundenzentrierten Produktentwicklung, der Selbstorganisation, der Selbstregulierung, der Selbstausführung und dem Selbstlernen. Ein datengetriebenes Smart Manufacturing Framework besteht aus den vier Modulen: Fertigungsmodul, Datentreibermodul, Echtzeitüberwachungsmodul und Problemverarbeitungsmodul. Datengetriebene Prozesse werden durch das Datentreiber Modul ermöglicht. In diesem werden direkt aus den physischen Prozessen ermittelte Daten übertragen. Die übertragenen Daten treiben verschiedene Aufgaben voran (Data Driven) nämlich die Datensammlung, die Datenintegration, die Datenvorverarbeitung und die Datenanalyse und -visualisierung. Gewonnene Erkenntnisse werden anschließend zur Entscheidungsfindung genutzt, welche direkt zur Optimierung der physischen Prozesse eingesetzt werden. Darüber hinaus kann das Data Driver Modul weitere Module antreiben, welche zum Beispiel zur direkten Steuerung von Prozessen eingesetzt werden (Echtzeitüberwachungsmodul) oder Probleme in Prozessen identifizieren und Problemlösungen erarbeiten (Problemverarbeitungsmodul). Dieses Framework ermöglicht eine Vielzahl von Anwendungen vor allem in der Qualitätskontrolle, im Smart Maintenance, dem Design, der Planung und Optimierung, der Materialdistribution und -verfolgung und der Prozessüberwachung.</p>
Kamble und Gunasekaran, 2020	<p>Big Data Analytics vereint die Datensammlung, Analysetools, Computeralgorithmen und Techniken, um Wissen aus großen Datenmengen zu extrahieren. Big Data Analytics (BDA) und Big Data Driven unterscheiden sich dadurch, dass sich Big Data Driven auf einen bestimmten Prozess bezieht, welcher BDA nutzt. Datengetrieben meint im Kontext von Big Data Driven Systems die Verwendung von BDA zur Verwaltung aller SC-Prozesse.</p>
Mamakoukas et al., 2019	<p>Datengesteuerte Modellierungs- und Steuerungsansätze können in Echtzeit ausgeführt werden. Zusammen mit Werkzeugen für maschinelles Lernen ermöglichen diese Rückkopplungen, welche Daten verwenden können, um ihre Modelle anzupassen, um ein möglichst genaues Modell bei einer effizienten Rechenleistung zu erhalten. Viele datengetriebene Ansätze erfordern intensive online Berechnungen bei der Steuerung von Prozessen.</p>

Qiao, Q. et al. 2019	Das digitale Zwillingsmodell integriert physikalische Objekte, virtuelle Modelle, Datenzusammenführung, eingebettete Sensoren intelligente Algorithmen und Datenbanken. Zudem nutzt es ein Deep Learning Ansatz zur Vorhersage von Zuständen. Data Driven Technologien sind zusammen mit Durchführungsregeln und innovativen Ausbildungstechniken die Faktoren, welche die Entwicklung digitaler Zwillinge ermöglichen. Diese datengetriebenen Ansätze werden für das Monitoring und unter Zuhilfenahme von Data Analytics zur Fehlerdiagnose verwendet. Zusammen mit statistischen Verfahren oder modellbasierten Ansätzen, können diese auch zur Vorhersage von Fehlern genutzt werden.
Zhang et al., 2019a	Ein datengetriebenes Framework für den digitalen Zwilling einer Fertigungszelle, ermöglicht Probleme auf Grundlage erfasster Daten eigenständig zu erkennen, zuverlässige Lösungen zu erarbeiten und diese autonom durch eine Steuerung des Prozesses umzusetzen. Auf der Grundlage von Echtzeitdaten werden Fertigungsprozesse erfasst und simuliert. Einheitliche und interoperable Modelle vernetzen den physischen und digitalen Raum. Unter Verwendung von KIs und extrahierten Informationen z. B. aus Prozessen und internen Datenbanken, werden die Modelle angelehnt zuverlässige Entscheidungen zu treffen.

Unter den Beiträgen zum Begriff Data Driven waren vor allem die Aussagen von Ph.D. Fei Tao, welcher an der Beihang University (BUAA) School of Automation and Electrical Engineering tätig ist, sehr aufschlussreich. Dieser hat laut Scopus gerade in den letzten vier Jahren mehrere hochzitierte Artikel zu den Themen Industrie 4.0, Smart Manufacturing und digitaler Zwilling publiziert, in denen er auch Bezug zu datengetriebenen Prozessen nimmt. Von allen Autoren\*innen konkretisiert Tao in seinem Fachartikel „Data Driven Smart Manufacturing“ (Tao et al., 2018b) den Begriff Data Driven am genauesten, indem er detailliert ein mögliches Framework einer datengetriebenen intelligenten Fabrik beschreibt und im Text und in Abbildungen deutlich auf die Charakteristiken des Begriffes Data Driven eingeht. Ob sich seine Auffassung des Begriffes Data Driven mit den Auffassungen der anderen Autoren\*innen deckt, wird im Folgenden untersucht. Um eine möglichst allgemeine Begriffsvorstellung von Data Driven zu erhalten, werden Gemeinsamkeiten und Unterschiede in den Kernaussagen der Autoren\*innen extrahiert. Dabei wird primär auf die Kernaussagen der sechs relevantesten Quellen (Tabelle 6) eingegangen.

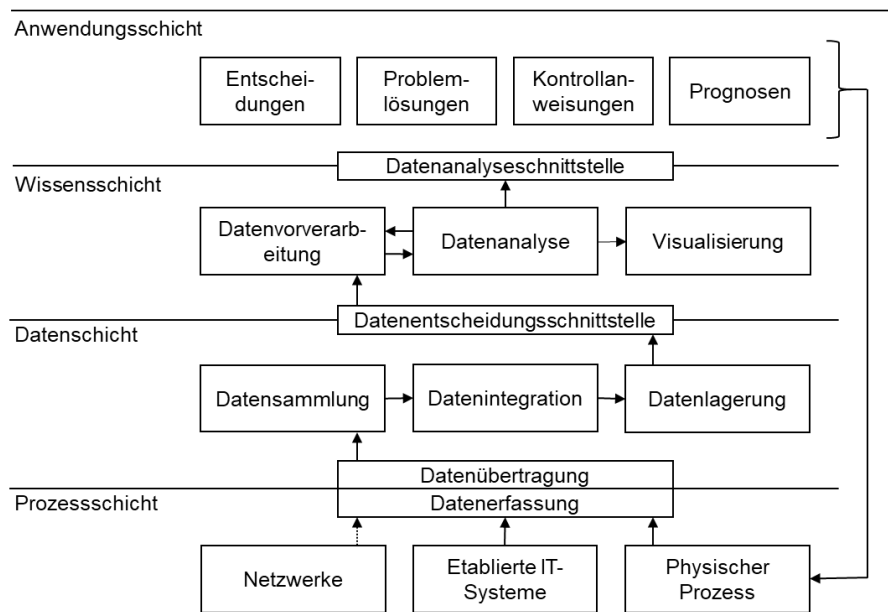
In ihrem Artikel schreiben Wen et al., 2018, dass datengetriebene Prozesse über die Zeit einen Wandel durchlebt haben. Während die traditionellen Prozesse noch auf Modelle und der Arbeit von Experten\*innen angewiesen waren, benötigen die aktuellen Prozesse keine manuelle Vorverarbeitung der zu analysierenden Daten mehr. Da in der Literaturanalyse aus-



schließlich gegenwärtige wissenschaftliche Quellen untersucht worden sind, bildet diese Arbeit nur die aktuelle Auffassung datengetriebener Prozesse ab. Aus diesem Grund wird im Folgenden zwischen einer **substanziellen Bedeutung des Begriffes Data Driven** und der Bedeutung im Zusammenhang mit der Beschreibung aktueller datengetriebener Prozesse im Umfeld von Produktion und Logistik unterschieden. Tao et al., 2018b unterscheiden in einem Framework einer datengetriebenen intelligenten Fabrik die beiden Bedeutungen, indem sie Data Driven als die direkt aus physischen Prozessen ermittelten und übertragenen Daten, welche verschiedene Aufgaben vorantreiben, beschreiben. Der gesamte datengetriebene Prozess hingegen umfasst nicht nur die Datenerfassung und Übertragung, sondern auch die Datenvorverarbeitung, die weiteren Analysen und alle weiteren nötigen Schritte bis hin zur Ausgabe der Ergebnisse. Vor allem meint Data Driven auch, bei Prozessen ein besonderes Augenmerk auf die Rolle der Daten zu legen (Cerquitelli et al., 2021). Dabei steht keine willkürliche Suche von Informationen im Vordergrund, sondern die Verwendung von Daten, um Wissen (Abschnitt 2.1) zu generieren, zur Erfüllung bestimmter Aufgaben (Kamble und Gunasekaran, 2020; Nguyen et al., 2021; Cerquitelli et al., 2021).

Doch für welche Aufgaben werden aktuelle datengetriebene Prozesse im Bereich von Produktion und Logistik eingesetzt und entwickelt, welche Vorteile bieten diese und welche Methoden und Technologien zeichnen diese aus? Zur Beantwortung dieser Fragen werden nachfolgend Aussagen zu aktuellen datengetriebenen Prozessen thematisiert. Einer der Hauptgründe zum Einsatz datengetriebener Prozesse besteht in einer gewünschten **Produktivitätssteigerung**. Datengetriebene Unternehmen schneiden hinsichtlich ihrer Produktivität besser ab. Zum einen können datengetriebene Prozesse direkt eine Steigerung der Produktivität bewirken (Miragliotta, G. et al. 2018), zum anderen werden datengetriebene Modelle (z. B. für Regelungsaufgaben) durch die Verwendung von Ein- und Ausgabedaten aufgebaut und haben dadurch einen geringeren Kostenaufwand im Gegensatz zur Erstellung physikalischer Modelle (Na und Peng, 2021). Die **Einsatzmöglichkeiten** datengetriebener Prozesse sind vielfältig. Wie in Abbildung 13 und Abbildung 14 zuvor ermittelt, werden datengetriebene Prozesse meist zur Diagnose operativer Prozesse eingesetzt, können aber auch genutzt werden, um Vorhersagen zu treffen. Auch Tao et al., 2018b gehen in ihrer Beschreibung einer datengetriebenen intelligenten Fertigung hauptsächlich auf operative Einsatzmöglichkeiten ein. Die Ergebnisse ihrer Analysen können beispielsweise zur Prozessüberwachung, zur Prozessoptimierung, zur Entscheidungsunterstützung, zur Problemanalyse und -lösung oder auch zur Steuerung physischer Prozesse genutzt werden (Tao et al., 2018b). Im Bereich Produktion wurden insbesondere Anwendungen in den Gebieten Fehlerprognose und -diagnose, Engpassüberwachung und Smart Factory genannt (Tao et al., 2018b; Subramaniyan et al., 2018; Na und Peng, 2021). Logistische Anwendungen betrafen unter anderem datengetriebene Supply Chains und einer Unterstützung des operativen SCM. Ein angeführter Vorteil, den aktuelle datengetriebene Systeme bieten, ist ihre **autonome Arbeitsweise**. Wie bereits beschrieben bauen diese komplett auf Daten auf und sind nicht mehr auf die Erfahrung von Ex-

perten\*innen zur Datenvorverarbeitung angewiesen. Probleme können eigenständig auf Grundlage erfasster Daten erkannt und zuverlässige Lösungen erarbeitet werden. Um bestimmte Muster in den Daten automatisch zu erkennen, beziehungsweise um aus Erfahrungswerten zu lernen, nutzen diese Technologien des maschinellen Lernens (Abschnitt 2.3) (Wen et al., 2018; Zhang et al., 2019a). Roh, P. et al. 2018 definieren den Begriff Data Driven überwiegend über die Möglichkeit die Datenerfassung und -analyse vollständig zu automatisieren und auch Tao et al., 2018b geben als Charakteristiken einer intelligenten datengetriebenen Fertigung neben der Kundenzentriertheit die Selbstorganisation, Selbstregelung, Selbstausführung und die Möglichkeit des Selbstlernens an. Dies bedeutet auch, dass vor der Durchführung datengetriebener Prozesse kein vorheriges Verständnis des zu behandelnden Problems benötigt wird (Kuo und Kusiak, 2019). Im Abschnitt 3.1.2 wurde bereits über den Einsatz diskutiert, physische Prozesse durch datengetriebene Prozesse mithilfe einer Rückkopplung zu automatisieren. Diese Möglichkeit scheint zu bestehen, doch wird dies nicht von allen Quellen als ein typisches Merkmal datengetriebener Prozesse angesehen. Die Charakteristiken einer autonomen Arbeitsweise implizieren, dass datengetriebene Prozesse gerade **für komplexe Systeme geeignet** sind, deren Modellierung eine große Herausforderung darstellt. Diese sind imstande auch ohne Signalmuster oder Modellen zu arbeiten und erlauben es darüber hinaus, stark verflochtene Muster in Datenmengen mit großen Volumen und komplexen und heterogenen Datenstrukturen und -formaten zu identifizieren (Kuo und Kusiak, 2019; Wen et al., 2018). Dadurch können datengetriebene Prozesse auch in komplexen und stark dynamischen Systemen dabei helfen verlässliche Daten zu gewinnen und diese eventuell auch zu steuern (Olusola. T. Faboya et al. 2018). Zur Bewältigung der zuvor beschriebenen Aufgaben nutzen datengetriebene Prozesse verschiedene **Methoden und Technologien**, welche sich je nach Ansatz voneinander unterscheiden können (Kuo und Kusiak, 2019). Zur Veranschaulichung der Arbeitsweise und des Aufbaus aktueller datengetriebener Prozesse und der in vielen Prozessen eingesetzten Technologien soll die folgende Abbildung dienen.



**Abbildung 15: Beispielmodell aktueller datengetriebener Prozesse (i. A. a. Tao et al., 2018b; Xu et al., 2020; Cerquitelli et al., 2021)**

Abbildung 15 zeigt ein Beispielmodell aktueller datengetriebener Prozesse. Wie bereits geschrieben existieren viele unterschiedliche Prozesse, deren Vorgehensweisen sich jedoch oft ähneln. Aus diesem Grund ist das Modell hinreichend gut geeignet, um einen kurzen Überblick über die aus den Quellen eruierten Aufgaben und Technologien datengetriebener Prozesse zu vermitteln. In Anlehnung an Xu et al., 2020 können datengetriebene Prozesse im Bereich der Produktion in die vier Schichten Prozessschicht, Datenschicht, Wissensschicht und Anwendungsschicht unterteilt werden. Laut Tao et al., 2018b sind es vor allem die Prozess- und die Datenschicht, welche den Begriff Data Driven ausmachen. Die Prozessschicht umfasst vor allem die zu analysierenden physischen Prozesse und etablierte IT-Systeme (Abschnitt 2.2). Manche Autoren\*innen, darunter zum Beispiel Kuo und Kusiak, 2019 und Jeong et al., 2020, nennen auch öffentliche Netzwerke oder Datenbanken. Die Datenerfassung erfasst Daten aus der Prozessschicht. Zur Erfassung von aktuellen Daten aus den physischen Prozessen in nahezu Echtzeit nutzt diese beispielsweise Sensoren und RFID Systeme (Abschnitt 3.1.3). Aus den etablierten IT-Systemen werden zum Beispiel Planungsdaten und historische Prozessdaten entnommen. Daten aus externen Systemen werden vornehmlich dazu genutzt KIs im eigenen System anzulernen (Ivanov und Dolgui, 2021). Laut Xu et al., 2020 werden die gesamten Daten an die Datenschicht übertragen, welche die Daten zunächst sammelt, danach zusammenführt, bereinigt und lagert. Die bereinigten Daten werden über eine Schnittstelle an die Wissensschicht übermittelt, welche sich mit der Extraktion von Wissen aus den Daten beschäftigt. Zunächst werden die Daten vorverarbeitet. Zur Filterung der Daten kommen zum Beispiel ML Technologien und statistische Data Analytics Methoden zum Einsatz, welche sowohl Daten aus vor- und den nachgelagerten Prozessen verwenden (Ivanov und Dolgui, 2021). Zur Datenanalyse nutzen datengetriebene Prozesse beispielsweise Simula-

tions- und Optimierungsmodelle, um Vorhersagen zu treffen, Datamining zur Wissenssuche oder ebenfalls ML Algorithmen und Big Data Analysen (Kamble und Gunasekaran, 2020; Qiao, Q. et al. 2019; Zhang et al., 2019a). Die Ergebnisse der Analyse werden primär zur Diagnose und Entscheidungsfindung eingesetzt (Abschnitt 3.1.2), ermöglichen aber auch Prognosen, Problemlösungen, Echtzeitmonitoring oder Kontrollanweisungen (Tao et al., 2018b). Die Ergebnisse des datengetriebenen Prozesses stellen wiederum Daten dar, welche zum physischen Prozess rückgekoppelt werden können, um einen geschlossenen Regelkreis zu ermöglichen (Tao et al., 2018b; Xu et al., 2020). In den Bereichen der Forschung und Entwicklung stehen datengetriebene Methoden aktuell primär im Fokus, da sie (als datengetriebenes Framework) die Entwicklung digitaler Zwillingmodelle ermöglichen. Diese stellen digitale Abbilder physischer Entitäten dar, welche den digitalen und den physischen Raum durch physische, virtuelle und verbundene Daten miteinander verbinden (Zhang et al., 2019a; Tao et al., 2018a).

Zusammenfassend lässt sich feststellen, dass sich die Aussagen der Autoren\*innen in mehreren Punkten überschneiden, aber gerade bei einer detaillierteren Beschreibung datengetriebener Prozesse weiter auseinander gehen. Daher lassen sich zwar häufig eingesetzte Methoden und Technologien (z. B. Data Analytics, ML) identifizieren, jedoch werden diese nicht als fester Bestandteil datengetriebener Prozesse beschrieben und können von Prozess zu Prozess variieren. Zudem scheinen datengetriebene Prozesse sich über eine Zielsetzung zu definieren, jedoch werden keine generellen Anwendungen benannt, daher scheinen diese auch variabel zu sein. Aus diesem Grund ist es schwer, allgemeine Aussagen über den Aufbau oder Anwendungen datengetriebener Prozesse zu treffen. Aus den Kernaussagen der Autoren\*innen lässt sich folgendes Zwischenfazit formulieren:

Der Begriff Data Driven kennzeichnet eine primär auf Daten fokussierte Betrachtung physischer Prozesse, zur Erfüllung konkretisierter vorwiegend operativer Aufgabenstellungen. Zu diesem Zweck werden Daten direkt aus physikalischen Prozessen erfasst und übertragen, um Analysen anzutreiben, welche gezielt auf die jeweilige Aufgabe bezogenes Wissen generieren. Informationstechnische Prozesse, welche Data Driven sind, bezeichnen sich als Data Driven Processes. Aktuelle Data Driven Processes können durch den Einsatz von Sensoren, AutoID Systemen und der Anbindung zu etablierten informationstechnischen Systemen große Mengen aktueller und historischer Daten erfassen und verarbeiten. Sie ermöglichen eine autonome Untersuchung dynamischer komplexer Systeme beispielsweise durch den Einsatz von Data Analytics und Machine Learning. Darüber hinaus können die, von den meist diagnostischen oder prädiktiven Analysen gelieferten, Erkenntnisse direkt zum physischen Prozess rückgekoppelt werden, um zum Beispiel diagnostizierte Fehler direkt zu beheben oder Prozessregelungen zu realisieren.

### 3.2 Literaturanalysegestützte Ableitung von Begriffsmerkmalen datengetriebener Prozesse

Unter Verwendung des zuvor aufgestellten Zwischenfazit (Abschnitt 3.1.3) und der Ergebnisse der Inhaltsanalysen (Abschnitt 3.1.2) wird das Ziel verfolgt Begriffsmerkmale (Abschnitt 2.6) für datengetriebene Prozesse abzuleiten. Dies dient die bisher gewonnenen Erkenntnisse zum Begriff Data Driven zusammenzufassen, um darzustellen, welche Merkmale einen datengetriebenen Prozess hinreichend genau beschreiben. Darüber hinaus wird mit Hilfe der Begriffsmerkmale ein Instrument entwickelt, welches Aufschluss darüber geben kann ob oder inwiefern ein betrachteter Prozess als datengetrieben bezeichnet werden kann.

Die Merkmale datengetriebener Prozesse werden zum einen aus den informalen Analyse-kriterien (Tabelle 3, Tabelle 4, Tabelle 5) und deren in Abschnitt 3.1.2 ermittelten Ausprägungen und zum anderen aus den Gemeinsamkeiten der Kernaussagen (Abschnitt 3.1.3) abgeleitet. Mithilfe der Merkmale und ihren Ausprägungen wird eine Art morphologischer Kasten entwickelt, welcher dazu dienen kann zu prüfen, inwiefern bestimmte Prozesse Data Driven sind. Die ersten Merkmale, die identifiziert werden können, sind die **Mandatorischen Merkmale**, welche notwendige Merkmale darstellen um einen Prozess datengetrieben nennen zu können. Diese können unmittelbar aus dem Zwischenfazit des Abschnitts 3.1.3 entnommen werden nämlich, dass datengetriebene Prozesse primär die Daten eines Prozesses betrachten, eine oder mehrere konkrete Aufgabenstellungen verfolgen, mit Hilfe ihrer Analysen Wissen generieren und Daten direkt aus den physischen Prozessen erfassen. Daraus folgt, dass Prozesse, welche nicht alle dieser vier Punkte erfüllen nicht als datengetrieben bezeichnet werden können.

Die weiteren Merkmale können nur identifizieren, inwieweit ein Prozess den Auffassungen aktueller datengetriebener Prozesse entspricht. Dazu werden starke und schwache Merkmalsausprägungen identifiziert und solche, welche laut Analyse gegen eine Bezeichnung als Data Driven sprechen. Als erstes werden mit Hilfe der informalen Kriterien aus Abschnitt 3.1.2 **Datenmerkmale** übernommen. Die Auswertung der Literaturanalyse wird als Grundlage für die Bewertung der Merkmalsausprägungen genutzt. Bei den Ausprägungen der Verarbeitung großer Datenvolumen und der Verarbeitung unterschiedlicher Datenstrukturen wurde zudem aus den Kernaussagen ermittelt, dass sich datengetriebene Prozesse nicht über diese definieren. Der Großteil der Quellen erwähnt zudem eine Datenvorverarbeitung, welche rein automatisch erfolgt. Eine manuell durch die Arbeit von Experten\*innen durchgeführte Vorverarbeitung von Daten ist laut Quellen kein Merkmal aktueller datengetriebener Prozesse, kann jedoch in älteren datengetriebenen Prozessen vorkommen. Bei der Datenerfassung ist die Datenextraktion aus den physischen Prozessen am bedeutendsten, gefolgt von der Datenübernahme aus etablierten IT-Systemen. Die Erfassung analoger Daten wird in der Literatur kaum erwähnt und manuell eingegebene Daten werden laut Quellen gar nicht genutzt. Auch eine Echtzeitdatenerfassung kann bedeutend für aktuelle datengetriebene Prozesse sein.



Optionale Merkmale	Datenmerkmale	Datenerfassung	Extrahiert Daten aus physischen Prozessen	✓	Übernimmt Daten etablierter IT-Systeme	●	Erfasst analoge Daten	Manuelle Eingabe von Daten durch Menschen	✗
		Echtzeiterfassung	Echtzeitdatenerfassung und Analysen möglich und bedeutend	●	Echtzeitdatenerfassung und Analysen möglich, aber unbedeutend		Echtzeitdatenerfassung und Analysen nicht möglich, aber bedeutend	Echtzeitdatenerfassung und Analysen nicht möglich und unbedeutend	
		Datenstruktur	Verarbeitung unterschiedlicher Datenstrukturen ist möglich, aber nicht primär bedeutend	●	Verarbeitung unterschiedlicher Datenstrukturen ist teilweise möglich		Verarbeitung unterschiedlicher Datenstrukturen ist nicht möglich	Verarbeitung unterschiedlicher Datenstrukturen steht im Vordergrund	✗
		Datenvolumen	Verarbeitung/ Erfassung großer Datenmengen ist möglich, aber nicht primär bedeutend	●	Verarbeitung/ Erfassung großer Datenmengen ist nicht möglich, aber bedeutend		Verarbeitung/ Erfassung großer Datenmengen ist nicht möglich und unbedeutend	Verarbeitung/ Erfassung großer Datenmengen steht im Vordergrund	✗
		Datenvorverarbeitung	Automatische Datenvorverarbeitung	✓	Teilweise automatische Datenvorverarbeitung		Keine Datenvorverarbeitung	Manuelle Datenvorverarbeitung durch Experten*innen	✗
	Aufgabenmerkmale	Aufgabenorientierung	Der Prozess wird für operative Aufgaben eingesetzt	✓	Der Prozess wird für taktische Aufgaben eingesetzt	●	Der Prozess wird für strategische Aufgaben eingesetzt	Der Prozess hat keine Aufgabenorientierung	
		Produktivität	Bewirkt eine deutliche Steigerung der Produktivität	●	Bewirkt eine geringe Steigerung der Produktivität	●	Bewirkt Keine Steigerung der Produktivität	Bewirkt eine Abnahme der Produktivität	
		Systemkomplexität	Ermöglicht Analysen hochkomplexer dynamischer Systeme	✓	Ermöglicht Analysen teils komplexer Systeme		Schwierigkeiten bei Analysen komplexer Systeme	Keine Analysen komplexer Systeme möglich	✗

Anwendungsmerkmale	Analyse-system	Das Analyse-system arbeitet diagnostisch	✓	Das Analyse-system arbeitet prädiktiv	●	Das Analyse-system arbeitet präskriptiv	●	Das Analyse-system arbeitet deskriptiv	
	Analyse-metho-den	Nutzen sta-tistischer Methoden (z. B. Data Analytics) und künstli-cher Intelli-genzen	✓	Nutzen statistischer Methoden oder künstli-cher Intelli-genzen	●	Nutzen von Modellen des physischen Systems		Nutzen keiner der genannten Methoden	
	Autonome Arbeitsweise	Prozess arbeitet voll-ständig autonom	✓	Prozess ar-beitet teil-weise autonom		Prozess ar-beitet so gut wie gar nicht autonom		Prozess arbeitet nicht autonom	✗
	Rückkopp-lung	Rückkopp-lung zum physischen System ist möglich und notwendig	●	Rückkopp-lung zum physischen System ist möglich, aber nicht notwendig	●	Rückkopp-lung zum physischen System ist nicht mög-lich, aber notwendig		Rückkopp-lung zum physischen System ist weder mög-lich noch notwendig	

✓ = trifft voll zu   ● = trifft teilweise zu   ✗ = trifft gar nicht zu

Tabelle 7 stellt einen **morphologischen Kasten aktueller datengetriebener Prozesse** dar. Dieser setzt sich zusammen aus den zuvor ermittelten Merkmalen und unterschiedlichen Merkmalsausprägungen. Die bereits eingetragenen Ausprägungen betreffen derer aktueller datengetriebener Prozesse, wobei zwischen voll zutreffenden- und teilweise zutreffenden Merkmalen unterschieden wird. Zusätzlich wurde erfasst welche Merkmale laut Analysen nicht auf diese Prozesse zutreffen. Die ersten (mandatorischen) Merkmale können gesondert betrachtet werden, da sie die elementarsten Anforderungen an den Begriff Data Driven kennzeichnen, welche durch die Literaturanalyse ermittelt werden konnten. Daraus folgt, dass alle vier Merkmale eines Prozesses mit den aufgeführten Ausprägungen übereinstimmen müssen, damit dieser als Data Driven bezeichnet werden kann. Für den Rest des Kastens kann angenommen werden, dass die Ausprägung, wie sehr ein Prozess der aktuellen Auffassung datengetriebener Prozesse gleicht, damit übereinstimmt, wie weit links des Kastens sich seine Ausprägungen befinden. Prozesse, welche Ausprägungen haben, die laut Tabelle 7 gar nicht auf aktuelle datengetriebene Prozesse zutreffen, können auch nicht als solche bezeichnet werden. Die Ausnahme bildet eine Merkmalsausprägung, welche die manuelle Vorverarbeitung von Daten betrifft. Diese Ausprägung schließt zwar aktuelle datengetriebene Prozesse aus, kann jedoch eine Ausprägung traditioneller datengetriebener Prozesse sein. Im weiteren Verlauf der Arbeit kann der Kasten genutzt werden, um informationstechnische Prozesse vom Begriff Data Driven abzugrenzen oder zu prüfen, inwieweit diese die Anforderungen aktueller datengetriebener Prozesse erfüllen.



### 3.3 Sichtweisen verschiedener Fachdisziplinen auf den Begriff Data Driven

Wie bereits in der Anwendungsphase in Abschnitt 3.1.2 Durchführung und Ergebnisdarstellung der Literaturanalyse zu datengetriebenen Prozessen erwähnt, wurden in der Literaturanalyse Artikel untersucht, welche von Wissenschaftlern\*innen unterschiedlicher Fachdisziplinen verfasst wurden. Die Fachdisziplinen der Autoren\*innen lassen sich grob in die Bereiche Maschinenbau, Logistik, Regelungstechnik und Informatik unterteilen.

Da in den Abschnitten 3.1.2 und 3.1.3 ungefiltert die Aussagen aller Autoren\*innen analysiert worden sind, repräsentieren die Merkmale aus Tabelle 7 eine gemeinsame Begriffsvorstellung der zuvor genannten Fachdisziplinen. Aufgrund der Literatursuche mit einem Fokus auf Ingenieurwissenschaften im Bereich von Produktion und Logistik lässt sich der Großteil der Verfasser dem Bereich des Maschinenbaus zuordnen. Eine weitere Auswirkung der Literatursuche und Auswahl ist, dass in der Literatur eine überwiegend technische Sichtweise des Begriffes untersucht wurde und infolgedessen eine wirtschaftswissenschaftliche Sichtweise vernachlässigt wurde. Daher ist die Aussage, dass der Einsatz datengetriebener Prozesse zu einer Steigerung der Produktivität führt, eine der wenigen Aussagen zu der wirtschaftlichen Wirkung dieser Prozesse. Ob beispielsweise datengetriebene Geschäftsmodelle über eine Optimierung der eigenen Prozesse hinaus, die Erschließung neuer Geschäftsmodelle ermöglichen, müsste in einer weiteren Literaturanalyse untersucht werden.

Unter den Aussagen der betrachteten Fachdisziplinen gab es in allem keine großen Unterschiede. Aus Sichtweise der Regelungstechnik scheint die Möglichkeit datengetrieben große Datenmengen zu erfassen und zu verarbeiten weniger Beachtung zu finden, da Autoren\*innen aus dieser Disziplin keine Aussagen zu dieser Thematik machen. Weitere auffällige Unterschiede kommen aus den Bereichen Logistik und Informatik. Ausschließlich Autoren\*innen dieser beiden Fachdisziplinen machten Aussagen zu einer möglichen strategischen Nutzung datengetriebener Prozesse. Dies könnte bedeuten, dass datengetriebene Prozesse eventuell auch im strategischen Bereich einen Nutzen haben, jedoch gerade im Maschinenbau und der Regelungstechnik eine vorwiegend operative Verwendung überwiegt. Dazu passend ist, dass alle Quellen aus den Bereichen Logistik und Informatik Angaben zu datengetriebenen prädiktiven Analysen machen, welche Planungsaufgaben unterstützen können.

### 3.4 Abgrenzung von Data Driven zu anderen Begriffen

Neben der Ableitung von Begriffsmerkmalen datengestützter Prozesse und einer Untersuchung unterschiedlicher Sichtweisen, war die Abgrenzung von Data Driven zu etablierten IT-Prozessen in der Produktion und Logistik und zu den Begriffen Industrie 4.0 und Big Data eine weitere Intention für die Durchführung der Literaturanalyse. Zur Erfüllung dieses Ziels wurden im Kapitel 2 die Grundlagen der abzugrenzenden Begriffe erläutert und bei der Durchführung der Literaturanalyse wurden anhand des informalen Analysekriteriums *Eingrenzung* (Tabelle 5) die Quellen gezielt auf Aussagen untersucht, welche Data Driven von den übrigen Begriffen unterscheiden. Zudem können die Ergebnisse aus den Abschnitten 3.1.2 und 3.1.3 zur Unterstützung der Abgrenzungen hinzugezogen werden. Eventuelle neue Erkenntnisse, die aus den Abgrenzungen gewonnen werden, könnten die Merkmale des morphologischen Kastens aktueller datengetriebener Prozesse (Tabelle 7) ergänzen und darüber hinaus zur Entwicklung eines eigenen Definitionsvorschlags hinzugezogen werden.

#### 3.4.1 Abgrenzung zu etablierten informationstechnischen Prozessen

Zu den Unterschieden zwischen datengetriebenen Prozessen und etablierten informationstechnischen Prozessen konnten in der Literaturanalyse Aussagen aus 15 der insgesamt 28 Quellen identifiziert werden. Darunter befinden sich auch Aussagen, welche aus vier der sechs relevantesten Quellen (Tabelle 2) stammen. Da sich einige der Aussagen gleichen und um die wichtigsten Aussagen zu erfassen, wird im Folgenden, der Übersicht halber, eine Tabelle mit den Aussagen jener vier Quellen vorgestellt. Für die weitere Betrachtung können jedoch die Aussagen der restlichen Quellen hinzugezogen werden. Eine gesamte Übersicht aller Aussagen befindet sich im digitalen Anhang in der Tabelle: Gesamtübersicht der Inhaltsanalyse.

**Tabelle 8: Abgrenzung zu etablierten IT-Prozessen**

Quellen	Abgrenzung zu etablierten IT-Prozessen
Tao et al., 2018b	Etablierte IT-Systeme gehören ins Informationszeitalter. Im Informationszeitalter wurden Daten in Computersystemen gespeichert und von Informationssystemen verwaltet. Die Analyseergebnisse mussten von menschlichen Bedienern*innen interpretiert werden. Zudem gab es keine effektiven Möglichkeiten unstrukturierte, verstreute, sich wiederholende und isolierte Daten zu analysieren.

Kamble und Gunasekaran, 2020	Die in traditionellen SC verwendeten Prozesse basieren auf historischen Informationen sind isoliert, statisch und liefern weniger effizient Informationen an die Entscheidungsträger. Das traditionelle Leistungsmessungssystem (PMS) ist darauf bedacht herauszufinden, was in der Vergangenheit passiert ist und sucht die Gründe dafür. Datengetriebene Systeme antizipieren dagegen zukünftige Leistungen und reagieren nicht erst auf Probleme, nachdem diese entstanden sind.
Mamakoukas et al., 2019	Datengetriebene Prozesse können unterstützen nicht lineare Systeme zu regeln.
Qiao, Q. et al. 2019	In der Vergangenheit basierten Fehlerdiagnosen und -prognosen auf Signalanalysen und den Erfahrungen von Experten*innen ohne die Involvierung von Big Data.

Die Ziele etablierter IT-Systeme (Abschnitt 2.2) und datengetriebener Prozesse (Abschnitt 3.1.3) sind grundlegend die Gleichen, nämlich eine Steigerung der Produktivität durch die Optimierung des Produktions- oder Logistiksystems. Besonders im operativen Bereich, in welchem viele datengetriebene Prozesse arbeiten, bieten diese einige Vorteile gegenüber den etablierten Prozessen. Viele dieser Vorteile haben ihren Ursprung in den unterschiedlichen Ansätzen der Analysensysteme. Etablierte IT-Systeme basieren größtenteils auf Modellen physikalischer Systeme (Abschnitt 2.2) wohingegen aktuelle datengetriebene Analysen meist auf statistische Methoden und ML basieren (Abschnitt 3.1.3) (Cerquitelli et al., 2021). Daraus folgt, dass auch die Betrachtung etablierter IT-Prozesse primär auf die physischen Prozesse und nicht auf Daten ausgerichtet ist. Bezogen auf die Datennutzung beschäftigen sich etablierte Systeme, auf operativer Ebene, überwiegend mit der Speicherung und der Verwaltung von Daten (Tao et al., 2018b). Ihre Analysensysteme arbeiten überwiegend deskriptiv-diagnostisch und nutzen historische Informationen (Kamble und Gunasekaran, 2020), um herauszufinden was und warum etwas in der Vergangenheit passiert ist. Datengetriebene Prozesse hingegen beziehen sich auf aktuelle Daten und Geschehnisse und antizipieren gegebenenfalls auch zukünftige Ergebnisse, indem sie dynamische Simulationsmodelle aufbauen, welche im Gegensatz zu herkömmlichen Simulationen den Einsatz in der Betriebsphase ermöglichen (Kamble und Gunasekaran, 2020; Jeong et al., 2020). Aufgrund der notwendigen Modellerstellung sind etablierte IT-Prozesse auch schlechter für den Einsatz in komplexen dynamischen Systemen geeignet (Na und Peng, 2021). Neben dem höheren Zeit- und Kostenaufwand sind deren Modelle meist statisch (Kamble und Gunasekaran, 2020) und unterstützen beispielsweise keine Regelung nicht linearer Systeme (Mamakoukas et al., 2019), welche intensive Onlineberechnungen voraussetzt. Erhöhte Aufwände sind nicht nur der Modellerstellung geschuldet, sondern auch weiterer Prozessschritte, welche die manuelle Arbeit menschlicher Bediener\*innen erfordern und nicht automatisiert ablaufen. So erfordern beispielsweise die Analyseergebnisse

etablierter IT-Prozesse zur Interpretation die Erfahrungen von Experten\*innen (Tao et al., 2018b; Qiao, Q. et al. 2019). Datengetriebene Prozesse benötigen zwar keine manuelle Extraktion von Merkmalen oder Interpretation von Experten\*innen, benötigen aber zum Anlernen ihrer Algorithmen auch Zeit und Trainingsdaten. Auch die Datenerfassung etablierter IT-Prozesse ist oft noch manuell (Roh, P. et al. 2018), kann jedoch durch den Einsatz von Sensoren auch automatisiert werden (Abschnitt 2.2). Dadurch können große Datenmengen erfasst, jedoch nicht analysiert werden (Kuo und Kusiak, 2019). Datengetriebene Analysen können durch die Involvierung von Big Data auch sehr große Datenmengen verarbeiten (Qiao, Q. et al. 2019). Zudem bieten etablierte IT-Prozesse im Gegensatz zu datengetriebenen Ansätzen keine effektiven Möglichkeiten unstrukturierte, verstreute, sich wiederholende und isolierte Daten zu analysieren (Tao et al., 2018b).

### 3.4.2 Abgrenzung zu Industrie 4.0

Wie bereits erwähnt nutzen datengetriebene Prozesse Technologien von Big Data, AutoID oder dem maschinellen Lernen. Industrie 4.0 ist ebenfalls ein Begriff, welcher diese aktuellen Technologien unter sich vereint (Abschnitt 2.3). Damit die beiden Begriffe voneinander abgegrenzt werden können, aber auch um ihre Beziehung zueinander zu untersuchen, wurden in der Literaturanalyse ebenfalls Aussagen zum Begriff Industrie 4.0 erfasst. Insgesamt wurden nur wenige Aussagen zu den Zusammenhängen beider Begriffe getätigt. Unter Berücksichtigung der Artikelrelevanz konnten folgende zwei Aussagen eruiert werden.

**Tabelle 9: Abgrenzung zu Industrie 4.0**

Quelle	Abgrenzung zu Industrie 4.0
Cerquitelli et al., 2021	Datengetriebene Modelle und Prozesse nutzen Technologien und Methoden der Industrie 4.0 und agieren in ihrer Umgebung.
Ma et al., 2020	Datengetriebene Prozesse bilden eine notwendige Voraussetzung für eine intelligente Fertigung und somit auch für die Industrie 4.0.

Aufgrund der Aussagen aus Tabelle 9 kann die Annahme getroffen werden, dass datengetriebene Prozesse in der Umgebung der Industrie 4.0 eingesetzt werden, oder sogar die Realisierung mancher ihrer Themen erst ermöglichen und somit als ein Teil dieser angesehen werden können. Ein großer Unterschied beider Begriffe besteht darin, dass die Industrie 4.0 ein Sammelbegriff an Themen und Technologien darstellt (Abschnitt 2.3), während unter Data Driven zwar auch unterschiedliche Technologien vereint werden, welche sich jedoch auf den Einsatz für spezifische Einzelanwendungen beziehen. Datengetriebene Prozesse greifen nicht

alle, aber mehrere der aktuellen Themenfelder der Industrie 4.0 auf, beispielsweise autonomes Handeln oder die Datenerfassung und -verarbeitung. Darüber hinaus stimmt der Begriff Data Driven teilweise mit der Beschreibung des Begriffes Industrie 4.0 des Bundesministeriums für Wirtschaft und Energie überein (Abschnitt 2.3). Bei dieser geht es vorzugsweise um die Realisierung einer intelligenten Fertigung, für diese laut Ma et al., 2020 datengetriebene Prozesse eine notwendige Voraussetzung bilden. Durch die Möglichkeit physische Prozesse nahezu in Echtzeit zu erfassen und zu analysieren, bieten datengetriebene Prozesse eine Lösung den physischen und den digitalen Raum zu vernetzen und damit datengetriebene digitale Zwillingmodelle zu entwickeln. Diese Modelle könnten bei der Realisierung cyberphysischer Systeme helfen, welche wiederum ein Thema im Zusammenhang mit intelligenten Fabriken sind. Neben den gemeinsamen Themenfelder, welche die Begriffe Data Driven und Industrie 4.0 verbinden, existieren dennoch etliche weitere Themen (Abbildung 5), zum Beispiel aus dem Bereich der Unternehmenstransformation, die durch datengetriebene Prozesse unberührt bleiben. Alles in allem bieten datengetriebene Prozesse im Umfeld von Produktion und Logistik Vorteile für die Umsetzung einiger Projekte der Industrie 4.0. Diese stellen keine eigenen neuen Technologien zur Verfügung, sondern stellen viel mehr eine Vorgehensweise oder einen Rahmen dar, welcher einige vorhandene Technologien der Industrie 4.0 (z. B. AutoID, ML, Big Data) zur Bearbeitung spezifischer Aufgabenstellungen nutzt.

### 3.4.3 Abgrenzung zu Big Data

Wie bereits im Abschnitt 2.5 beschrieben, existieren zwischen den Begriffen Big Data und Data Driven viele Parallelen. Aufgrund der hohen Ähnlichkeit ist es schwer beide Begriffe klar voneinander zu differenzieren. Dies ist wahrscheinlich auch einer der Auslöser dafür, dass in der Literatur die Begriffe nicht klar voneinander abgegrenzt werden. Vor allem aktuelle datengetriebene Prozesse bedienen sich meist Technologien und Methoden von Big Data (z. B. Big Data Analytics), weshalb in diesem Zusammenhang oft der Begriff *Big Data Driven* verwendet wird (Kamble und Gunasekaran, 2020). Infolge der beschriebenen Problematik konnten aus der Literaturanalyse lediglich zwei Aussagen zur Abgrenzung der Begriffe Big Data und Data Driven identifiziert werden, welche aber von zweien der sechs relevantesten Quellen (Tabelle 2) stammen.

**Tabelle 10: Abgrenzung zu Big Data**

Quelle	Abgrenzung zu Big Data
Tao et al., 2018b	Big Data versetzt Unternehmen in die Lage datengetriebene Strategien anzuwenden, um die Wettbewerbsfähigkeit zu steigern.

Kamble und Gunasekaran, 2020	Unterschied über den Prozessbezug. Big Data Analytics und Big Data Driven unterscheiden sich dadurch, dass sich Big Data Driven auf einen bestimmten Prozess (hier eine Lieferkette) bezieht, welcher Big Data Analytics nutzt.
------------------------------	---

Sowohl Big Data als auch Data Driven lassen sich dadurch beschreiben, dass sie das Ziel verfolgen einen bestimmten Mehrwert zu generieren (Abschnitt 2.4, Abschnitt 3.1.3). Zur Steigerung der Produktivität bearbeiten datengetriebene Prozesse konkrete Aufgabenstellungen (Abschnitt 3.1.3). In diesem Punkt scheint sich Big Data zu unterscheiden, da dessen Nutzen im Vorhinein meist nicht klar erkennbar ist (Abschnitt 2.4). Auch Kamble und Gunasekaran unterscheiden reine Big Data Analysen und Big Data Driven über den Prozessbezug (Tabelle 10). Dies bedeutet auch, dass Big Data als ein Teil aktueller datengetriebener Prozesse, welche Big Data Analysen nutzen, angesehen werden kann und Big Data sogar eine Voraussetzung für die Anwendung aktueller datengetriebener Strategien im Bereich von Produktion und Logistik ist (Holland, 2020; Tao et al., 2018b). Bezogen auf die Verwendung werden Big Data Analysen auch für die Erschließung neuer Geschäftsmodelle genutzt (Abschnitt 2.4). Die in der Literaturanalyse betrachteten datengetriebenen Prozesse wurden primär zur Optimierung vorhandener Prozesse auf operativer Ebene genutzt, jedoch wurden in der Analyse keine datengetriebene Geschäftsprozesse betrachtet, welche eventuell ebenfalls zur Erschließung neuer Geschäftsmodelle geeignet sein könnten. Da Big Data auch Entscheidungen auf strategischer Ebene unterstützt, hat die Möglichkeit vielfältige Datenquellen und Datenstrukturen zu nutzen wahrscheinlich einen höheren Stellenwert als bei datengetriebenen Prozessen, welche primär die Daten vorhandener Prozesse erfassen und weniger nach neuen Quellen suchen. Die Echtzeiterfassung von Daten ist für beide Begriffe für die Realisierung von Anwendungsfällen bedeutend, jedoch ist eine direkte Erfassung von Daten aus physischen Prozessen nicht für alle Big Data Anwendungen von Belangen. Vor allem die Möglichkeit große Datenmengen zu erfassen und zu verarbeiten ist eine der zentralen Aspekte von Big Data. Aktuelle datengetriebene Prozesse bieten zwar ebenfalls eine Möglichkeit große Mengen an Daten zu erfassen und unter Zuhilfenahme von Big Data Technologien zu analysieren, jedoch wird diese Möglichkeit in der Literatur nicht als einer ihrer zentralen Aspekte angesehen. Das Ziel datengetriebener Prozesse ist damit nicht große Mengen an Daten zu verarbeiten, sondern so viele Mengen an Daten zu erfassen und zu analysieren, wie es für die Erfüllung ihrer Ziele notwendig ist.

### 3.5 Diskussion und Definitionsvorschlag für den Begriff Data Driven

Das Hauptziel dieses Kapitels war es den Begriff Data Driven genauer zu spezifizieren, indem eine möglichst allgemeine Vorstellung des Begriffes erfasst wurde und daraus Merkmale datengetriebener Prozesse erarbeitet wurden. Im folgenden abschließenden Abschnitt werden die zuvor gewonnenen Erkenntnisse diskutiert, um auf deren Grundlage einen Definitionsvorschlag für den Begriff Data Driven im Umfeld von Produktion und Logistik zu verfassen.

Zur Erreichung dieses Ziels wurde zuerst eine Literaturanalyse durchgeführt, welche quantitative und qualitative Aussagen zu datengetriebenen Prozessen erfasste. Diese lieferte vor allem Erkenntnisse zum Einsatz von Daten und zur Anwendung datengetriebener Prozesse. Im Laufe der Literaturanalyse wurden zudem Kernaussagen zum Begriff Data Driven zusammengefasst, welche als Zwischenergebnis einen ersten Entwurf eines Definitionsvorschlages lieferten. Die zuvor gewonnenen Erkenntnisse wurden zusammengeführt und es wurden Begriffsmerkmale des Begriffes Data Driven abgeleitet. Die Merkmale wurden in einem morphologischen Kasten festgehalten, welcher als Entscheidungstool dienen könnte, um zu bestimmen ob und inwieweit Prozesse einer aktuellen Begriffsvorstellung datengetriebener Prozesse im Umfeld von Produktion und Logistik entsprechen. Mit Hilfe der gewonnenen Ergebnisse wurden mögliche Unterschiede in den Sichtweisen verschiedener Fachdisziplinen auf den Begriff Data Driven untersucht. Abschließend wurde der Begriff Data Driven von etablierten IT-Prozessen, der Industrie 4.0 und dem Begriff Big Data abgegrenzt.

Eine der wichtigsten gewonnenen Erkenntnisse zu datengetriebenen Prozessen ist, dass nicht der eine datengetriebene Prozess existiert. Es gibt viele unterschiedliche Arten und Weisen datengetriebene Prozesse zu realisieren, welche ebenso zur Bearbeitung unterschiedlicher Aufgabenstellungen geeignet sind. Die aktuellen datengetriebenen Prozesse stellen außerdem keine neuen eigenen Technologien zur Verfügung, sondern nutzen hauptsächlich verschiedene Technologien und Ansätze von Big Data und der Industrie 4.0, darunter meist AutoID Systeme, Data Analysis und Methoden des maschinellen Lernens. Damit geben datengetriebene Prozesse eher eine Art Rahmen vor, welcher vorgibt wie und wofür die vereinten Technologien genutzt werden. Im Zusammenhang mit Data Driven steht vor allem eine proaktive Nutzung von Daten im Vordergrund, um die Produktivität und Qualität zu steigern. Daraus folgt die primär auf Daten fokussierte Betrachtung physischer Prozesse. Daneben hängt der Begriff Data Driven mit dem Einsatz zur Erfüllung konkretisierter Aufgabenstellungen zusammen. Laut Analysen scheinen datengetriebene Prozesse vorwiegend zu operativen Zwecken verwendet zu werden, bieten aber scheinbar auch einen Nutzen für Planungsaufgaben (zumindest auf taktischer Ebene). Datengetriebene Prozesse erfassen und übertragen Daten direkt aus physischen Prozessen. Die erfassten Daten treiben daraufhin die Analysesysteme an (datengetrieben), welche auf eine zuvor festgelegte Aufgabenstellung bezogenes Wissen generieren. Die Ausgabe der Prozesse reicht zum Beispiel von der Visualisierung, Entscheidungsfindungen, Problemlösungen bis hin zu Kontrollanweisungen, welche zu Steuerungszwecken direkt

zum physischen System rückgekoppelt werden können. Vor allem die Erfassung und Verarbeitung von großen Datenmengen in nahezu Echtzeit ist ein Thema im Zusammenhang mit aktuellen datengetriebenen Prozessen im Umfeld von Produktion und Logistik. Dies wird, laut Literatur, durch die Nutzung von AutoID Systemen, Big Data Analysen und dem Einsatz von ML erreicht. Durch den Einsatz hauptsächlich statistischer Methoden in Verbindung mit dem maschinellen Lernen, können aktuelle datengetriebene Prozesse völlig auf die Erstellung physikalischer Modelle verzichten. Dadurch wird eine autonome Arbeitsweise und Untersuchungen dynamischer komplexer Systeme ermöglicht. Indem datengetriebene Prozesse zum Aufbau dynamischer Simulationsmodelle eingesetzt werden, antizipieren diese gegebenenfalls auch zukünftige Ereignisse und ermöglichen den Einsatz von Simulationen in der Betriebsphase. Ein weiteres und häufig diskutiertes Einsatzgebiet datengetriebener Prozesse ist die Entwicklung digitaler Zwillinge, welche insbesondere von den Echtzeitanalysen und der hohen Dynamik dieser Prozesse profitieren.

Aus den gewonnenen Erkenntnissen der Arbeit wird zusammenfassend der folgende **Definitionsvorschlag** für den Begriff Data Driven und für aktuelle Data Driven Processes im Umfeld von Produktion und Logistik formuliert. Eine detaillierte Herleitung einzelner Aussagen des Definitionsvorschlages inklusive derer Quellen, befindet sich im Anhang dieser Arbeit (Tabelle A. 2: Definitionsvorschlag Herleitungstabelle):

Der Begriff **Data Driven** im Umfeld von Produktion und Logistik kennzeichnet eine primär auf Daten fokussierte Betrachtung physischer Prozesse zur proaktiven Erfüllung konkretisierter vorwiegend operativer Aufgabenstellungen. Zu diesem Zweck werden Daten direkt aus physischen Prozessen extrahiert und übertragen, um Analysen anzutreiben, welche gezielt auf die jeweilige Aufgabe bezogenes Wissen generieren.

Informationstechnische Prozesse, welche Data Driven sind, bezeichnen sich als **Data Driven Processes**. Aktuelle Data Driven Processes im Umfeld von Produktion und Logistik bieten eine Möglichkeit, durch den Einsatz von Sensoren, AutoID Systemen und Anbindungen an etablierte informationstechnische Systeme, große Mengen aktueller und historischer Daten in nahezu Echtzeit zu erfassen und zu verarbeiten. Der Aufbau dieser Prozesse kann sich abhängig von der Aufgabenstellung und ihren Einsatzgebieten unterscheiden. Ihre Analysen, welche meist statistische Methoden wie Data Analytics und Methoden des maschinellen Lernens vereinen, ermöglichen eine autonome Untersuchung dynamischer komplexer Systeme. Dazu stellen sie keine neuen eigenen Technologien zur Verfügung, sondern nutzen hauptsächlich verschiedene Technologien und Ansätze von Big Data und der Industrie 4.0. Darüber hinaus können die, von den vorwiegend diagnostischen oder prädiktiven Analysen gelieferten, Erkenntnisse direkt zum physischen Prozess rückgekoppelt werden, um beispielsweise diagnostizierte Fehler direkt zu beheben oder um eine Prozessregelung zu realisieren.



Da bei den erarbeiteten Begriffsmerkmalen datengetriebener Prozesse zwischen mandatorischen und optionalen Merkmalen unterschieden wurde, wurde auch bei der Verfassung des Definitionsvorschlages eine Aufteilung in zwei Teile unternommen. Der erste Teil der Definition befasst sich mit dem Begriff Data Driven und leitet sich hauptsächlich aus den mandatorischen Merkmalen her. Dieser bildet somit den Kern des Begriffes Data Driven im Umfeld von Produktion und Logistik ab, welcher wahrscheinlich von allen datengetriebenen Prozessen in der Produktion und Logistik erfüllt wird und darüber hinaus zeitlich gesehen wahrscheinlich keine großen Veränderungen aufweist. Der zweite Teil des Definitionsvorschlages erläutert aktuelle datengetriebene Prozesse im Umfeld von Produktion und Logistik. Dieser zweite Teil der Definition schließt den ersten Teil mit ein und vereint darüber hinaus die erfassten optionalen Begriffsmerkmale aktueller datengetriebener Prozesse und die Abgrenzungen von Data Driven zur Industrie 4.0, zu Big Data und zu etablierten IT-Prozessen in Produktion und Logistik. Der zweite Teil des Definitionsvorschlags zeichnet sich dadurch aus, dass dieser detaillierter auf die Datennutzung, Anwendung und Verwendung aktueller datengetriebener Prozesse eingeht. Durch den stärkeren Bezug zu aktuellen Prozessen und der Vernachlässigung eines vergangenen und eventuell zukünftigen Wandels dieser Prozesse, bildet dieser Teil des Definitionsvorschlages eine sehr aktuelle aber eventuell auch zeitlich unbeständigere Sicht auf den Begriff Data Driven.

Der gesamte Definitionsvorschlag kann somit als eine komprimierte Zusammenfassung des in der aktuellen Fachliteratur vorherrschenden Begriffsverständnisses zu Data Driven verstanden werden. Dadurch kann dieser einen grundlegenden Überblick über aktuelle datengetriebene Prozesse vermitteln, um darauf aufbauend weitere Forschung zu betreiben oder für praktische Anwendungen eventuelle Nutzenpotentiale zu identifizieren. Zudem kann das Vorgehen der Literaturanalyse und der Definitionsvorschlag als Vorlage dienen, eine allgemeine Definition für den Begriff Data Driven im Umfeld von Produktion und Logistik zu entwickeln. Die grundlegenden Kernaussagen des Definitionsvorschlages werden in den Begriffsmerkmalen des entwickelten morphologischen Kastens (Tabelle 7) festgehalten. Der Kasten könnte dazu genutzt werden, um zu überprüfen, inwiefern bestimmte Prozesse der aktuellen Auffassung datengetriebener Prozesse und somit auch dem Definitionsvorschlag entsprechen. Doch, ob dieser dafür geeignet ist, beziehungsweise ob sich Prozesse auf die unterschiedlichen Merkmalsausprägungen hin überhaupt überprüfen lassen, muss in einer abschließenden Evaluierung untersucht werden.

## 4 Evaluierung der Begriffskriterien und des Definitionsvorschlages

Um die reine Anwendbarkeit des entwickelten morphologischen Kastens datengetriebener Prozesse (Tabelle 7) zu überprüfen, folgt im folgenden Kapitel eine Evaluierung der Ergebnisse. Da es sich bei der Arbeit um eine Grundlagenforschung handelt, ist eine Überprüfung ob und inwieweit die erarbeiteten Merkmale und der Definitionsvorschlag (Abschnitt 3.5) den Begriff Data Driven hinreichend genau abbilden nicht ohne weitere Forschung auf diesem Gebiet möglich. Zunächst wird in Abschnitt 4.1 Evaluierungskonzept auf die Vorgehensweise der Evaluierung eingegangen. Daraufhin werden Beispielprozesse eingeführt (Abschnitt 4.2), anhand denen anschließend die Evaluierung ergebnisoffen durchgeführt wird (Abschnitt 4.3). Zum Abschluss folgt eine Bewertung des Definitionsvorschlags und der Begriffsmerkmale auf Grundlage der Evaluierungsergebnisse (Abschnitt 4.4).

### 4.1 Evaluierungskonzept

Die Ziele der Evaluierung bestehen vor allem darin eine Bewertungsgrundlage für die Verwendbarkeit der zuvor erarbeiteten Begriffsmerkmale (Tabelle 7) und des damit zusammenhängenden Definitionsvorschlages des Begriffes Data Driven (Abschnitt 3.5) zu schaffen. Darüber hinaus können die Ergebnisse der Evaluierung zur Verbreiterung der Wissensbasis genutzt werden, um zukünftige Aufgaben der Entwicklung einer allgemeinen Definition zu identifizieren. Daraus folgt, dass die Evaluierung summativ ist und ein zusammenfassendes Urteil über die Ergebnisse der Arbeit bildet und keinen direkten Einfluss auf die Ergebnisse ausübt. Aufgrund des qualitativen Charakters der Begriffsmerkmale wird ebenfalls eine eher qualitative, vergleichende und relative Evaluierung durchgeführt, welche eingeschränkt messbare Ergebnisse liefert.

Evaluierungsgegenstand ist Tabelle 7: Morphologischer Kasten aktueller datengetriebener Prozesse. Mit der Hilfe von Beispielprozessen wird geprüft, ob die Begriffsmerkmale aus Tabelle 7 auf herkömmliche informationstechnische Prozesse angewandt werden können und darüber hinaus, falls diese anwendbar sind, inwiefern die Ergebnisse möglichen Erwartungen entsprechen. Die Prüfung wird durchgeführt, indem die zutreffenden Merkmale der Beispielprozesse ermittelt und in Tabelle 7 eingetragen werden. Anschließend wird untersucht ob alle Merkmale erfasst werden konnten oder ob einige der Merkmale Probleme bei der Anwendbarkeit bereiten. Ein besonderes Augenmerk wird dabei auf die Erfüllung der mandatorischen Merkmale gelegt. Für den Fall, dass ausreichend viele Merkmale ermittelt werden können, kann der durch die Merkmale der Beispielprozesse ermittelte Verlauf mit dem, aus den Analysen ermittelten, idealen Verlauf verglichen werden. Dabei wird darauf geachtet ob und wie

viele der voll zutreffenden und der gar nicht zutreffenden Merkmale erfüllt werden. Außerdem wird gesondert beachtet ob alle mandatorischen Merkmale erfüllt werden, um den Prozess Data Driven zu nennen. Sollte einer der Beispielprozesse nicht alle der mandatorischen Merkmale erfüllen, wird dieser als nicht Data Driven eingestuft. Wenn einer der Beispielprozesse neben der Erfüllung aller mandatorischen Merkmale ein nicht zutreffendes Merkmal erfüllt, wird dieser nicht als ein aktueller datengetriebener Prozess eingestuft. Für den Fall, dass ein Beispielprozess alle voll zutreffenden und keine nichtzutreffenden Merkmale erfüllt, wird dieser als datengetriebener Prozess eingestuft. Falls keines der zuvor genannten Szenarien eintritt, ist es möglich, dass es sich beim Beispielprozess um einen aktuellen datengetriebenen Prozess handelt. Die zuvor beschriebenen Fälle können in der folgenden Matrix (Abbildung 16) festgehalten werden:

		<i>alle mandatorischen Merkmale</i>			<i>gar nicht zutreffendes Merkmal</i>			<i>alle voll zutreffenden Merkmale</i>			<i>Grad von Data Driven</i>		
Erfüllung	X	X	X	kein aktueller datengetriebener Prozess									
	X	X	-	kein aktueller datengetriebener Prozess									
	X	-	X	aktueller datengetriebener Prozess									
	X	-	-	möglicherweise aktueller datengetriebener Prozess									
	-	X	X	nicht Data Driven									
	-	X	-										
	-	-	X										
	-	-	-										

Abbildung 16: Evaluierungsmatrix

In Abbildung 16 wird festgehalten, welcher Grad von Data Driven, je nach Erfüllung bestimmter Merkmale, in der Evaluierung als mögliches Ergebnis angenommen wird. Für den Fall, dass nicht klar beantwortet werden kann, ob ein Beispielprozess als aktueller datengetriebener Prozess bezeichnet werden kann (möglicherweise aktueller datengetriebener Prozess), können die Muster des idealen Prozesses und des Beispielprozesses verglichen werden, um abschätzen zu können, inwieweit diese übereinstimmen, beziehungsweise wo die größten Unterschiede herrschen. Das Ziel der Evaluierung gilt als erreicht, wenn der morphologische Kasten alle mandatorischen Merkmale und einen Großteil der weiteren optionalen Merkmale der Beispielprozesse erfassen kann. Darüber hinaus kann geprüft werden, inwiefern ein zuvor angenommener Grad von Data Driven der Beispielprozesse mit dem in der Evaluierung ermittelten Grad übereinstimmt.

## 4.2 Einführung von Beispielprozessen

Für die Durchführung der Evaluierung werden im Folgenden drei Beispielprozesse eingeführt. Diese wurden aus der Literatur eruiert. Bei der Prozessauswahl wird versucht Prozesse zu erfassen, zu denen mögliche Annahmen zum Grad von Data Driven (Abbildung 16) im Vorfeld getroffen werden können. Dabei besteht allerdings das Problem, dass Prozesse beispielsweise datengetrieben sein könnten, selbst wenn diese zuvor nicht als solche bezeichnet worden sind. Wichtig für die Überprüfung der Anwendbarkeit sind aber vor allem Angaben zur Datennutzung, zur Anwendung und zu den Aufgaben der Beispielprozesse, da diese in den Merkmalen erfasst werden.

### **Beispielprozess A:** Knowledge Discovery in Databases (KDD)

KDD beschäftigt sich mit dem gesamten Prozess der Wissensentdeckung aus Daten, vom Abrufen und Vorbereiten von Daten bis zum Bereitstellen und Interpretieren der Ergebnisse (Fayyad, 1996). Der KDD sucht mit seinen Analysen nach bislang unbekanntem Mustern in großen und teils komplexen Datenbeständen. Die Analysen werden mit Data Mining Methoden durchgeführt, welche auch Data Analytics und Methoden des maschinellen Lernens nutzen (Fayyad, 1996). Damit ähneln KDD Prozesse der Beschreibung aktueller datengetriebener Prozesse. Fayyad, 1996 beschreibt KDD Prozesse als *User Driven*, da der\*die Benutzer\*in einen Einfluss auf die Datenvorverarbeitung, der Auswahl der Analysen und der Ergebnisinterpretation aufweist. An anderer Stelle bezeichnet er KDD jedoch indirekt als Data Driven mit dem Verweis darauf, dass Wissen das Endprodukt einer datengetriebenen Suche ist (Fayyad, 1996). Aus den Aussagen lässt sich die Vermutung aufstellen, dass es sich bei KDD wahrscheinlich um einen nicht aktuellen datengetriebenen Prozess handeln könnte.

### **Beispielprozess B:** Speicherprogrammierbare Steuerung (SPS)

SPS gehören zu den etablierten IT-Systemen (Abschnitt 2.2) und stellen rechnergestützte Automatisierungseinrichtungen dar, welche derzeit die Standardtechnologie für industrielle Steuerungsaufgaben sind (Bernstein, H. 2018). Diese setzen sich aus Signalbaugruppen, einer Zentraleinheit, Programmspeicher und Feldbuskopplern zusammen und besitzen eine Schnittstelle zu einem Programmiergerät (Bernstein, H. 2018). Das Programmiergerät gibt der Steuerung Anweisungen vor, welche zyklisch oder ereignisgesteuert von dieser ausgeführt werden. Da sich die Programme der SPS auf Abläufe des physischen Systems beziehen, sind SPS wahrscheinlich nicht Data Driven.

### Beispielprozess C: Data Driven Process Design (DDPD)

Im Bereich der Produktion ermöglicht ein DDPD Daten aus der Fertigung für das Design beispielsweise neuer Fertigungsprozesse zu nutzen. Anstelle eines Vorwärtsdesigns, welches zum Festlegen neuer Prozessparameter auf Experimenten und Simulationen aufbaut, verwendet das DDPD ein von den Daten ausgehendes Rückwärtsdesign. Das Rückwärtsdesign analysiert und verarbeitet mit Hilfe von Data Mining Methoden vorliegende Prozessdaten (z. B. aus einem MES), um Zusammenhänge zwischen Prozessqualität und Prozessparameter zu finden. Darüber hinaus erfasst das DDPD Fertigungsdaten direkt aus den physischen Prozessen. Das gewonnene Wissen aus den Analysen und die erfassten Daten zum Produktionszustand werden beide zum Design des neuen Prozesses verwendet, welcher wiederum Fertigungsdaten erzeugt, die erfasst werden (Wei et al., 2020). Aufgrund der Bezeichnung, wird beim DDPD davon ausgegangen, dass es sich dabei um einen datengetriebenen Prozess handelt.

## 4.3 Durchführung der Evaluierung

Bei der Durchführung der Evaluierung werden die zuvor in Abschnitt 4.2 beschriebene Beispielprozesse nacheinander untersucht. Zunächst wird die Anwendbarkeit der Tabelle 7: Morphologischer Kasten aktueller datengetriebener Prozesse geprüft, indem die entsprechenden Merkmale der Beispielprozesse ermittelt und in Tabelle 7 eingetragen werden. Geprüft wird, ob alle Merkmale des Kastens ermittelt werden können. Anschließend wird mit Hilfe von Abbildung 16: Evaluierungsmatrix getestet, ob die Ergebnisse des Grades von Data Driven mit den zuvor getroffenen Annahmen der Beispielprozesse übereinstimmen, um eventuelle Verbesserungspotentiale des morphologischen Kastens zu identifizieren.

### Beispielprozess A: Knowledge Discovery in Databases (KDD)

Tabelle 11: Morphologischer Kasten Beispielprozess A

Kategorie	Merkmal	Merkmalsausprägungen			Kommentare
Mandatorische Merkmale	Prozessbetrachtung	Primär auf Daten ausgerichtet	✓		KDD fokussiert sich auf den Gesamtprozess der Wissensuche aus Daten. (S. 39)
	Aufgabenstellung	Eine oder mehrere konkrete Aufgabenstellungen	✓		Zuerst wird das Ziel des KDD aus Kundensicht identifiziert. (S. 42)

	End- oder Zwischenprodukt des Prozesses	Wissen	✓		KDD fokussiert sich auf den Gesamtprozess der Wissensuche aus Daten. (S. 39)	
	Datenextraktion	Indirekte Datenerfassung aus physischen Prozessen	✗		Nutzen von Datenbanken und manuelle Erstellung eines Zieldatensatzes. (S. 42)	
Optionale Merkmale	Datenmerkmale	Datenerfassung	Übernimmt Daten etablierter IT-Systeme	●		Nutzen von Datenbanken. (S. 42)
		Echtzeiterfassung	Echtzeitdatenerfassung und Analysen möglich und bedeutend	●		Eine Echtzeitdatenerfassung ist durch die Integration von Sensorsystemen möglich. (S. 50)
		Datenstruktur	Verarbeitung unterschiedlicher Datenstrukturen ist teilweise möglich			Die Datentransformation ist ein Teil des KDD. Zudem können unterschiedliche Datentypen verarbeitet werden. (S. 48)
		Datenvolumen	Verarbeitung/ Erfassung großer Datenmengen ist möglich, aber nicht primär bedeutend	●		Die Verarbeitung der großen und schnell anwachsenden Datenmengen, ist auch ein Thema des KDD. (S. 37)
		Datenvorverarbeitung	Manuelle Datenvorverarbeitung durch Experten*innen	✗		Eine Datenvorverarbeitung ist nötig, diese wird jedoch nicht automatisiert. Z. B. manuelle Auswahl von Zieldatensätze oder auch Transformationsoperationen. (S. 50)
	Aufgabenmerkmale	Aufgabenorientierung	Der Prozess wird für taktische Aufgaben eingesetzt	●	Der Prozess wird für strategische Aufgaben eingesetzt	Nutzen des gewonnenen Wissens beispielsweise für Planungsaufgaben. Wird aber auch zur Verbesserung der Mensch-Maschine Interaktion verwendet (evtl. auch operativer Nutzen). (S. 39-40)
		Produktivität	-		-	-
		Systemkomplexität	Ermöglicht Analysen teils komplexer Systeme			Verarbeitung auch nicht linearer viel dimensionierter Systeme. (S. 43)

Anwendungsmerkmale	Analyse-system	Das Analysesystem arbeitet prädiktiv	●	Das Analysesystem arbeitet deskriptiv	Das Ziel der Wissenssuche lässt sich in prädiktive und deskriptive Ziele unterteilen. (S. 43)
	Analysemethoden	Nutzen statistischer Methoden (z. B. Data Analytics) und künstlicher Intelligenzen	✓		KDD Analysen nutzen Data Mining, welches unter anderem statistische Datenanalysen und auch Methoden des maschinellen Lernens verwenden. (S. 43)
	Autonome Arbeitsweise	Prozess arbeitet teilweise autonom			Die Analysesysteme arbeiten teils autonom. Jedoch die Vorverarbeitung und die Ergebnisinterpretation, erfordern Expertenwissen. (S. 48)
	Rückkopplung	Rückkopplung zum physischen System ist möglich, aber nicht notwendig	●		Das gewonnene Wissen kann direkt im System einbezogen werden. (S. 42)

✓ = trifft voll zu   ● = trifft teilweise zu   ✗ = trifft gar nicht zu

In Tabelle 11 sind die erfassten Merkmale des KDD dargestellt, welche den Aussagen von Fayyad, 1996 entnommen werden konnten, inklusive Kommentare zur Begründung der getroffenen Merkmalsausprägungen. Bezogen auf die Anwendbarkeit des morphologischen Kastens konnten alle Merkmale bis auf eines identifiziert werden. Zum Aufgabenmerkmal der Produktivität lassen sich keine direkten Aussagen aus dem Text eruieren. Zu weiteren Merkmalen, wie zum Beispiel dem Analysesystem, konnten gleich mehrere Merkmalsausprägungen gefunden werden, was aber keinen Widerspruch in der Anwendbarkeit bedeutet. Da der KDD nicht alle mandatorischen Merkmale erfüllt, wird dieser laut Abbildung 16 als nicht Data Driven eingestuft. Laut Aussage von Fayyad, 1996 wird der Prozess jedoch als Data Driven bezeichnet. Daraus lässt sich folgern, dass es sich beim nicht erfüllten mandatorischen Merkmal der Datenextraktion eventuell um ein nicht notwendiges Merkmal von Data Driven handeln könnte, sondern vielleicht eher um ein Datenmerkmal. Um dies zu überprüfen könnten weitere Prozesse getestet werden, welche laut Literatur datengetrieben sind, aber wahrscheinlich nicht der Auffassung aktueller datengetriebener Prozesse. Wären alle der mandatorischen Merkmale des Prozesses erfüllt, wäre der KDD Data Driven aber würde nicht die Merkmale aktueller datengetriebener Prozesse im Bereich von Produktion und Logistik erfüllen, da die Datenvorverarbeitung laut Literatur manuell durch Experten durchgeführt wird. Auch die deskriptiven Analysen sind ein Indiz, welches eher gegen einen aktuellen datengetriebenen Prozess spricht.

**Beispielprozess B: Speicherprogrammierbare Steuerung (SPS)****Tabelle 12: Morphologischer Kasten Beispielprozess B**

Kategorie	Merkmal	Merkmalsausprägungen			Kommentare	
Mandatorische Merkmale	Prozessbetrachtung	Primär auf den Prozess ausgerichtet	✘		Programmierung der SPS bildet mit Hilfe einfacher logischer Operatoren Anweisungen der Prozesse ab (S. 33)	
	Aufgabenstellung	Eine oder mehrere konkrete Aufgabenstellungen	✓		SPS hat die konkrete Aufgabe physische Prozesse zu steuern (S. 12-13)	
	End- oder Zwischenprodukt des Prozesses	Geht nicht über Informationen hinaus	✘		Ausgangssignale stellen kein neu gewonnenes Wissen dar, sondern sind das Ergebnis abgearbeiteter Anweisungen (S. 18-23)	
	Datenextraktion	Direkte Datenerfassung aus physischen Prozessen	✓		Ermitteln mit Hilfe von Sensoren Daten unmittelbar aus physischen Prozessen (S. 13)	
Optionale Merkmale	Datenmerkmale	Datenerfassung	Extrahiert Daten aus physischen Prozessen	✓		Ermitteln mit Hilfe von Sensoren Daten unmittelbar aus physischen Prozessen (S. 13)
		Echtzeiterfassung	Echtzeitdatenerfassung und Analysen möglich und bedeutend	●		Eine möglichst kurze Reaktionszeit auf ein eingehendes Signal ist bedeutend. Sensoren übermitteln Daten direkt.
		Datenstruktur	Verarbeitung unterschiedlicher Datenstrukturen ist nicht möglich			Verarbeitung binärer oder analoger Eingangssignale (S. 7-12)
		Datenvolumen	Verarbeitung/ Erfassung großer Datenmengen ist nicht möglich und unbedeutend			In der Regel ist die Verarbeitung großer Datenmengen nicht bedeutend. Die Verarbeitung größerer Datenmengen führt zu einer Verringerung der Geschwindigkeit (S. 30)



Aufgabenmerkmale	Datenvorverarbeitung	Keine Datenvorverarbeitung			Im Falle binärer Eingangssignale keine Datenverarbeitung nötig. Bei analogen Eingangsdaten: digitalisieren. (S. 11)
	Aufgabenorientierung	Der Prozess wird für operative Aufgaben eingesetzt	✓		überwiegend dient eine SPS zur Steuerung operativer Prozesse (S. 12-13)
	Produktivität	-		-	-
	Systemkomplexität	Ermöglicht Analysen teils komplexer Systeme			SPS sind für komplexe Automatisierungsaufgaben geeignet. (S. 43-44)
Anwendungsmerkmale	Analysesystem	Das Analysesystem arbeitet präskriptiv		Das Analysesystem arbeitet deskriptiv ●	SPS schreibt präskriptiv Anweisungen vor und kann darüber hinaus lassen sich Operationen im Programmablaufplan anschaulich wiedergeben (S. 448)
	Analysemethoden	Nutzen von Modellen des physischen Systems			SPS Programme werden durch textuelle oder grafische Programmierung realisiert, welche Wissen über das zu steuernde physische System voraussetzen.
	Autonome Arbeitsweise	Prozess arbeitet vollständig autonom	✓		Automatisierungsgeräts fragt die Eingänge auf die Zustände „Spannung vorhanden“ und „Spannung nicht vorhanden“ ab (S. 59)
	Rückkopplung	Rückkopplung zum physischen System ist möglich und notwendig		●	Ausgang der SPS direkt mit den Aktoren des physischen Systems verbunden (S. 13)

✓= trifft voll zu   ●= trifft teilweise zu   ✗= trifft gar nicht zu

In Tabelle 12 sind die erfassten Merkmale der SPS dargestellt, welche den Aussagen von Bernstein, H. 2018 entnommen werden konnten, inklusive Kommentare zur Begründung der getroffenen Merkmalsausprägungen. Wie beim Beispielprozess A konnten, bezogen auf die Anwendbarkeit des morphologischen Kastens, alle Merkmale bis auf das Aufgabenmerkmal der Produktivität identifiziert werden. Zwei der mandatorischen Merkmale sind nicht erfüllt, daher entspricht laut Abbildung 16 das Ergebnis des Grades von Data Driven den Erwartungen nämlich, dass eine SPS nicht Data Driven ist.

**Beispielprozess C: Data Driven Process Design (DDPD)****Tabelle 13: Morphologischer Kasten Beispielprozess C**

Kategorie	Merkmal	Merkmalsausprägungen			Kommentare
Mandatorische Merkmale	Prozessbe- trachtung	Primär auf Daten ausgerichtet	✓		Die Abläufe des physi- schen Prozesses sind für die Analysen unbedeu- tend. Das DDPD verwen- det hauptsächlich erfasste Daten des physischen Prozesses.
	Aufgabenstel- lung	Eine oder mehrere konkrete Aufgaben- stellungen	✓		Zum Unterstützen von Designentscheidungen und zur Optimierung von Designs (S. 731)
	End- oder Zwischenpro- dukt des Pro- zesses	Wissen	✓		Verwenden von durch Data Mining erzeugtem Wissen und Muster. (S. 731)
	Datenextrak- tion	Direkte Datenerfas- sung aus physischen Prozessen	✓		direkte Datenentnahme aus z. B. MES Systemen und direkt aus dem physi- schen Prozess (S. 731)
Optionale Merkmale	Datenmerkmale	Datenerfas- sung	✓	Übernimmt Daten etab- lierter IT- Systeme	• direkte Datenentnahme aus z. B. MES Systemen und direkt aus dem physi- schen Prozess (S. 731)
		Echtzeiterfas- sung			Direkte Datenentnahme, autonome Arbeitsweise und der Einsatz von Deep Learning ermöglichen theoretisch schnelle Da- tenanalysen, welche fürs DDPD nicht zwingend notwendig sind.
		Datenstruktur		•	Strukturierte- und un- strukturierte Daten wer- den gegeben Falls zur weiteren Verarbeitung fusioniert. (S. 729)

	Datenvolumen	Verarbeitung/ Erfassung großer Datenmengen ist möglich, aber nicht primär bedeutend	●		große Mengen an Fertigungsdaten, welche ein großes Volumen, hohe Geschwindigkeiten und eine hohe Vielfalt aufweisen. (S. 728)
	Datenvorverarbeitung	Automatische Datenverarbeitung	✓		Datenvorverarbeitung zum Beispiel mit Hilfe von Cluster-Algorithmen. (S. 730)
Aufgabenmerkmale	Aufgabenorientierung	Der Prozess wird für taktische Aufgaben eingesetzt	●		Zum Entwurf von Fertigungsprozessen (daher eher taktisch).
	Produktivität	-	-	-	-
	Systemkomplexität	Ermöglicht Analysen teils komplexer Systeme			Ermöglicht die Analysen höher dimensionierter Daten, durch die Unterstützung von Data Mining und Deep Learning (S. 730-733)
Anwendungsmerkmale	Analysesystem	Das Analysesystem arbeitet prädiktiv	●		Nutzt Regressions- und Klassifikationsmethoden, welche prädiktiv eingesetzt werden. (S. 730)
	Analysemethoden	Nutzen statistischer Methoden (z. B. Data Analytics) und künstlicher Intelligenzen	✓		Nutzt Data Mining, welches statistische Methoden verwendet. Verwendet künstliche neuronale Netze zur automatischen Fehlererkennung. (S. 731-733)
	Autonome Arbeitsweise	Prozess arbeitet vollständig autonom	✓		Mit Ausnahme der aus den Erkenntnissen gewonnenen Designentscheidungen, arbeitet der Prozess autonom. (S. 729-733)
	Rückkopplung	Rückkopplung zum physischen System ist möglich und notwendig	●		gewonnenen Erkenntnisse werden zum physischen System rückgekoppelt, um dieses weiter zu optimieren. (S. 729)

✓ = trifft voll zu    ● = trifft teilweise zu    ✗ = trifft gar nicht zu

In Tabelle 13 sind die erfassten Merkmale des DDPD dargestellt, welche den Aussagen von Wei et al., 2020 entnommen werden konnten. Auch bei der Untersuchung dieses Prozesses konnten keine Aussagen bezogen auf das Merkmal Produktivität aus der untersuchten Quelle eruiert werden. Zum mandatorischen Merkmal der Datenextraktion hätte ebenfalls die indirekte Datenerfassung aus physischen Prozessen vermerkt werden können, da der Prozess sowohl Daten aus etablierten IT-Systemen wie auch direkt aus dem Prozess ermittelt. Jedoch überwiegt die Erfüllung des Merkmals gegenüber der nicht Erfüllung. So wie in Abschnitt 4.2 erwartet erfüllt das DDPD alle vier mandatorischen Merkmale, weshalb dieser laut Abbildung 16 als Data Driven bezeichnet werden kann. Darüber hinaus sind alle weiteren Merkmalausprägungen links angeordnet und keines dieser Merkmale sind welche, die gar nicht zutreffen. Dies deutet darauf hin, dass es sich beim DDPD sehr wahrscheinlich um einen aktuellen datengetriebenen Prozess handelt. Dies würde auch den Erwartungen aus Abschnitt 4.2 entsprechen, welche davon ausgegangen sind, dass es sich beim DDPD um einen aktuellen datengetriebenen Prozess handelt.

#### 4.4 Bewertung der Begriffskriterien und des Definitionsvorschlages

Auf Grundlage der zuvor durchgeführten Evaluierung (Abschnitt 4.3) folgt eine Einschätzung der Anwendbarkeit des morphologischen Kastens aktueller datengetriebener Prozesse (Tabelle 7). Darüber hinaus wird diskutiert, ob die aufgestellten Merkmale und somit auch der Definitionsvorschlag als nützlich erachtet werden, um ein besseres Verständnis für den Begriff Data Driven im Umfeld von Produktion und Logistik aufzubauen und an welchen Stellen durch die Evaluierung eventuelle Verbesserungspotentiale und Aufgaben für weitere Forschung auf diesem Gebiet identifiziert werden konnten.

In der Evaluierungsdurchführung wurde der morphologische Kasten mit der Hilfe drei zuvor eingeführter Beispielprozesse auf seine Anwendbarkeit überprüft. Bei fast allen Merkmalen gab es bei der Ermittlung der spezifischen Merkmalsausprägungen der Beispielprozesse keine Probleme. Bei der Ermittlung der Ausprägung der Produktivität gab es bei allen drei Beispielprozessen entweder keine oder zu ungenaue Angaben, um eine bestimmte Ausprägung zuzuweisen. Dies könnte zum einen daran liegen, dass eine Produktivitätssteigerung beim Einsatz mancher Prozesse als selbstverständlich angenommen wird und daher nicht immer speziell in der Literatur erwähnt wird. Zum anderen ist die Merkmalsausprägung durch die Worte *deutliche Steigerung* und *geringe Steigerung* subjektiv behaftet, was genaue Angaben der Ausprägung erschwert. Dies trifft ebenso auf die Systemkomplexität und auf das Datenvolumen zu. Die Systemkomplexität wird unter anderem in *hochkomplexe Systeme* und *teils komplexe Systeme* unterschieden. So wie bei Big Data ist auch bei aktuellen datengetriebenen Prozessen die Verarbeitbarkeit großer Datenmengen ein Thema. Doch auf die Frage ab wann eine Datenmenge als groß gilt, gibt es keine klare Antwort. Daher konnten Aussagen zur ana-

lysierbaren Systemkomplexität und zum Datenvolumen zwar der Literatur entnommen werden, diese sind jedoch subjektiv und zeit- und kontextabhängig. Daher kann der morphologische Kasten höchstens zur begrifflichen Einordnung von Prozessen in Hinblick auf den Begriff Data Driven, nicht aber zum Vergleich der Prozesse untereinander verwendet werden. Abgesehen vom Merkmal Produktivität ist der morphologische Kasten laut Evaluierung anwendbar. Dies ist jedoch auch aufgrund der geringen Anzahl an durchgeführten Tests keine gesicherte Erkenntnis, sondern nur eine erste Einschätzung, die getroffen werden kann. Zur Sicherstellung und Verbesserung der Anwendbarkeit, sollten weitere Beispielprozesse zu Testzwecken herangezogen werden.

Da keine allgemeine Definition des Begriffes Data Driven existiert, ist es ohne weiteres nicht möglich die erarbeiteten Begriffsmerkmale datengetriebener Prozesse auf ihren Wahrheitsgehalt hin zu überprüfen. Bei den Beispielprozessen (Abschnitt 4.2) wurden zwar Annahmen zu ihren Grad von Data Driven getroffen, jedoch sind dies keine festen Tatsachen, welche eine Beweisführung ermöglichen würden. Dennoch können durch Vergleiche der zuvor getroffenen Annahmen und der in der Evaluierung ermittelten Grade von Data Driven Vermutungen bezüglich Unstimmigkeiten der Begriffsmerkmale und des Definitionsvorschlags aufgestellt werden, um daraus mögliche Konsequenzen für zukünftige Aufgaben auf dem Weg zu einer allgemeinen Begriffsbestimmung von Data Driven abzuleiten. Die folgenden Grade von Data Driven wurden jedoch ohne Beachtung des Merkmals Produktivität ermittelt, da für dieses kein Merkmal bestimmt werden konnte. Sowohl die zuvor getroffenen Annahmen von Beispielprozess B als auch die Annahmen von Beispielprozess C stimmten mit den in der Evaluierung, durch Anwendung des morphologischen Kastens und der Abbildung 16: Evaluierungsmatrix, ermittelten Graden von Data Driven überein. Die SPS erfüllte nicht alle mandatorischen Merkmale und wurde somit als nicht Data Driven eingestuft. Das DDPD erfüllte alle mandatorischen Merkmale und erfüllte einen großen Teil der optionalen Merkmale, sodass es sich bei diesem Prozess, laut Abbildung 16, sehr wahrscheinlich um einen aktuellen datengetriebenen Prozess handelt. Beim Beispielprozess A hingegen wurden die zuvor getroffenen Annahmen bei der Anwendung des morphologischen Kastens nicht bestätigt. Durch die hohe Ähnlichkeit zu den in Kapitel 3 untersuchten Prozessen und den von Fayyad, 1996 angeführten Verweis, dass Wissen das Endprodukt einer datengetriebenen Suche ist, wurde zuvor angenommen, dass der KDD wahrscheinlich Data Driven aber kein aktueller datengetriebener Prozess ist. Dieser erfüllt jedoch nicht das mandatorische Merkmal der Datenextraktion und wird laut Abbildung 16 daher nicht als Data Driven bezeichnet. Ob dieses durch die Evaluierung gewonnene Ergebnis der Wahrheit entspricht, kann an dieser Stelle nicht ermittelt werden. Die angenommenen Unstimmigkeiten bezogen auf das Merkmal Datenextraktion können aber auch darauf hinweisen, dass es sich bei diesem Merkmal eher um ein schwächeres optionales Merkmal handeln könnte. Die hohe Gewichtung der direkten Datenerfassung aus physischen Prozessen kann beispielsweise das Ergebnis der Literatursuche für die Begriffsbestimmung sein. In dieser wurden hauptsächlich aktuelle und darunter viele

operative Anwendungen in der Produktion betrachtet. Dadurch könnte beispielsweise das Kriterium der direkten Datenextraktion als notwendig erachtet worden sein, obwohl dies ebenso als ein Merkmal einer intelligenten Fertigung angesehen werden könnte. Ein eventuell besseres Merkmal könnte die Anforderung an einen durchgängigen Datenfluss sein, aber ebenso wie die Richtigkeit des Merkmals Datenextraktion muss dies durch weitere Forschung ermittelt werden.

Zusammenfassend kann, nach einer ersten kurzen Evaluierung, dem morphologischen Kasten datengetriebener Prozesse eine gute Anwendbarkeit bestätigt werden. Bis auf wenige Ausnahmen entsprachen die erfassten Merkmalsausprägungen und die erfüllten Grade von Data Driven auch weitestgehend den zuvor getroffenen Annahmen. Aus diesem Grund kann die Vermutung aufgestellt werden, dass die Begriffsmerkmale datengetriebener Prozesse und der Definitionsvorschlag des Begriffes Data Driven als Grundlage für weitere tiefergehende Forschungen auf dem Gebiet der datengetriebenen Prozesse geeignet sind. Durch weitere Literatur- und Prozessanalysen kann der Kasten eventuell um weitere Merkmale ergänzt werden. Zudem könnte in Zukunft eine genauere Gewichtung der Merkmale vorgenommen werden. In einem weiteren Schritt müssten die Merkmale in Unternehmen an laufenden informationstechnischen Prozessen geprüft werden, um letztendlich zu einem allgemeinen Begriffsverständnis von Data Driven im Umfeld von Produktion und Logistik beitragen zu können.

## 5 Zusammenfassung und Ausblick

Die Problemstellung, mit der sich diese Arbeit beschäftigt, ist das Fehlen einer eindeutigen Definition des Begriffes Data Driven im Umfeld von Produktion und Logistik und das damit einhergehende subjektive Begriffsverständnis. Damit stellt die Arbeit eine Grundlagenforschung auf dem Gebiet datengetriebener Prozesse dar, welche die Problematik, der derzeit stetig wachsenden und sich verändernden Begriffswelt infolge der fortschreitenden Digitalisierung aufgreift. Die wesentlichen Ziele der Arbeit sind einen Beitrag zu einem besseren Verständnis aktueller datengetriebener Prozesse im Umfeld von Produktion und Logistik zu leisten und die Abgrenzung vom Begriff Data Driven zur Industrie 4.0, zu Big Data und zu etablierten informationstechnischen Prozessen in der Produktion und Logistik. Zur Erfüllung dieser Aufgabenstellung wurde eine systematische Literaturanalyse, angelehnt an die Methode der qualitativen Inhaltsanalyse nach Mayring, durchgeführt und parallel wurden Kernaussagen zum Begriff Data Driven zusammengefasst. Auf Grundlage der Analyseergebnisse und der Kernaussagen wurden Begriffsmerkmale erarbeitet, die als Ausgangspunkt für die Entwicklung eines Definitionsvorschlags dienen.

Die Ergebnisse der Arbeit lassen sich in die folgenden drei Elemente unterteilen:

- Begriffsmerkmale und Morphologischer Kasten aktueller datengetriebener Prozesse (Tabelle 7)
- Abgrenzung des Begriffes Data Driven zu etablierten IT-Prozessen in der Produktion und Logistik, der Industrie 4.0 und Big Data (Abschnitt 3.4)
- Definitionsvorschlag für den Begriff Data Driven im Umfeld von Produktion und Logistik (Abschnitt 3.5)

Die erarbeiteten Begriffsmerkmale und ihre Ausprägungen bilden das aktuelle in der Fachliteratur vorherrschende Begriffsverständnis datengetriebener Prozesse ab. Diese unterscheiden speziell mandatorische und optionale Merkmale. Mit Hilfe der Begriffsmerkmale wurde mit dem morphologischen Kasten aktueller datengetriebener Prozesse ein Entscheidungstool entwickelt, um zu überprüfen, inwieweit einzelne Prozesse mit der aktuellen Auffassung datengetriebener Prozesse übereinstimmen. Die zuvor gewonnenen Erkenntnisse wurden zur Abgrenzung des Begriffes Data Driven zu anderen Begriffen genutzt. Dabei wurde vor allem festgestellt, dass datengetriebene Prozesse im Gegensatz zu etablierten Prozessen keine Modelle physischer Prozesse, sondern statistische Analysen und Methoden des maschinellen Lernens nutzen. Diese können als Teil der Industrie 4.0 angesehen werden, da sie die Umsetzung mancher Themen der Industrie 4.0 ermöglichen. Die Begriffe Data Driven und Big Data ähneln sich sehr. Einer der größeren Unterschiede konnte im Prozessbezug datengetriebener Prozesse erfasst werden. Unter Zuhilfenahme der Begriffsmerkmale, der Abgrenzung zu anderen Begriffen und eines zuvor als Zwischenfazit formulierten vorläufigen Definitionsvorschlags (Abschnitt 3.1.3), wurde ein Definitionsvorschlag für den Begriff Data Driven im

Umfeld von Produktion und Logistik entwickelt. Dieser greift die Unterteilung der Begriffsmerkmale in mandatorische und optionale Merkmale auf und erläutert auf der einen Seite den Begriff Data Driven und auf der anderen Seite eine aktuelle Auffassung datengetriebener Prozesse im Umfeld von Produktion und Logistik.

Bezüglich der Ergebnisinterpretation konnte aus den gewonnenen Erkenntnissen zwar keine allgemeine Begriffsdefinition, aber ein Definitionsvorschlag des Begriffes Data Driven und aktueller datengetriebener Prozesse im Umfeld von Produktion und Logistik verfasst werden. Der entwickelte Definitionsvorschlag bietet eine wichtige Grundlage, um Data Driven auch in weiteren Forschungen gedanklich besser einzuordnen. Laut Literaturanalyse haben datengetriebene Prozesse mit dem Aufkommen neuer für sie nutzbarer Technologien einen Wandel durchlaufen, sodass nicht ausgeschlossen werden kann, dass sich der Begriff infolge der fortschreitenden Digitalisierung weiter verändert. Der erarbeitete morphologische Kasten datengetriebener Prozesse (Tabelle 7) kann zukünftig dabei helfen ein allgemeines Verständnis für datengetriebene Prozesse aufzubauen und Prozesse auf ihren Bezug zu Data Driven hin zu untersuchen. Seine Anwendbarkeit wurde in Kapitel 4 anhand dreier Beispielprozesse überprüft. Die Evaluierung hat gezeigt, dass der morphologische Kasten eine sehr gute Anwendbarkeit hat. Darüber hinaus konnte der Kasten zuvor getroffene Annahmen zum Grad von Data Driven weitestgehend bestätigen, wodurch angenommen wird, dass die erarbeiteten Begriffsmerkmale und der verfasste Definitionsvorschlag den Begriff Data Driven relativ gut abbilden. Somit dient der morphologische Kasten nicht nur als Grundlage, zur Vermittlung eines ersten Eindruckes zu aktuellen datengetriebenen Prozessen, sondern auch als Ausgangspunkt für weitere Forschung auf diesem Gebiet. Die Literaturanalyse und die Entwicklung des morphologischen Kastens können als Vorbild für die Ableitung weiterer Begriffsmerkmale dienen, um eine genauere und stets aktuelle Auffassung des Begriffes Data Driven abzubilden. Dadurch, dass sich datengetriebene Prozesse auch untereinander unterscheiden können, ist es schwer allgemeine Aussagen in der Literatur zu finden, oder darüber hinaus selbst zusammenzutragen. Die Aussagen beziehen sich meist auf spezielle Prozesse. Da sich darüber hinaus die Angaben von Autoren\*innen der betrachteten Fachdisziplinen sehr ähneln, ist es schwer die Unterschiede in deren Auffassungen zum Begriff Data Driven zu erfassen. Eine Abgrenzung von Data Driven zu etablierten IT-Prozessen und der Industrie 4.0 konnte leicht vorgenommen werden. Eine Abgrenzung zu Big Data fiel eher schwer, aufgrund einer hohen Ähnlichkeit beider Begriffe. Aktuelle datengetriebene Prozesse scheinen sogar teilweise von Big Data abhängig zu sein, da diese Voraussetzungen für die Verwirklichung datengetriebener Strategien im Bereich von Produktion und Logistik bieten. Durch die gewonnenen Erkenntnisse aus der Literaturanalyse, lässt sich der Begriff Data Driven vor allem über den speziellen Bezug zu seiner Aufgabe abgrenzen.

Eine allgemeine Auffassung des Begriffes Data Driven kann die Arbeit schon allein aus dem Grund nicht abbilden, da durch die Formulierung der Aufgabenstellung der Arbeit an sich aber auch durch zum Beispiel Eingrenzungen bei der Literaturanalyse Beschränkungen



der Forschung vorgenommen wurden. Die erste Begrenzung betrifft den Untersuchungsbe-  
reich, welcher auf das Umfeld von Produktion und Logistik festgelegt wurde. Die Ergebnisse  
liefern daher keine Aussagen zur Nutzung datengetriebener Prozesse außerhalb dieses Be-  
reichs. Durch eine Eingrenzung der Suchbegriffe und Suchkriterien, um Literatur zu datenge-  
triebenen Prozessen im genannten Umfeld zu finden, wurde überwiegend Literatur ermittelt,  
welche tiefergehende Beschreibungen spezifischer Anwendungen datengetriebener Prozesse  
liefern. Aufgrund dessen liefern die Aussagen der Autoren\*innen eine eher technische Sicht-  
weise. Sie beziehen sich auf spezifische Prozesse und liefern wenig allgemeine und manchmal  
keine auf die Analysekriterien bezogenen Aussagen zum Begriff Data Driven. Daher wäre für  
einige der quantitativen Analysekriterien eine größere Anzahl an analysierter Literatur von  
Vorteil, was aufgrund begrenzter zeitlicher Kapazität nicht möglich war. Für die Eruierung  
qualitativer Aussagen war die Anzahl jedoch ausreichend. Als letzte Beschränkung lässt sich  
die zeitliche Beschränkung der Quellen angeben, welche auf Veröffentlichungen ab dem Jahr  
2018 begrenzt wurden. Daher liefert die Arbeit nur Aussagen zu aktuellen datengetriebenen  
Prozessen und macht keine Angaben dazu, wie sich der Begriff zeitlich gewandelt hat. Durch  
die überwiegend technische Sicht wird wenig beleuchtet ob es auch weitere abweichende Be-  
griffsvorstellungen zum Beispiel aus dem Gebiet der Wirtschaftswissenschaften gibt, welches  
in der Literaturanalyse aufgrund der Beschränkungen vernachlässigt wurde. Eine weitere Fra-  
ge die offen bleibt ist, welche Bedeutung datengetriebene Prozesse für heutige Unternehmen  
haben. Da es sich bei der Arbeit um eine Grundlagenforschung handelt, konnte die Evaluie-  
rung lediglich die Anwendbarkeit bestätigen, nicht aber eindeutige Aussagen über die Kor-  
rektheit der einzelnen Begriffsmerkmale liefern. Trotz einiger unbeantworteter Fragen bietet  
die Arbeit einen grundlegenden Überblick aktueller datengetriebener Prozesse im Bereich von  
Produktion und Logistik und kann als Grundlage weiterer Forschung auf diesem Gebiet die-  
nen. Zudem zeigt sie die Notwendigkeit, sich auch mit weiteren Begriffen der Digitalisierung  
nähergehend zu beschäftigen.

Für die weiterführende Forschung kann die Arbeit somit als Vorlage für Begriffsbestim-  
mungen derzeitiger oder auch zukünftiger Begriffe, vor dem Hintergrund einer ständig wach-  
senden und sich verändernden Begriffswelt infolge der fortschreitenden Digitalisierung, ver-  
wendet werden. Als Anknüpfungspunkt für zukünftige Forschung kann der Definitionsvorschlag  
eine Grundlage für die Entwicklung einer allgemeinen Definition des Begriffes Data Driven  
bieten. Bezüglich des praktischen Nutzens ist die Erforschung von möglichen Einsatzmög-  
lichkeiten aktueller datengetriebener Prozesse ein weiteres wissenschaftlich relevantes The-  
ma. Vor allem die Möglichkeiten zur Betreuung von Echtzeitanalysen oder die Entwicklung  
datengetriebener Zwillingsmodelle sind aktuell zwei Themen, welche durch den Einsatz aktu-  
eller datengetriebener Prozesse umgesetzt werden. Die Verwendbarkeit des morphologischen  
Kastens und der Merkmale muss in Unternehmen an Prozessen geprüft werden, um diesen zu  
prüfen, zu verbessern und eventuell weitere Merkmale zu ergänzen.

---

Aufgrund der hohen Aktualität des digitalen Wandels von Unternehmen, der Industrie 4.0 und den Entwicklungen im Bereich intelligenter Fabriken, sind datengetriebene Prozesse im Umfeld von Produktion und Logistik ein Thema, welches zukünftig sehr wahrscheinlich eine größere Bedeutung in der Praxis und auch für die Forschung einnehmen wird. Dies expliziert die Relevanz einer zeitnahen Auseinandersetzung mit dem Begriff Data Driven, um die Potentiale datengetriebener Prozesse aufzudecken und für die Praxis nutzbar zu machen.

## Literaturverzeichnis

- Appelfeller, W.; Feldmann, C.: Die digitale Transformation des Unternehmens. Berlin, Heidelberg: Springer Berlin Heidelberg 2018.
- Arnold, D.; Furmans, K.: Materialfluss in Logistiksystemen. Berlin, Heidelberg: Springer Berlin Heidelberg 2019.
- Bachmann, R.; Kemper, G.; Gerzer, T.: Big Data - Fluch oder Segen?: Unternehmen im Spiegel gesellschaftlichen Wandels. Heidelberg: Hüthig Jehle Rehm GmbH 2014.
- Bernstein, H. (Hg.): Speicherprogrammierbare Steuerung - SPS. Berlin: De Gruyter 2018.
- Bhattacharyya, S. (Hg.): Deep learning: Research and applications. Boston: De Gruyter 2020.
- BITKOM, 2014: Big-Data-Technologien-Wissen für Entscheider, Leitfaden. Hg. v. Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V. Berlin. Online verfügbar unter <https://www.bitkom.org/sites/default/files/file/import/140228-Big-Data-Technologien-Wissen-fuer-Entscheider.pdf>, zuletzt geprüft am 06.07.2021.
- Bundesministerium für Wirtschaft und Energie: Leitbild 2030 für Industrie 4.0. Plattform Industrie 4.0. Online verfügbar unter <https://www.plattform-i40.de/PI40/Navigation/DE/Industrie40/Leitbild2030/leitbild-2030.html>, zuletzt geprüft am 01.07.2021.
- Bundesministerium für Wirtschaft und Energie: Was ist Industrie 4.0. Plattform Industrie 4.0. Online verfügbar unter <https://www.plattform-i40.de/PI40/Navigation/DE/Industrie40/WasIndustrie40/was-ist-industrie-40.html>, zuletzt geprüft am 01.07.2021.
- Cerquitelli, T.; Pagliari, D.J.; Calimera, A.; Bottaccioli, L.; Patti, E.; Acquaviva, A.; Poncino, M.: Manufacturing as a Data-Driven Practice: Methodologies, Technologies, and Tools. Proceedings of the IEEE 109 (2021) 4, S. 399–422.
- Clarivate, 2021: Journal Citation Reports. Hg. v. Clarivate. Online verfügbar unter <https://jcr.clarivate.com/jcr/browse-journals?Init=Yes&SrcApp=IC2LS>, zuletzt geprüft am 07.08.2021.
- Cleve, J.; Lämmel, U.: 2 Grundlagen des Data Mining. In: Cleve, J.; Lämmel, U. (Hrsg.): Data Mining: De Gruyter 2014, S. 37–56.
- Corsten, H.; Gössinger, R.; Spengler, T. (Hrsg.): Handbuch Produktions- und Logistikmanagement in Wertschöpfungsnetzwerken. Berlin: Walter de Gruyter GmbH 2018.
- Craig, I.D.; Ferguson, L.; Finch, A.T.: Journals ranking and impact factors: how the performance of journals is measured. In: The Future of the Academic Journal: Elsevier 2014, S. 259–298.

- DIN Deutsches Institut für Normung e. V. DIN EN 62264-1: Integration von Unternehmensführungs- und Leitsystemen. Berlin: Beuth Verlag GmbH, 2013.
- DIN Deutsches Institut für Normung e. V. DIN IEC 60050-351: Internationales Elektrotechnisches Wörterbuch. Berlin: Beuth Verlag GmbH, 2014.
- DIN Deutsches Institut für Normung e. V. DIN EN ISO 9000: Qualitätsmanagementsysteme. Berlin: Beuth Verlag GmbH, November 2015.
- DIN Deutsches Institut für Normung e. V. DIN 55350: Begriffe zu Qualitätsmanagement und Statistik. Berlin: Beuth Verlag GmbH, September 2020.
- Dornseif, M.; Schumann, K.: Probleme des Datenbegriffs im Rahmen des § 269 StGB. Juristische Rundschau (2002) 2, S. 52–56.
- Duden online, 2021: Da-ten, die. Hg. v. Dudenredaktion. Online verfügbar unter <https://www.duden.de/node/30506/revision/30535>, zuletzt aktualisiert am 08.06.2021.
- Fasel, D.; Meier, A.: Big Data: Grundlagen, Systeme und Nutzungspotenziale. Wiesbaden: Springer Fachmedien Wiesbaden 2016.
- Fayyad, U.: From Data Mining to Knowledge Discovery in Databases. AI Magazine 17 (1996) 3, S. 37–54.
- Frauenhofer IOSB: Industrie 4.0. visIT (2017) 18.
- Früh, W.: Inhaltsanalyse: Theorie und Praxis. Konstanz, München: UVK Verlagsgesellschaft mbH; UVK/Lucius 2017.
- Gadatsch, A.; Landrock, H.: Big Data für Entscheider: Entwicklung und Umsetzung datengetriebener Geschäftsmodelle. Wiesbaden: Springer Fachmedien Wiesbaden 2017.
- Hammond, P.: From computer-assisted to data-driven: Journalism and Big Data. Journalism 18 (2017) 4, S. 408–424.
- Han, T.; Liu, C.; Yang, W.; Jiang, D.: Deep Transfer Network with Joint Distribution Adaptation: A New Intelligent Fault Diagnosis Framework for Industry Application. ISA Transactions 97 (2020) 6, S. 269–281.
- Heinbach, C.; Hagen, S.; Thomas, O.: Freight-Logistics-as-a-Service – Innovative Geschäftsmodelle für ein datengetriebenes Transportmanagement. HMD Praxis der Wirtschaftsinformatik 58 (2021) 3, S. 580–594.
- Hilber, M.: Teil 1. Technische und wirtschaftliche Grundlagen. In: Hilber, M. (Hrsg.): Handbuch Cloud Computing. Köln: Verlag Dr. Otto Schmidt 2014, S. 1–130.
- Hirsch, J.E.: An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America (2005) 102, S. 16569–16572.

- Holland, H.: Big Data. In: Holland, H. (Hrsg.): Digitales Dialogmarketing. Wiesbaden: Springer Fachmedien Wiesbaden 2020, S. 1–23.
- Holland, H.: Digitales Dialogmarketing. Wiesbaden: Springer Fachmedien Wiesbaden 2021.
- ISO/IEC JTC 1 ISO/IEC 2382: Information technology — Vocabulary, Mai 2015.
- Ivanov, D.; Dolgui, A.: A digital supply chain twin for managing the disruption risks and resilience in the era of Industry 4.0. *Production Planning & Control* 32 (2021) 9, S. 775–788.
- Jeong, Y.; Singh, A.; Zafarzadeh, M.; Wiktorsson, M.; Baalsrud Hauge, J.: Data-Driven Manufacturing Simulation: Towards a CPS-Based Approach. In: Säfsten, K.; Elgh, F. (Hrsg.): SPS2020: IOS Press 2020,
- Jiang, Y.; Yin, S.; Kaynak, O.: Data-Driven Monitoring and Safety Control of Industrial Cyber-Physical Systems: Basics and Beyond. *IEEE Access* 6 (2018), S. 47374–47384.
- Kamble, S.S.; Gunasekaran, A.: Big data-driven supply chain performance measurement system: a review and framework for implementation. *International Journal of Production Research* 58 (2020) 1, S. 65–86.
- Kollmann, T.: Handbuch Digitale Wirtschaft. Wiesbaden: Springer Fachmedien Wiesbaden 2020.
- Kollmann, T.; Kuckertz, A.; Stöckmann, C.: Das 1 x 1 des Wissenschaftlichen Arbeitens: Von der Idee bis zur Abgabe. Wiesbaden: Springer Fachmedien Wiesbaden 2016.
- Kuckartz, U.: Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. Weinheim, Basel: Beltz Verlag 2012.
- Kuo, Y.-H.; Kusiak, A.: From data to big data in production research: the past and future trends. *International Journal of Production Research* 57 (2019) 15-16, S. 4828–4853.
- Lassmann, W.; Schwarzer, J.; Rogge, R.; Kaftan, H.-J. (Hrsg.): Wirtschaftsinformatik: Nachschlagewerk für Studium und Praxis ; [Lehrbuch - Bachelor geeignet. Wiesbaden: Gabler 2006.
- Li, X.; Xu, Z.; Wang, M.; Cui, J.; Yang, X.: A Pattern-moving-Based Data-driven Control Method for a Kind of Industrial Production Processes. In: 2020 IEEE 9th Data Driven Control and Learning Systems Conference (DDCLS), Liuzhou, China, 20.11.2020 - 22.11.2020, 2020, S. 590–594.
- Ma, S.; Zhang, Y.; Liu, Y.; Yang, H.; Lv, J.; Ren, S.: Data-driven sustainable intelligent manufacturing based on demand response for energy-intensive industries. *Journal of Cleaner Production* 274 (2020), S. 123155.
- Mamakoukas, G.; Castano, M.; Tan, X.; Murphey, T.: Local Koopman Operators for Data-Driven Control of Robotic Systems. In: Robotics: Science and Systems XV, June 22-26, 2019, 2019,

- Mathar, H.-J.; Scheuring, J.: Unternehmenslogistik: Grundlagen. Bodenheim: Herdt 2012.
- Mayring, P.: Qualitative Inhaltsanalyse: Grundlagen und Techniken. Weinheim, Basel: Beltz Verlag 2015.
- Men, H.: Data-Driven Learning in Enhancing Learners' Language Idiomaticity. *International Journal of Emerging Technologies in Learning (iJET)* 15 (2020) 23, S. 27.
- Mertens, P.; Bodendorf, F.; König, W.; Schumann, M.; Hess, T.; Buxmann, P.: Grundzüge der Wirtschaftsinformatik. Berlin, Heidelberg: Springer Berlin Heidelberg 2017.
- Miragliotta, G.; Sianesi, A.; Convertini, E.; Distante, R. (Hrsg.): Data driven management in Industry 4.0: a method to measure Data Productivity. Amsterdam: Elsevier 2018.
- Mitchell, T.M.: Machine learning. New York: McGraw-Hill 1997.
- Na, Y.; Peng, S.: A Data-driven Approach for Identification and Detection of Intermittent Faults. In: 2021 4th IEEE International Conference on Industrial Cyber-Physical Systems (ICPS), Victoria, BC, Canada, 10.05.2021 - 12.05.2021, 2021, S. 455–460.
- Nguyen, D.T.; Adulyasak, Y.; Cordeau, J.-F.; Ponce, S.I.: Data-driven operations and supply chain management: established research clusters from 2000 to early 2020. *International Journal of Production Research* (2021), S. 1–25.
- North, K.: Wissensorientierte Unternehmensführung: Wissensmanagement im digitalen Wandel. Wiesbaden: Springer Fachmedien Wiesbaden 2021.
- Notté, K.: Wissensmanagement im Vertrieb. Wiesbaden: Springer Fachmedien Wiesbaden 2013.
- Olusola, T. Faboya; Graziola P. Figueredo; Brendan Ryan; Peer-Olaf Siebers (Hrsg.): The Usefulness of Data-driven, Intelligent Agent-Based Modelling for Transport Infrastructure Management. Piscataway, NJ: IEEE 2018.
- Pietsch, W.: Big Data: Cambridge University Press 2021.
- Qiao, Q.; Wang, J.; Ye, L.; Gao, R. (Hrsg.): Digital Twin for Machining Tool Condition Prediction. Amsterdam: Elsevier 2019.
- Roh, P.; Kunz, A.; Netland, T. (Hrsg.): Data-driven detection of moving bottlenecks in multi-variant production lines. Amsterdam: Elsevier 2018.
- Rössler, P.: Inhaltsanalyse. Konstanz, München: UVK Verlagsgesellschaft mbH mit UVK/Lucius; UVK/Lucius 2017.
- Springer, N.; Koschel, F.; Fahr, A.; Pürer, H.: Empirische Methoden der Kommunikationswissenschaft. Konstanz, München: UVK Verlagsgesellschaft mbH; UVK/Lucius 2015.
- Subramaniyan, M.; Skoogh, A.; Salomonsson, H.; Bangalore, P.; Bokrantz, J.: A data-driven algorithm to predict throughput bottlenecks in a production system based on active periods of the machines. *Computers & Industrial Engineering* 125 (2018), S. 533–544.

- Sun, Z., 2018: 10 Bigs: Big Data and Its Ten Big Characteristics, PNG University of Technology. Online verfügbar unter [https://www.researchgate.net/publication/322592851\\_10\\_Bigs\\_Big\\_Data\\_and\\_Its\\_Ten\\_Big\\_Characteristics/link/5ffcf49fa6fdccdc84a3dd8/download](https://www.researchgate.net/publication/322592851_10_Bigs_Big_Data_and_Its_Ten_Big_Characteristics/link/5ffcf49fa6fdccdc84a3dd8/download), zuletzt aktualisiert am 12.01.2021, zuletzt geprüft am 30.06.2021.
- Tao, F.; Cheng, J.; Qi, Q.; Zhang, M.; Zhang, H.; Sui, F.: Digital twin-driven product design, manufacturing and service with big data. *The International Journal of Advanced Manufacturing Technology* 94 (2018a) 9-12, S. 3563–3576.
- Tao, F.; Qi, Q.; Liu, A.; Kusiak, A.: Data-driven smart manufacturing. *Journal of Manufacturing Systems* 48 (2018b), S. 157–169.
- ten\_Hompel, M.; Bauernhansl, T.; Vogel-Heuser, B.: *Handbuch Industrie 4.0*. Berlin, Heidelberg: Springer Berlin Heidelberg 2020.
- ten\_Hompel, M.; Büchter, H.; Franzke, U.: *Identifikationssysteme und Automatisierung*. Berlin, Heidelberg: Springer Berlin Heidelberg 2008.
- ten\_Hompel, M.; Heidenblut, V.: *Taschenlexikon Logistik*. Berlin, Heidelberg: Springer Berlin Heidelberg 2011.
- Ungermann, F.; Kuhnle, A.; Stricker, N.; Lanza, G.: Data Analytics for Manufacturing Systems – A Data-Driven Approach for Process Optimization. *Procedia CIRP* 81 (2019), S. 369–374.
- Voss, R.: *Wissenschaftliches Arbeiten: ... leicht verständlich!* München: UVK Verlagsgesellschaft mbH 2020.
- Wei, W.; Yuan, J.; Liu, A.: Manufacturing data-driven process adaptive design method. *Procedia CIRP* 91 (2020), S. 728–734.
- Weiss, B.M.; Hamel, J.M.; Ganter, M.A.; Storti, D.W.: Data-Driven Additive Manufacturing Constraints for Topology Optimization. *Journal of Manufacturing Science and Engineering* 143 (2021) 2.
- Wen, L.; Li, X.; Gao, L.; Zhang, Y.: A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method. *IEEE Transactions on Industrial Electronics* 65 (2018) 7, S. 5990–5998.
- Xu, K.; Li, Y.; Liu, C.; Liu, X.; Hao, X.; Gao, J.; Maropoulos, P.G.: Advanced Data Collection and Analysis in Data-Driven Manufacturing Process. *Chinese Journal of Mechanical Engineering* 33 (2020) 1.
- Yan, W.; Lin, S.; Kafka, O.L.; Lian, Y.; Yu, C.; Liu, Z.; Yan, J.; Wolff, S.; Wu, H.; Ndip-Abbor, E.; Mozaffar, M.; Ehmann, K.; Cao, J.; Wagner, G.J.; Liu, W.K.: Data-driven multi-scale multi-physics models to derive process–structure–property relationships for additive manufacturing. *Computational Mechanics* 61 (2018) 5, S. 521–541.

- 
- Zhang, C.; Zhou, G.; He, J.; Li, Z.; Cheng, W.: A data- and knowledge-driven framework for digital twin manufacturing cell. *Procedia CIRP* 83 (2019a), S. 345–350.
- Zhang, Y.; Cheng, Y.; Wang, X.V.; Zhong, R.Y.; Zhang, Y.; Tao, F.: Data-driven smart production line and its common factors. *The International Journal of Advanced Manufacturing Technology* 103 (2019b) 1-4, S. 1211–1223.
- Zhong, R. (Hg.): *RFID Data Driven Performance Evaluation in Production Systems*. Amsterdam: Elsevier 2019.



## Anhang

### Anhang 1: Liste des digitalen Anhangs

- Citavi Export
- Tabelle: Gesamtübersicht der Inhaltsanalyse

### Anhang 2: Ergänzende Tabellen

Tabelle A. 1: Ranking aller Quellen<sup>1</sup>

Titel	Name	Typ	JIF	Zitationen	h-Index	Rangwert
Digital twin-driven product design, manufacturing and service with big data	Tao, Fei; Cheng, Ji-angfeng; Qi, Qinglin; et al.	J	3,226	753	61	7,193
A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method	Wen, Long; Li, Xinyu; Gao, Liang; et al.	J	8,236	588	56	6,768
Data-driven smart manufacturing	Tao, Fei; Qi, Qinglin; Liu, Ang; et al.	J	8,633	443	65	5,991
Big data-driven supply chain performance measurement system: a review and framework for implementation	Kamble, Sachin S.; Gunasekaran, Angappa	J	8,568	35	85	3,492
Deep Transfer Network with Joint Distribution Adaptation: A New Intelligent	Han, Te; Liu, Chao; Yang, Wenguang; et al.	J	5,468	126	79	3,485

<sup>1</sup>Die Tabelle listet alle analysierten Quellen aufgeteilt in Journale und Conference Paper. Diese sind sortiert nach ihren berechneten Rangwerten. Die Berechnung der Rangwerte der Journale und der Conference Paper erfolgte unabhängig voneinander. Der Ausschluss aller Journale mit einem JIF unter 8,000, ist in dieser Tabelle unberücksichtigt.

From data to big data in production research: the past and future trends	Kuo, Yong-Hong; Kusiak, Andrew	J	8,568	62	65	3,265
A digital supply chain twin for managing the disruption risks and resilience in the era of Industry 4.0	Ivanov, Dmitry; Dolgui, Alexandre	J	7,044	91	48	2,856
Data-driven operations and supply chain management: established research clusters from 2000 to early 2020	Nguyen, Duy Tan; Adulyasak, Yossiri; Cordeau, Jean-François; et al.	J	8,568	0	56	2,635
Data-Driven Monitoring and Safety Control of Industrial Cyber-Physical Systems: Basics and Beyond	Jiang, Yuchen; Yin, Shen; Kaynak, Okyay	J	3,367	114	55	2,539
Data-driven sustainable intelligent manufacturing based on demand response	Ma, Shuaiyin; Zhang, Yingfeng; Liu, Yang; et al.	J	9,297	23	37	2,526
Manufacturing as a Data-Driven Practice: Methodologies, Technologies, and Tools	Cerquitelli, Tania; Pagliari, Daniele Jahier; Calimera, Andrea; et al.	J	10,96	0	31	2,52
Data-driven multi-scale multi-physics models to derive process–structure–property relationships for additive manufacturing	Yan, Wentao; Lin, Stephen; Kafka, Orion L.; et al.	J	4	99	41	2,249
Data-driven smart production line and its common factors	Zhang, Yongping; Cheng, Ying; Wang, Xi Vincent; et al.	J	3,226	7	61	1,878
A data-driven algorithm to predict throughput bottlenecks in a production system based on active periods of the machines	Subramaniyan, Mukund; Skoogh, Anders; Salomonsson, Hans; et al.	J	5,431	30	16	1,476

Data-Driven Additive Manufacturing Constraints for Topology Optimization	Weiss, Benjamin M.; Hamel, Joshua M.; Ganter, Mark A.; et al.	J	3,033	4	20	0,965
Advanced Data Collection and Analysis in Data-Driven Manufacturing Process	Xu, Ke; Li, Yingguang; Liu, Changqing; et al.	J	1,936	11	24	0,911
Freight-Logistics-as-a-Service – Innovative Geschäftsmodelle für ein datengetriebenes Transportmanagement	Heinbach, Christoph; Hagen, Simon; Thomas, Oliver	J	-	0	12	0,251
Local Koopman Operators for Data-Driven Control of Robotic Systems	Mamakoukas, Giorgos; Castano, Maria; Tan, Xiaobo; et al.	CP	-	25	36	5,297
Digital Twin for Machining Tool Condition Prediction	Qiao, Qianzhe; Wang, Jinjiang; Ye, Lunkuan; et al.	CP	-	21	45	5,211
A data- and knowledge-driven framework for digital twin manufacturing cell	Zhang, Chao; Zhou, Guanghui; He, Jun; et al.	CP	-	13	16	2,614
Data driven management in Industry 4.0: a method to measure Data Productivity	Miragliotta, Giovanni; Sianesi, Andrea; Convertini, Elisa; et al.	CP	-	12	12	2,27
RFID Data Driven Performance Evaluation in Production Systems	Zhong, Ray Y.	CP	-	3	34	2,168
The Usefulness of Data-driven, Intelligent Agent-Based Modelling for Transport Infrastructure Management	Olusola. T. Faboya; Graziela P. Figueredo; Brendan Ryan; et al.	CP	-	4	13	1,221
Data Analytics for Manufacturing Systems – A Data-Driven Approach for Process Optimization	Ungermann, Florian; Kuhnle, Andreas; Stricker, Nicole; et al.	CP	-	2	18	1,205

Data-driven detection of moving bottlenecks in multi-variant production lines	Roh, P.; Kunz, A.; Netland, T.	CP	-	0	15	0,775
A Pattern-moving-Based Data-driven Control Method for a Kind of Industrial Production Processes	Li, Xiangquan; Xu, Zhengguang; Wang, Mushu; et al.	CP	-	0	12	0,62
Data-Driven Manufacturing Simulation: Towards a CPS-Based Approach	Jeong, Yongkuk; Singh, Amita; Zafarzadeh, Masoud; et al.	CP	-	0	10	0,516
A Data-driven Approach for Identification and Detection of Intermittent Faults	Na, Yuhong; Peng, Shichen	CP	-	0	2	0,103

**Tabelle A. 2: Definitionsvorschlag Herleitungstabelle**

Teil des Definitionsvorschlags	Quellen	Aussagen
primär auf Daten fokussierte Betrachtung physischer Prozesse	Cerquitelli et al., 2021, S. 401	"We adopt a datadriven view of the overall system, by putting emphasis on the role of data in the various steps of processing."
	Na und Peng, 2021, S. 455	"[...] the databased SKR method is selected [...] by using input and output data."
	Nguyen et al., 2021, S. 2	"We define data-driven or data-based OSCM broadly as the use of data for OSCM practices [...]"
	Tao et al., 2018a, S. 3567	"[...] the big data-driven product design process mainly puts emphasis on the analysis of physical data [...]"
	Xu et al., 2020, S. 3	"[...] data generated through manufacturing activities is fully utilized [...]"
	Zhang et al., 2019b, S. 1213	"Data-driven is a closed-loop method that not only starts with data but also ends with data."

proaktive Erfüllung konkretisierter Aufgabenstellungen	Kamble und Gunasekaran, 2020, S. 65	"The supply chains adopting the BDA are referred to as big data-driven supply chains"
	Nguyen et al., 2021, S. 2	"We define data-driven or data-based OSCM broadly as the use of data for OSCM practices [...]"
	Tao et al., 2018b, S. 6	"Afterwards, explicit information and actionable recommendations exploited from different kinds of raw data are used to direct the actions [...]"
	Zhang et al., 2019b, S. 1211	"Then, a data-driven method including data self-perception, data understanding, decision-making, and precise control for implementing SPL is proposed."
vorwiegend operative Aufgabenstellungen	Abschnitt 3.1.2 (unter Verwendung)	20 der 23 aus der Literatur ermittelten Aussagen zur Aufgabenorientierung datengetriebener Prozesse, benennen operative Aufgaben. Nur 13 benennen taktische Aufgaben und nur 3 strategische.
direkte Datenerfassung und -übertragung aus physischen Prozessen	Cerquitelli et al., 2021, S. 401	"[...] data coming directly from the production machines (production line), for instance, obtained by built-in sensors and extracted by PLCs [...] data collected through ad hoc sensors added to the machines (IoT Devices)."
	Ivanov und Dolgui, 2021, S. 775	"These models need data on disruptions which happened in the past to construct disruption scenarios, and real-time data on disruptions [...]"
	Roh, P. et al. 2018, S. 159	"[...] using automatically collected data [...]"
	Tao et al., 2018b, S. 6	"During the input-output transformation process, various data is collected from human operators, production equipment, information systems, and industrial networks."
	Wen et al., 2018, S. 5990	"Deep learning (DL) provides an effective way to extract the features of raw data automatically."
	Abschnitt 3.1.2 (unter Daten)	21 der 22 aus der Literatur ermittelten Aussagen zur Datenerfassung datengetriebener Prozesse, benennen die Datenextraktion aus physischen Prozessen.
	Daten treiben Analysen an	Cerquitelli et al., 2021, S. 402
Subramaniyan et al., 2018, S. 533		"To enable such data-driven decision making, companies need informative analytical algorithms to turn high volumes of fast-moving data into meaningful insights."

	Tao et al., 2018b, S. 6	"As inputs, the data from the manufacturing module is transmitted to cloud-based data centers to be further analyzed. Afterwards, explicit information and actionable recommendations exploited from different kinds of raw data are used to direct the actions [...] in the manufacturing module. The real-time monitoring module and problem-processing module are also both powered by the data driver module."
Analysen generieren gezielt auf die Aufgabe bezogenes Wissen	Cerquitelli et al., 2021, S. 399	"[...] along with data-driven methodologies to extract useful and hidden knowledge supporting the decisionmaking process."
	Kamble und Gunasekaran, 2020, S. 65	"[...] to derive meaningful insights, patterns from the collected large data sets [...]"
	Kuo und Kusiak, 2019, S. 4840	"Most customer research is datadriven, as the required knowledge for the research must be captured [...]"
	Subramaniyan et al., 2018, S. 533	"To enable such data-driven decision making, companies need informative analytical algorithms to turn high volumes of fast-moving data into meaningful insights."
	Tao et al., 2018b, S. 4	"Through the above data processing efforts, understandable knowledge can be derived from a large quantity of dynamic and ambiguous raw data."
	Wen et al., 2018, S. 5990	"The knowledge-based methods, which are also named as data-driven methods [...]"
	Xu et al., 2020, S. 3	"[...] raw data transformed into insightful features and knowledge via data processing technologies."
Einsatz von Sensoren, AutoID Systemen und Anbindung zu etablierten IT-Systemen	Cerquitelli et al., 2021, S. 401	"[...] preexisting data (additional data sources) available in the manufacturing company, [...] data coming directly from the production machines (production line), for instance, obtained by built-in sensors and [...] data collected through ad hoc sensors added to the machines (IoT Devices)."
	Ivanov und Dolgui, 2021, S. 780	"Data-driven support can be classified as external and internal information control loops. External support is mainly concerned with disruption data, both anticipated and actual, aggregated through a firm's own historical data and partner risk data, early warning systems, and T&T systems. Internal support encompasses the use of sensors and RFID to update SC data, such as capacity and inventory."

	Jeong et al., 2020, S. 590-591	"Means of data gathering can be IoT which collects data in real-time. Examples are Radio Frequency Identification (RFID), Real-time Locating System (RTLS), smart sensors and actuators. [...] The real-time data gathered from devices and historical data gathered from legacy systems such as Enterprise Resource Planning (ERP), Manufacturing Execution System (MES) [...]"
	Subramaniyan et al., 2018, S. 534	"[...] datadriven algorithms that use digital historical machine data collected by MES [...]"
	Tao et al., 2018b, S. 1-4	"[...] sensors in manufacturing to collect real-time manufacturing data. [...] the manufacturing data comes from equipment, products, human operators, information systems, and networks [...]"
	Xu et al., 2020, S. 2	"[...] data collector, e.g., sensor is designated to capture useful physical values [...]"
	Abschnitt 3.1.2 (unter Daten)	21 der 22 aus der Literatur ermittelten Aussagen zur Datenerfassung datengetriebener Prozesse, benennen die Datenextraktion aus physischen Prozessen. 14 der 22 Aussagen benennen übernommene Daten aus weiteren IT-Systemen.
Möglichkeit große Datenmengen zu verarbeiten	Abschnitt 3.1.2 (unter Daten)	16 der 22 analysierten Quellen machen Aussagen zur Aufnahme und Analyse großer Datenmengen.
Nutzung aktueller und historischer Daten	Cerquitelli et al., 2021, S. 402	"Fig. 1. Reference architecture used in this work."
	Jeong et al., 2020, S. 591	"The real-time data gathered from devices and historical data gathered from legacy systems [...]"
	Subramaniyan et al., 2018, S. 534	"[...] datadriven algorithms that use digital historical machine data [...] data-driven decision making using the real-time MES data."
	Tao et al., 2018b, S. 7	"Fig. 3 The framework of data-driven smart manufacturing"
	Wen et al., 2018, S. 5990-5991	"Deep learning (DL) provides an effective way to extract the features of raw data automatically. [...] The intelligent learning from the large volume of historic data is essential for the data-driven fault diagnosis method."
	Zhang et al., 2019a, S. 346	"Based on the real-time manufacturing data published by data space and historical knowledge in knowledge space [...]"
	Zhang et al., 2019b, S. 1216	"Fig. 5 A data-driven method to realize SPL"

Datenerfassung und -verarbeitung in nahezu Echtzeit	Abschnitt 3.1.2 (unter Daten)	18 der 28 analysierten Quellen (darunter fünf der sechs am höchsten gerankten) machen explizite Angaben dazu, dass datengetriebene Prozesse Daten in nahezu Echtzeit erfassen und geben falls auch verarbeiten.
Aufbau ist abhängig von Einsatzgebiet und Aufgabenstellung	Cerquitelli et al., 2021, S. 400	"[...] a global overview of datadriven methodologies, technologies, and tools [...]"
	Kuo und Kusiak, 2019, S. 4838	"4.1.1.4 Miscellaneous data-driven approaches. [...]"
Analysen verwenden meist statistische Methoden und Methoden des maschinellen Lernens	Cerquitelli et al., 2021, S. 409	"[...] data-driven, i.e., based on statistical or ML algorithms."
	Ivanov und Dolgui, 2021, S. 781	"Figure 2. Digital supply chain twin for managing disruption risks."
	Kamble und Gunasekaran, 2020, S. 65	"The supply chains adopting the BDA are referred to as big data-driven supply chains."
	Kuo und Kusiak, 2019, S. 4835	"Data-driven approaches allow discovery of non-trivial patterns from data. Data mining algorithms are the primary model and knowledge discovery tools."
	Wen et al., 2018, S. 5991	"Further, the statistical and machine learning methods are often utilized jointly."
	Zhong, R. 2019, S. 24	"Big data analytic tools such as statistics approaches, data mining models and machine learning algorithms, provide a good opportunity for processing the data."
	Abschnitt 3.1.2 (unter Methoden und Technologien)	Ein großer Teil der analysierten Quellen benennt die Nutzung statistischer Methoden (Data Analytics: 7 Nennungen, Data Mining: 7 Nennungen) und Methoden des maschinellen Lernens, künstlicher Intelligenzen oder deep learning (15 Nennungen)
autonome Arbeitsweise	Roh, P. et al. 2018, S. 160	"The human agency impact of the proposed method is low, as the whole processes of data collection and analysis can be fully automated."
	Tao et al., 2018b, S. 6	"The data-driven smart manufacturing shares the following five characteristics (see Fig. 4): [...] customer-centric product development [...] self-organization [...] self-execution [...] self-regulation [...] self-learning and self-adaption [...]"
	Wen et al., 2018, S. 5990	"Deep learning (DL) provides an effective way to extract the features of raw data automatically."
	Zhang et al., 2019a, S. 346	"[...] that could support the autonomous operations of manufacturing process, namely operating separate from human control."
Untersuchung dynamischer komplexer Systeme	Han et al., 2020, S. 106	"[...] data-driven methods get rid of the complexity of creating physical or statistical models for different engineering systems."



	Kuo und Kusiak, 2019, S. 4835	"The structure and format of data-sets became more complex and heterogeneous. Traditional statistical approaches faced the challenge of high dimensionality and interactions among the variables. [...] Data-driven approaches allow discovery of non-trivial patterns from data."
	Li et al., 2020, S. 590	"Due to the existence of a kind of industrial production process which is too complex to determine the model structure and identify the system parameters, a data-driven control method is proposed based on pattern moving."
	Mamakoukas et al., 2019, S. 1	"[...] the authors exploit the Koopman operator to develop a systematic, data-driven approach for constructing a linear representation in terms of higher order derivatives of the underlying nonlinear dynamics."
	Wen et al., 2018, S. 5990	"They are very suitable for the complex systems which are difficult to establish the explicit models or signal symptoms."
stellen keine neuen Technologien zur Verfügung	Cerquitelli et al., 2021, S. 400	"[...] a global overview of datadriven methodologies, technologies, and tools [...]"
	Kamble und Gunasekaran, 2020, S. 65	"Big data analytics (BDA) defined as collection of data, analytical tools, computer algorithms and techniques to derive meaningful insights [...] The supply chains adopting the BDA are referred to as big data-driven supply chains."
	Kuo und Kusiak, 2019, S. 4838	"4.1.1.4 Miscellaneous data-driven approaches. [...]"
vorwiegend diagnostische oder prädiktive Analysen	Abschnitt 3.1.2 (unter Verwendung)	26 der 26 aus der Literatur ermittelten Aussagen zu den Analysensystemen datengetriebener Prozesse, betreffen diagnostische Analysen. 16 Quellen nennen prädiktive Analysen. 12 nennen präskriptive- und nur 3 nennen deskriptive Analysen.
Rückkopplung	Abschnitt 3.1.2 (unter Verwendung)	10 der 28 analysierten Quellen geben an, dass die Analyseergebnisse datengetriebener Prozesse, bzw. gewonnene Erkenntnisse oder auch abgeleitete Handlungen zum physischen System rückgekoppelt werden können.
direkte Fehlerbehebung oder Prozessregelung	Jeong et al., 2020, S. 593	"In the second phase, results from data analysis are fed into physical systems in the real world. The results are used as input to machines and actuators in real-time resulting in a self-improvement process for the complete cyber-physical simulation model."
	Li et al., 2020, S. 590	"a data-driven control method is proposed based on pattern moving."
	Tao et al., 2018b, S. 7	"Fig. 3. The framework of data-driven smart manufacturing."