

1 Einleitung

Infolge der Globalisierung ist die Logistik zu einem der größten Wirtschaftssegmente in Deutschland herangewachsen (vgl. Steglich et al. 2016; Wannenwetsch 2010). Zu den Aufgaben der Logistik zählen nach Jünemann (1989, S. 18), „die richtige Menge, der richtigen Objekte als Gegenstände der Logistik (Güter, Personen, Energie, Informationen), am richtigen Ort (Quelle, Senke) im System, zum richtigen Zeitpunkt, in der richtigen Qualität, zu den richtigen Kosten zur Verfügung zu stellen“. Ein wesentlicher Aspekt in der Logistik ist die Wertschöpfung mittels globaler Netzwerke, an denen unterschiedliche Teilnehmer partizipieren. Diese Netzwerke sind unter dem Begriff Supply Chain (SC, Plural: SCs) zusammengefasst und ein elementarer Bestandteil der heutigen Prozesslandschaft. Zur SC gehören alle Unternehmen bzw. eigenständig handelnden Unternehmensteile, die zur Entwicklung, Erstellung und Lieferung von Produkten oder Dienstleistungen beitragen (Beckmann 2012). In den letzten Jahren führte die permanente Zunahme der globalen Vernetzung zu einem Komplexitätsanstieg in der SC (Pfeiffer et al. 2013). Ein Ende dieser Entwicklung ist nicht abzusehen. Aufgrund des Komplexitätsanstieges entstehen beispielsweise höhere Kosten oder zeitliche Verzögerungen in der logistischen Kette, wodurch Unsicherheiten in der SC zunehmen. Dadurch ist die Notwendigkeit gegeben, die Unsicherheiten zu reduzieren, um die zunehmende Komplexität in der SC zu beherrschen (Wilding 1998). Um die Komplexität zu kontrollieren, ist konkretes Wissen über die SC notwendig. Die Fähigkeit, dieses Wissen zu nutzen sowie in geeigneter Weise zu integrieren, ersetzt mittlerweile bisherige Optimierungsziele für Unternehmen (Kemppainen und Vepsäläinen 2003) und stellt eine wichtige Unternehmensressource dar (Reiber 2013; Wenzel, Abel et al. 2011).

Eine spezielle Form von Wissen, die sowohl für die SCs als auch für das Supply Chain Management (SCM) von größter Wichtigkeit ist, sind Wirkzusammenhänge (Harland 1996). Wirkzusammenhänge sind immer dann von Relevanz, wenn es um die Abbildung von Systemverhalten geht. Das Systemverhalten, das zur SC-Kontrolle in der Intralogistik dient, wird z. B. mittels Simulation (Rabe und Scheidler 2014) oder logistischer Assistenzsysteme modelliert (Kuhn et al. 2008). Fachlich können diese Wirkzusammenhänge beispielsweise beschreiben, unter welchen Konstellationen Abweichungen von der durchschnittlichen Lieferzeit besonders häufig auftreten. Die Lieferung von Waren von und zu unterschiedlichen Standorten innerhalb eines logistischen Systems wird allgemein als Transaktion bezeichnet (Corsten und Gössinger 2008). Sowohl innerhalb einzelner als auch zwischen unterschiedlichen Transaktionen treten jeweils verschiedene Wirkzusammenhänge auf. Ein elementarer Schritt zur Beherrschbarkeit von SCs ist die Identifikation von Wirkzusammenhängen der logistischen Transaktionen und deren Modellierung. Dadurch besteht die Möglichkeit, Kenntnisse über das Systemverhalten bereitzustellen.

In der SC werden traditionell Materialflüsse, Finanzflüsse und Informationsflüsse unterschieden. Die Betrachtung der Flüsse erfolgt sowohl in Richtung des Lieferanten als auch des Kunden. Die Flüsse sind durch verschiedenartige Prozesse gekennzeichnet. Dementsprechend werden Materialflussprozesse in Transport-, Lager-, Umschlag- und Sortierprozesse unterteilt (Arnold und Furmans 2009). Grundlage dieser Prozesse ist eine Datenbasis, mit der die Prozessausführung ermöglicht wird. Während dieser Ausführung generiert der Prozess Daten. Ein Großteil der erzeugten Daten basiert auf Transaktionen, die Zustandsänderungen beschreiben, die mittels Mengenangaben und Zeitstempeln erfasst werden. Die gestiegene Komplexität der Netzwerke in Verbindung mit der Verbesserung der Informations- und Kommunikationstechnologien führt zu einem Anstieg der zu speichernden Datensätze und deren Detaillierungsgrad (Harrison und van Hoek 2008). Aufgrund der unüberschaubaren Datenmenge kann die Entdeckung von komplexen Wirkzusammenhängen nicht mehr manuell erfolgen.

Eine Herausforderung ist die Wissensentdeckung im Bereich der SC, da bereits bestehende Lösungen aus dem Feld der Wissensentdeckung nicht einfach übertragbar sind (Wyatt et al. 2014). Ein allgemeiner Ansatz, um Wissen wie Wirkzusammenhänge zu identifizieren, ist das Knowledge Discovery in Databases (KDD). Im KDD existieren etablierte Vorgehensmodelle zur Wissensentdeckung wie beispielsweise das Cross Industry Standard Process for Data Mining (Crisp-DM) (Gabriel et al. 2009) oder das Modell von Fayyad (Fayyad et al. 1996b). Der zentrale Aspekt des KDD ist die Wissensentdeckung mittels Data Mining. Hierbei bezeichnet Data Mining eine Sammlung von Verfahren zur Extraktion von Wissen. Klassische Problemstellungen im Bereich der Logistik nutzen bereits erfolgreich diese Art der Wissensentdeckung. Hierzu gehören Kundensegmentierung zu Marketingzwecken, Prognosen im Bereich des Controllings und Unterstützung der Produktion im Bereich der Materialbedarfsplanung (Gabriel et al. 2009). Spezifische Data-Mining-Methoden konnten ebenfalls für die SC adaptiert werden. Exemplarisch sei der Beitrag von Kamble et al. (2015) erwähnt, der Vorhersageverfahren des Data Minings für die SC genutzt hat, um Aspekte wie den Bullwhip-Effekt zu berücksichtigen. Allerdings zeigt die jüngste Studie des Fraunhofer IPA (Weskamp et al. 2014), dass hauptsächlich Unternehmen aus Branchen mit starkem Endkundenbezug KDD-Techniken nutzen. In dieser Studie empfehlen zudem 34% der befragten Unternehmen die Logistik-Branche als großen Potentialträger für das KDD. KDD hat zum Ziel, mittels eines geeigneten zyklischen Vorgehensmodells Muster, wie beispielsweise Wirkzusammenhänge, in Daten zu entdecken. Im Vorgehensmodell des KDD gibt es unterschiedliche Musterbegriffe, die aus den verschiedenen Teilschritten resultieren. Diese Muster können auf unterschiedliche Art dargestellt werden (Klößen und Zytkow 1996). Verifikation und Validierung (V&V) sind aufgrund der induktiven Prozess-Vorgehensweise keine Eigenschaft des KDD (Düsing 2010). Aus diesem Grund muss das KDD-Vorgehensmodell um die V&V erweitert werden. Der wachsende Datenbestand erweist sich im Zusammenhang mit dem KDD sowie der V&V als Vorteil, denn sowohl für die Methoden der Wissensent-

deckung als auch die Validierung des Wissens ist eine ausreichende Datenbasis notwendig.

Das Ziel der Arbeit ist die Entwicklung einer Methode zur Wissensentdeckung in SC-Datenbeständen, die als Element ein domänenspezifisches KDD-Vorgehensmodell beinhaltet. Für die Modellentwicklung müssen als Teilziele die bereits bestehenden Vorgehensmodelle wie Crisp-DM oder das Fayyad-Modell bezüglich der spezifischen Anforderungen und Ziele der SC analysiert und gegebenenfalls entsprechend angepasst werden. Insbesondere ist zu beachten, dass eine bedarfsgerechte Vorverarbeitung der Daten und eine Anpassung der Algorithmen zur Wissensentdeckung zu berücksichtigen sind. Ein weiteres Teilziel besteht in der Aufschlüsselung der Bedeutung des dispositiven Kontextwissens im SC-Umfeld sowie dessen anschließender Integration in die Problemkodierung der Algorithmen. Hierfür werden spezifische Charakteristika der SC wie u. a. die Berücksichtigung unterschiedlicher Transportmittel in der Vorverarbeitung entwickelt. Ein zusätzliches Ziel liegt in der Wissensrepräsentation durch Muster, wobei der Musterbegriff für die spezifischen Bedürfnisse der SC angepasst und erweitert werden muss. Dies bedeutet insbesondere, dass das Kontextwissen in den Musterbegriff des KDD-Vorgehensmodells integriert wird. Ein weiteres Ziel ist die Ermöglichung der Verwendung der entwickelten Methode bei einer unzureichenden Datengrundlage. Zu diesem Zweck wird eine Datengenerierung durch Simulation in der Methode eingeführt. Ferner ist ein weiteres Ziel dieser Arbeit die Integration einer begleitenden V&V in das zu entwickelnde Vorgehensmodell der Methode. Hierbei wird die V&V durch einen neuartigen Einsatz der Simulation unterstützt. Das Erreichen der aufgeführten Forschungsziele ist im Hinblick auf die anstehende digitale Transformation und zunehmende Automatisierung in den Unternehmen erforderlich, da dadurch eine Grundlage zur Beherrschung der Komplexität im SC-Umfeld geschaffen wird.

Zur Erreichung der zuvor genannten Ziele werden in dieser Arbeit Datenbestände von verschiedenen SCs untersucht und mögliche logistische Fragestellungen zu diesen Datenbeständen identifiziert. Aus der Datengrundlage in Verbindung mit möglichen SC-Fragestellungen werden Anforderungen abgeleitet. Basierend auf diesen Anforderungen wird ein KDD-Vorgehensmodell aus dem Bestand der existierenden Vorgehensmodelle ausgewählt. Auf Grundlage dieses Vorgehensmodells wird im Anschluss das domänenspezifische KDD-Vorgehensmodell für die SC entwickelt. Entsprechend muss hierfür insbesondere das Verhältnis von bedarfsgerechter Vorverarbeitung zu anderen Prozessphasen bestimmt werden. Die bedarfsgerechte Vorverarbeitung im SC-Kontext umfasst beispielsweise spezifische Transformationen von Daten, da Transaktionsdaten sowohl numerische Daten (z. B. Zeitstempel) als auch eine Vielzahl von nominalen Daten (z. B. Standorte) enthalten. Dies bedeutet eine zusätzliche, zeitintensive Vorverarbeitung der Daten für die Algorithmen. Aus der fachlichen Perspektive ist insbesondere der Schritt der Partitionierung relevant, der oftmals bei großen Datenbeständen für die effiziente Suche notwendig ist. Die Partitionierung ermöglicht die effiziente Suche über grö-

ßere Zeiträume, indem diese den Suchraum nach geeigneten Kriterien aufteilt. Je nach Transaktionsfrequenz und Transaktionsgranularität des Unternehmens erfordert die Größe des Datenbestandes diesen Vorgang. Im Rahmen der Vorgehensmodellerstellung sind geeignete Musterbegriffe zu definieren. Diese sind an entsprechender Stelle in das Vorgehensmodell zu integrieren. Hierbei ist insbesondere die Wechselwirkung zwischen Mustern und Kontextwissen zu beachten und für das Vorgehensmodell zu adaptieren. Unter diesem Aspekt ist mittels V&V zu prüfen, ob die entdeckten Muster relevante Zusammenhänge beschreiben und diese im realen System zu beobachten sind. Für die modellbegleitende V&V wird in der vorliegenden Arbeit ein neues Konzept in das zu entwickelnde KDD-Vorgehensmodell integriert. Zudem werden die bestehenden V&V-Techniken um die ereignisdiskrete Simulation (Discrete Event Simulation, DES) erweitert, die im Kontext der SC bereits vielfältig Anwendung findet. DES kann neben der Simulation der kompletten SC auch die konkreten Auswirkungen einzelner Zusammenhänge innerhalb des zu betrachtenden Systems isoliert untersuchen sowie die Analyse von Abhängigkeiten unterstützen. Die DES ist im Rahmen der KDD-Vorgehensmodelle ein neuer Ansatz, um inhärente Verfahrensschwächen im SC-Kontext auszugleichen. Da insbesondere in der Planungsphase von SCs oftmals keine ausreichende Datenbasis vorhanden ist, muss zusätzlich eine Möglichkeit zu einer künstlichen Generierung von Daten entwickelt werden. Dieser Vorgang ist im Kontext des DES als Data Farming bekannt und wird hier für den Einsatz in Vorgehensmodellen adaptiert.

Um die praktische Anwendbarkeit der in dieser Arbeit entwickelten Methode zu demonstrieren, wird diese in zwei Anwendungsfeldern angewandt. Das erste Anwendungsfeld zeigt die Ausführung des entwickelten Vorgehensmodells auf Transaktionsdaten in einem konkreten Praxisfall. Hier dient das Vorgehensmodell sowie die gewonnenen Erkenntnisse bei der Ausführung des Modells als Basis für die Konzeptionierung eines unternehmensweiten Datencockpits. Das zweite Anwendungsfeld stellt ein Data-Farming-Modell vor, mit dem die Datengenerierung innerhalb des entwickelten Methode demonstriert wird. Zusätzlich wird exemplarisch dargestellt, wie die Wirkzusammenhänge der generierten Daten mittels der in dieser Arbeit entwickelten simulationsunterstützten Validierung geprüft werden können.

2 Wissensgewinn im Kontext der Supply Chain

In diesem Kapitel werden die Grundlagen für die in dieser Arbeit entwickelten Methode vorgestellt. Zentrale Themen sind der Wissensbegriff, die SC und ihre Datenlage sowie Verfahren, die im Kontext der Wissensgewinnung Anwendung finden. Aus den hier diskutierten Problemstellungen ergibt sich sowohl eine thematische Eingrenzung als auch der Handlungsbedarf, der am Ende des Kapitels zu Forschungsfragen zusammengefasst wird. Dafür werden zunächst die zentralen Begriffe, die sich aus dem Thema dieser Arbeit ergeben, erläutert.

Allgemein beschreibt eine Methode ein planmäßiges Vorgehen mit überprüfbaren Ergebnissen und bildet somit den Ausgangspunkt ingenieurmäßigen Vorgehens. Der Begriff Methode als „an approach to perform a systems development consisting of directions and rules, structured in a systematic way in development activities with corresponding development products“ (Brinkkemper 1996, S. 275) stammt aus dem Method Engineering. Methoden lassen sich nach Prozeduren, Notationen und deren Mischform, Konzepte, unterteilen (Goldkuhl et al. 1998), wobei diesen wiederum einzelne Elemente zugeordnet werden können. Das zentrale Methodenelement ist das Vorgehensmodell (Braun et al. 2004), dessen Entwicklung ein Ziel dieser Arbeit ist. Ein Vorgehensmodell beschreibt nach Sharafi (2013) die einzelnen Phasen eines Prozesses und strukturiert die damit verbundenen Maßnahmen. Das Vorgehensmodell, in Funktion eines Rahmenwerkes, muss deutlich vom Prozessmodell, in Funktion eines Regelwerkes, unterschieden werden. In der Literatur findet sich oftmals eine synonyme Verwendung beider Begriffe. In dieser Arbeit wird jedoch der korrekte Begriff Vorgehensmodell genutzt und nur bei direkten Zitaten oder Eigennamen der teilweise unsaubere Begriff der Literatur übernommen. Bei der praktischen Evaluierung des Vorgehensmodells wird die Durchführung einzelner Phasen im Modell diskutiert. In diesem Kontext wird der Begriff Prozess genutzt, wenn die sequentielle Ausführung von einzelnen Phasen verdeutlicht werden soll.

Der Term „Erschließen von Wissen“ beschreibt das Ziel der Methode, die Extraktion und anschließende Interpretation von Informationen, die unter gewissen Gegebenheiten, hauptsächlich Vernetzung (Dengel 2012), Wissen darstellen. Der Begriff Wissen wird in dieser Arbeit im ingenieurwissenschaftlichen Kontext untersucht. Grundlegende Betrachtungen wie die Erkenntnistheorie aus der Philosophie, die Frage, unter welchen Umständen der Begriff Wissen angewandt wird oder die Erforschung, aus welchen Grundlagen sich der Wissensbegriff historisch zusammensetzt, werden nicht diskutiert. Grundlage für die Wissensextraktion sind Datenmuster, eine Informationsrepräsentation, die aus der Verfahrenswelt des KDD, einem Vorgehensmodell zur Wissensentdeckung, stammt. Somit stellen die Muster den Betrachtungsgegenstand der zu entwickelnden Wissensentdeckung dar. Dar-

über hinaus stellt der Datenmusterspekt in Verbindung mit dem Term „Erschließung von Wissen“ den Bezug zum KDD bezüglich des zu entwickelten Vorgehensmodells her. Viele der im KDD eingesetzten Verfahren zur Wissensentdeckung stammen ursprünglich aus der Mathematik oder der Statistik oder sind aus anderen Informatik-Disziplinen bekannt (Fockel 2009). Folglich kann konstatiert werden, dass der Begriff Wissensentdeckung in unterschiedlichen Disziplinen angesiedelt ist und praktische Lösungen in der Industrie oftmals einen interdisziplinären Ansatz erzwingen. Diese Arbeit startet in der Begriffswelt des KDD; eine darüberhinausgehende Betrachtung anderer Disziplinen sowie verwandter Gebiete der Informatik wird nur bei komplexen Begriffsklärungen berücksichtigt. Der letzte Teil des Titels „Supply-Chain-Datenbanken“ beschreibt den Kontext der zu entwickelnden Methode, deren Gültigkeit für den Einsatz in SCs gezeigt wird, jedoch nicht notwendigerweise auf diese beschränkt ist.

2.1 Wissen

Im Folgenden wird eine Einführung der Begriffe Wissen und Wissensmanagement gegeben. Im Anschluss werden verschiedene Kategorisierungsmöglichkeiten des Wissensbegriffs erörtert und aufgezeigt, welche Kategorisierung für den Kontext der SC Anwendung finden kann.

2.1.1 Definition von Wissen

In der Literatur finden sich unterschiedliche Definitionsansätze des Begriffs Wissen. So gibt es im Allgemeinen ein Verständnis des Begriffs Wissen, doch unterschiedliche Fachdisziplinen verwenden eine Vielzahl eigener Definitionen. Vorherrschend ist eine Abgrenzung der Begriffe Daten, Informationen und Wissen (Probst et al. 2006). Einige Autoren erwähnen noch zusätzlich den Begriff der Zeichen (Krcmar 2011). Diese Grundbegriffe Daten, Informationen und Wissen finden sich in der Wissenstreppe von North (2011) wieder. Abbildung 2.1 zeigt den Zusammenhang der Grundbegriffe nach diesem Abbildungsmodell. Hierbei stellen die Zeichen die unterste Ebene dar, die durch Ordnungsregeln (Syntax) in Daten überführt werden. North gibt an, dass die Anreicherung von Daten mit Bedeutung zu Information führt und verweist auf die hierarchische Beziehung zwischen den Begriffen, die von Krcmar (2011) dargelegt wurde.

Definition 2.1 Information: Zeichen werden durch Syntaxregeln zu Daten, welche interpretierbar sind und damit für den Empfänger Information darstellen (vgl. North 2011; Probst et al. 2006).

In der angeführten Definition wurde der Begriff Kontext vermieden, obwohl Krcmar (2011) und Probst et al. (2006) von Kontextbezug sprechen. Die Begründung hier-

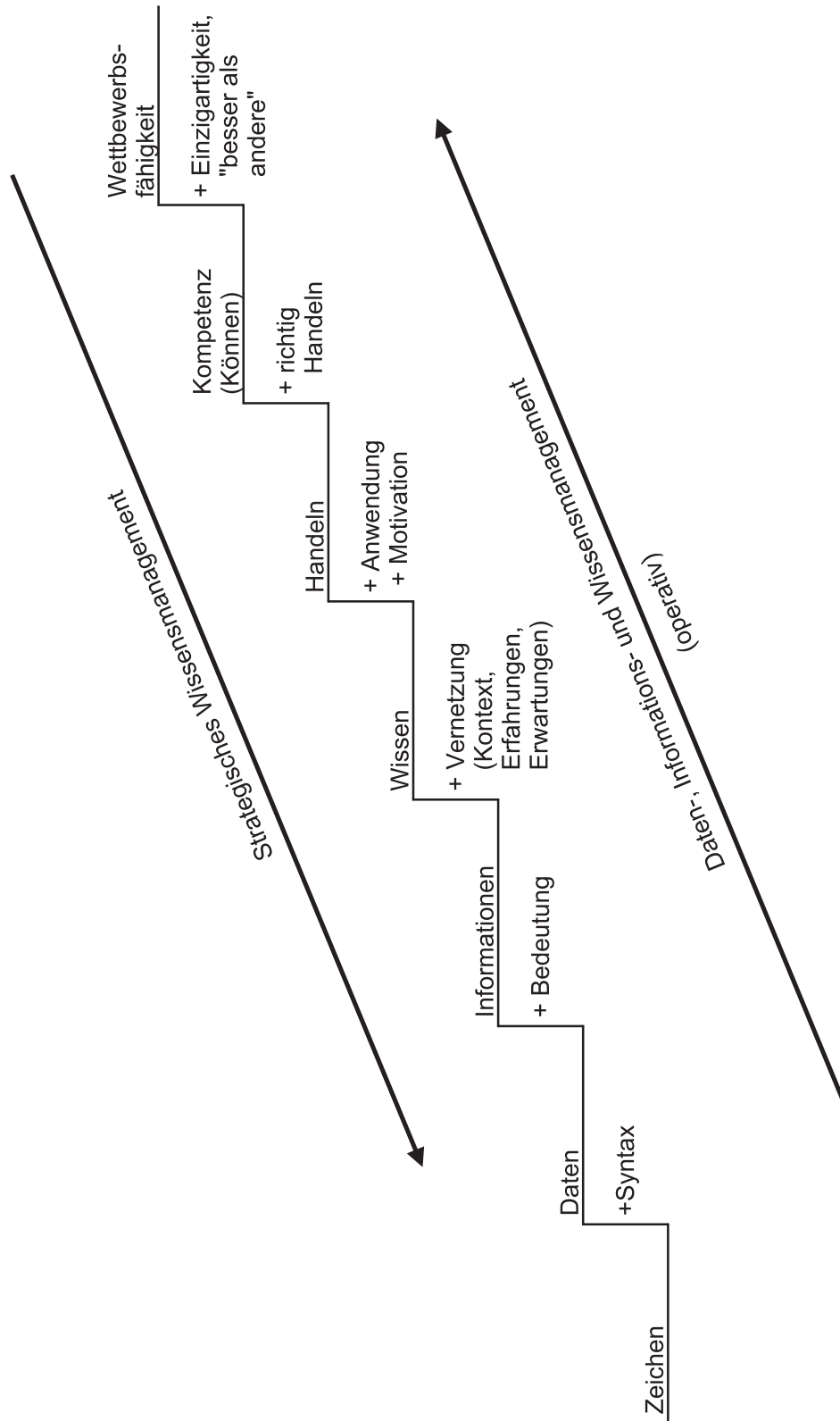


Abbildung 2.1: Wissenstreppe nach North (2011, S. 37)



für ist, dass der Kontextbezug in der Wissenstreppe erst bei der Überführung der Information von zentraler Bedeutung ist. Nach North (2011) entsteht Wissen, wenn Information mit weiteren Informationen, Erfahrungen oder Kontext vernetzt wird. Das wird auch in dem Zitat des Computerpioniers Feigenbaum deutlich: „Knowledge is not the same as information. Knowledge is information that has been pared, shaped, interpreted, selected, and transformed“ (Feigenbaum und McCorduck 1983, S. 121). Demgegenüber stehen spezifische Wissensdefinitionen, wie beispielsweise die von Sowa (2000), im Kontext der künstlichen Intelligenz. Hier wird Wissen als die Fähigkeit beschrieben, ein mentales Modell zu erstellen, welches die Teilaspekte der Wirklichkeit in geeigneter Weise repräsentiert. Der Begriff Wissen muss immer im Problemkontext betrachtet werden, da verschiedene Fachdisziplinen den Begriff für ihre Bedürfnisse adaptiert haben. Zusätzlich ist die Suche nach beliebigen Strukturen und somit unspezifischem Wissen basierend auf den heutigen Methoden unmöglich (Küppers 1999).

Die vorliegende Arbeit folgt in der Begriffsdefinition von „Wissen“ der weit verbreiteten Definition nach Probst, der die zentralen Begriffe Daten und Informationen aufnimmt. Er ergänzt diese Begriffe um den Begriff des „Individuums“, um die menschliche Interaktion zu verdeutlichen. Die Ergänzung um die Interaktionsmöglichkeit wird als wesentlich erachtet, da die Wissensentdeckung in Funktion eines Prozesses immer manuelle Komponenten wie die menschliche Interpretation beinhaltet. Diese Definition bildet auch eine der Grundlagen für die Wissenstreppe, die in Abbildung 2.1 dargestellt ist. Das aufgeführte Originalzitat in Definition 2.2 wurde zum späteren Zeitpunkt auch in die deutsche Spezifikation des Wissensmanagements aufgenommen (DIN SPEC 91281:2012-04), hier wird jedoch die Originalquelle zitiert.

Definition 2.2 Wissen: „Wissen bezeichnet die Gesamtheit der Kenntnisse und Fähigkeiten, die Individuen zur Lösung von Problemen einsetzen. Dies umfasst sowohl theoretische Erkenntnisse als auch praktische Alltagsregeln und Handlungsanweisungen. Wissen stützt sich auf Daten und Informationen, ist im Gegensatz zu diesen jedoch immer an Personen gebunden. Es wird von Individuen konstruiert und repräsentiert deren Erwartungen über Ursache-Wirkungs-Zusammenhänge“ (Probst et al. 2006, S. 44).

Die nächsten drei Treppenstufen der Wissenstreppe in Abbildung 2.1 verdeutlichen, wie ein Unternehmen das Wissen verwerten kann, um Wettbewerbsfähigkeit zu erlangen. Dieser Treppenbereich ist dem Wissensmanagement zugeordnet. Unter Wissensmanagement ist die Nutzung des Wissens als Ressource der organisationalen Wissensbasis sowie die systematische Planung, Steuerung und Organisation der Wissensbedarfe der Organisation zu verstehen (Schaffranietz und Neumann 2009). Das Verständnis dieser Arbeit richtet sich nach der verbreiteten Spezifikation von Heisig:

Definition 2.3 Wissensmanagement: „Wissensmanagement umfasst Verfahren, Methoden, Instrumente und Werkzeuge, die einen systematischen, methodengestützten Umgang mit Wissen in allen Bereichen und auf allen Ebenen der Organisation realisieren, um die organisatorische Leistungsfähigkeit der Geschäftsprozesse zu verbessern und zur Erreichung der Organisationsziele beizutragen“ (Heisig 2005, S. 18).

Die Notwendigkeit des Einsatzes von Wissensmanagementsystemen als einer Methode der Standardisierung und Systematisierung von Wissen begründet sich nach Wenzel, Abel et al. (2011) in der Komplexität der Wissensarbeit. Es gibt unterschiedliche Wissensmanagement-Modelle, die sich in Beschreibung der Aufgaben und Aktivitäten deutlich voneinander unterscheiden. Diese Modelle sollen jedoch im Kontext der vorliegenden Arbeit nicht diskutiert werden, da der Fokus auf der Wissensentdeckung liegt. In der Folge ist das Grundverständnis von der Existenz des Wissensmanagements ausreichend. Die Wissensentdeckung umfasst im Wesentlichen die Treppenstufen der Daten, Information und des Wissens in Abbildung 2.1, wobei die Unterscheidung zwischen Daten und Information im Bezug auf die Wissensentdeckung nicht immer eindeutig ist. Dies ist darin begründet, dass die Ebenen der Daten und der Informationen oftmals im Umfeld des KDDs nicht disjunkt sind, da die Zuordnung vom Betrachter abhängig ist. Folglich wird der Begriff Information nur dann verwendet, wenn eine Interpretierbarkeit der Daten durch zusätzliche Randbedingungen angezeigt ist. Dies ist oftmals in späteren Phasen des KDDs der Fall. Auf Ebene der Datenbank und den Rohdaten findet der Begriff Daten Anwendung, da hier häufig die technischen Gegebenheiten und nicht die inhaltlichen Bedeutungen der Daten im Vordergrund stehen.

2.1.2 Systematisierung des Wissens

In der zuvor aufgeführten Definition des Wissensmanagements wird Wissen als Ressource bezeichnet. Diese Ressource kann aus Unternehmenssicht in verschiedene Kategorien unterteilt werden. Die gebräuchlichsten Kategorien sind Explizierungsgrad (implizit oder explizit), Anwendbarkeit (kontextbezogen oder dekontextualisiert) und Zugänglichkeit (individuell oder kollektiv) (Bullinger et al. 2009). Probst et al. (2006) schlagen abweichende Kategorisierungen vor, die auf die Ebenen Daten, Information und Wissen projiziert werden. Hierbei werden beispielsweise Daten als kontextunabhängig, Wissen jedoch als kontextabhängig gesehen. Da in der Kategorisierung von Probst der Explizierungsgrad keine Berücksichtigung findet, der insbesondere für die Darstellung von Wissen von großer Bedeutung ist, dient als Grundlage die Kategorisierung von Bullinger. Abbildung 2.2 gibt eine Übersicht der wesentlichen Kategorien.

Der Begriff explizites Wissen drückt aus, dass etwas eindeutig kodierbar und somit eindeutig kommunizierbar ist. Den Gegensatz hierzu bildet das implizite (tazite) Wissen, das nicht einfach fassbar ist; beispielsweise, weil es erfahrungsbezogen oder

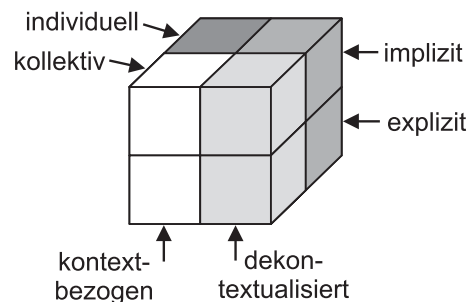


Abbildung 2.2: Ressource Wissen nach Bullinger et al. (2009, S. 703)

nicht formalisierbar ist. Kontextbezogenes Wissen ist spezifisch angewandtes Wissen wie Aufgaben-, Prozess- oder Organisationswissen. Eine Übertragbarkeit auf andere Objekte oder Situationen ist im Allgemeinen nicht möglich. Im Gegensatz dazu beschreibt dekontextualisiertes Wissen grundlegende Vorgehensweisen und Handlungen, die auf unterschiedliche Fragestellungen angewandt werden können. Individuelles Wissen ist nur einzelnen Personen zugänglich, kollektives Wissen immer mehreren Personen. Für eine über diese Definitionen hinausgehende Betrachtung der Kategorien und deren Kombinationsmöglichkeiten sei auf Heisig (2005) verwiesen.

2.2 Supply Chains

Im folgenden Abschnitt werden die SC, das SCM und die aktuelle Datenlage in der SC diskutiert, um diese mit dem zuvor eingeführten Wissensbegriff zu verknüpfen. Neben dem grundlegenden Verständnis von SC und SCM liegt der Fokus auf den Datenbeständen in der SC, welche die Grundlage der Wissensentdeckung bilden. Um in den nachfolgenden Kapiteln ein gemeinsames Verständnis der Sachverhalte zu gewährleisten, ist die informationstechnische Begriffserklärung das wesentliche Ziel dieser Diskussion.

2.2.1 Grundlagen der Supply Chain

SCs sind ein zentraler Bestandteil der Produktionslandschaft (Stevens 1989). Zur SC gehören alle Unternehmen bzw. eigenständig handelnde Unternehmensteile, die zur Entwicklung, Erstellung und Lieferung von Produkten oder Dienstleistungen beitragen und reichen von einer Quelle (Zulieferer der Zulieferer) bis zu einer Senke (Kunden der Kunden) (Beckmann 2012). In der SC werden traditionell Materialflüsse (bzw. Dienstleistungsflüsse), Finanzflüsse und Informationsflüsse unterschieden. Die Betrachtung der Flüsse erfolgt sowohl stromaufwärts zum Lieferanten als auch stromabwärts zum Kunden. Das Ziel einer SC ist, dem Kunden die richti-

gen Güter am richtigen Ort und zum richtigen Zeitpunkt zur Verfügung zu stellen (Jünemann 1989).

Der Begriff SC bedeutet wörtlich übersetzt Versorgungs- oder Lieferkette (Wien-dahl 2004). Es hat sich jedoch keine einheitliche Bedeutung durchgesetzt und die klassische SC findet sich in der deutschen Literatur am ehesten unter dem Begriff der Wertschöpfungskette, die in den 80er Jahren von Porter etabliert wurde (Arnold, Isermann et al. 2008). Jedoch mag die wörtliche Übersetzung mit dem Begriff Kette irreführend sein, wenn beachtet wird, dass die Literatur in ihren Definitionen der SC zumeist den Begriff des Netzwerks heranzieht (Arndt 2013; Arnold, Isermann et al. 2008; Günther und Tempelmeier 2011; Vahrenkamp und Kotzab 2012). So trifft Lambert (2005, S. 2) die Aussage „Strictly speaking, the supply chain is not a chain of business, but a network of business and relationships“ . Blackstone (2010, S. 213) definiert die SC als „The global network used to deliver products and services from raw materials to end customers through an engineered flow of information, physical distribution, and cash“ . Lee und Ng (1997) bezeichnen die SC als Netzwerk von Entitäten. Diese Formulierung greifen auch Fleischmann und Meyr (2001) auf. Sie bezeichnen die SC als eng verflochtenes Netz von Entitäten, in dem Produkte oder Dienstleistungen hergestellt und zum Abnehmer geliefert werden. Auch Riha (2009, S. 78) greift den Netzwerkcharakter auf, in dem die SC als „eine auf Logistikdienstleistung spezialisierte Form eines Netzwerks“ bezeichnet wird. Nach Corsten und Gössinger (2008) entsteht die SC durch den Zusammenschluss von Unternehmen zu Netzwerken, verbunden mit einer unternehmensübergreifenden Betrachtungsweise. Günther und Tempelmeier (2011) lassen als Übersetzung der logistischen Kette die Begriffe SC und Supply Network gleichbedeutend nebeneinander stehen. Christopher (2011) schlägt sogar vor, den Begriff SC um den Zusatz Network zu erweitern, um die Tatsache zu berücksichtigen, dass heute oftmals keine linearen Ketten vorliegen, sondern vielmehr Geflechte von vielen Herstellern, vielen Kunden und Kunden der Kunden. Ähnlich sehen es weitere Autoren, die versucht haben, den Begriff SC um die Komponente des Netzwerks zu ergänzen oder abzuändern (vgl. Chopra und Meindl 2014 oder Stolzle und Otto 2003).

In dieser Arbeit wird dennoch der Begriff SC verwendet, da er die Literatur bis zum heutigen Zeitpunkt dominiert, wohl wissend, dass es sich bei der Kette in der Realität zumeist um ein Netzwerk handelt. Die angeführten Überlegungen und Ansätze finden Berücksichtigung in der Definition von Christopher, die in dieser Arbeit dem Grundverständnis dient.

Definition 2.4 Supply Chain: „The supply chain is the network of organizations that are involved, through upstream and downstream linkages, in the different processes and activities that produce value in the form of products and services in the hands of the ultimate consumer“ (Christopher 1998, S. 15).

Mittels Typologisierung lassen sich verschiedene Arten von SCs durch ihre Merkmale abgrenzen. Hierbei wird anhand von spezifischen Merkmalen und Merkmalsausprägungen zwischen verschiedenen Typen der SCs unterschieden, sodass darauf aufbauend beispielsweise eine Abgrenzung der Anwendungsbereiche von Methoden oder Softwaresystemen durchgeführt werden kann. Da es je nach Sichtweise eine Vielzahl von Merkmalen und Anordnungsmöglichkeiten gibt, haben sich im Verlauf der Zeit sehr unterschiedliche SC-Typologien entwickelt (Giese 2012). Im Kontext der Wissensentdeckung, die als Grundlage Datenbestände benötigt, sind viele der Typologien nicht zielführend. Diese Typologien verwenden Unterscheidungsmerkmale, die keinen Bezug zu den Datenbeständen oder Wissensentdeckungsaufgaben aufweisen. Hier fehlt generell eine Beschreibung von unterschiedlichen SC-Merkmalen und ihr Bezug zu SC-Datenbanken in der Literatur. Basierend auf dem Fokus der Arbeit wurde eine Typisierung der SC nach Meyr und Stadler (2005) durchgeführt, die eine Reihe von funktionellen und strukturellen Merkmalen sowie Unterkategorien entwickelt haben. Während sich die funktionalen Merkmale in vier Kategorien aufgliedern (Art der Beschaffung, Art der Produktion, Art der Distribution und Art des Absatzes), werden strukturelle Merkmale in die Merkmalskategorien der Topografie der SC, Integration und Koordination eingeordnet. Eine Auswahl dieser Kategorien wurde zur Typisierung der hier behandelten SCs genutzt (vgl. Tabelle 2.1), da sich aus der Analyse von Transaktionsdaten Informationen über diese Kategorieausprägungen ableiten lassen. So können beispielsweise Distributionsmuster aus Zeitstempeln für Transporte abgeleitet werden. Kategorien, deren Ausprägungen nicht auf Datenebene zu erfassen sind, finden in der Tabelle 2.1 keine Berücksichtigung.

Tabelle 2.1 zeigt die wichtigsten Merkmalsausprägungen der verwendeten SC-Daten, die der weiteren Arbeit als Einschränkung zugrundegelegt werden. Es wird ausdrücklich darauf hingewiesen, dass branchenspezifische Supply Chains z. B. aus dem Energiesektor aufgrund ihrer anfallenden Daten und Merkmalsausprägungen nicht im Rahmen der vorliegenden Arbeit diskutiert werden.

2.2.2 Datenaufkommen in Supply Chains

Die Veränderungstreiber der Logistik (z. B. Globalisierung und Regionalisierung, Informations- und Kommunikationstechnologien (IuK) oder Kooperationen) wirken auf alle ihr zugeordneten Bereiche, so auch auf die SC. Bedingt durch Globalisierung und exponentielle Verbesserung der IuK steigt die Netzwerkgröße und die Komplexität der SCs signifikant (Serdarasan 2013). Der Anstieg der Netzwerkgröße und die steigende Komplexität beeinflussen unmittelbar die durch die Flüsse gekennzeichneten Prozesse der SC. Die Prozesse wiederum verwenden in der Folge eine größere Datenbasis und erzeugen selbst auch größere Datenbestände, als dies noch vor einigen Jahren der Fall war (European Commission December 2012). Die Größe der Datenbasis ist durch die zwei Dimensionen Volumen und Komplexität gekennzeichnet. Die Dimensionen und ihre Verortung in einer Tabelle können der

Tabelle 2.1: Merkmalsausprägungen der untersuchten Supply Chains nach Meyr und Stadtler (2005)

Merkmalskategorie	Unterkategorien	Typische Merkmalsausprägung
Art der Produktion	Wiederholung der Abläufe	Losfertigung
	Umrüstaufwand	Hoch, abfolgeabhängige Rüstzeiten und -kosten
Art der Distribution	Distributionsstruktur	3 Ebenen
	Distributionsmuster	Dynamisch
	Einsatz von Transportmitteln	Unbegrenzt
Art des Absatzes	Verfügbarkeit zukünftiger Nachfrage	Vorhersage
	Anzahl der Produktarten	Mehrere Produkte
	Grad der individuellen Anpassung	Standardprodukte
	Stückliste	Divergent
	Anteil an Dienstleistungen	Nicht vorhanden
Topographie der SC	Netzwerkstruktur	Gemischt (divergente und konvergente Elemente)
	Grad der Globalisierung	Mehrere Länder
	Position der Entkopplungspunkte	Auftragsproduktion, Produktion in Projekten

Abbildung 2.3 entnommen werden. Die Dimension Volumen bedeutet, dass die Anzahl der einzelnen Datensätze von Relevanz ist. Die Dimension Komplexität, die auch als Varietät bezeichnet wird (Che et al. 2013), kennzeichnet, dass die Anzahl der Einzelelemente innerhalb eines Datensatzes betrachtet wird.

Die SC-Datenbestände sind inhomogen bezüglich der verwendeten Datenkategorien, da Daten sowohl in numerischer als auch alphanumerischer Form enthalten sind. Die unterschiedlichen Kategorien, in denen die Daten vorliegen, stehen in Wechselwirkung zu den Methoden der Wissensentdeckung. Das ist damit begründet, dass alle Methoden spezifische Eingabedaten benötigen und somit Anforderungen an die zugrundeliegende Datenbasis stellen. Neben möglichen Datenkategorien

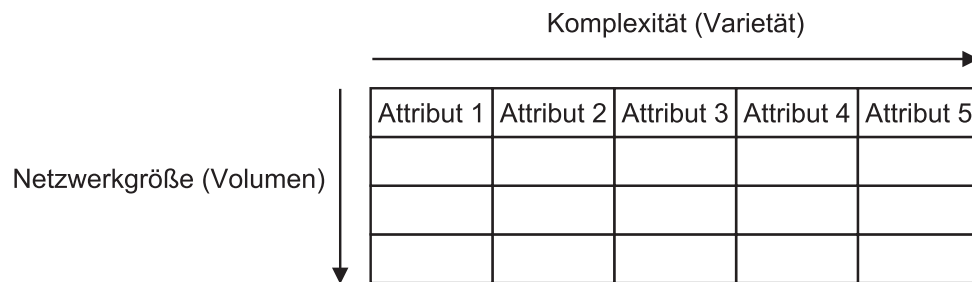


Abbildung 2.3: Netzwerkgröße und Komplexität in Zuordnung zur Tabledarstellung

ist die Datenqualität ein maßgeblicher Einflussfaktor, denn ohne hinreichende Datenqualität sind die Ergebnisse der Wissensentdeckung nicht aussagekräftig. Aufgrund der angesprochenen Wechselwirkungen sollen in den folgenden Abschnitten die Daten, ihre mögliche Kategorisierung und die Datenqualität im Kontext der SC thematisiert werden.

2.2.2.1 Datendefinitionen im Kontext der SC

Im Kontext der Datenbasis als Grundlage der Wissensentdeckung wird der Begriff „Datenbestand“ im Plural verwendet, sofern nicht ein spezifischer Datenbestand im Rahmen von Verfahrensbeschreibungen oder durchgeführten Experimenten adressiert ist. Die Begründung liegt darin, dass im Regelfall mehrere Datenbestände aus separaten Systemen eine Datenbasis bilden. Ein Datenbestand besteht aus einzelnen Datensätzen, die wiederum aus Daten bzw. Datenfeldern zusammengesetzt sind. Zwischen Daten und Datenfeldern lässt sich inhaltlich keine Unterscheidung treffen. So nutzt Mertens (Mertens, Bodendorf et al. 2012) den Begriff Datenfelder als kleinste Einheit, während das Lexikon für Wirtschaftsinformatik (Stickel et al. 1997) in einer äquivalenten Formulierung Daten als kleinste Einheit definiert. Hier soll als kleinste Einheit der Begriff Datenfelder gewählt werden, um eine deutliche Unterscheidung zum Begriff Daten zu ermöglichen.

Datensätze können in Dateien oder Datenbanken gespeichert werden. Im Umfeld der SC sind Datenbanken, insbesondere relationale Datenbanken, von zentraler Bedeutung. Datenbanken werden aus den Rohdaten der unterschiedlichen operativen SC-Systeme mittels Extraktion-Transformation-Laden (ETL) befüllt (Mertens, Back et al. 2001). Der Speichermechanismus, der sich lediglich ein oder mehreren Dateien als Speicherort bedient, findet in komplexen Systemen häufig nur beim Datenaustausch Verwendung. Dieser Datenaustausch kann beispielsweise zwischen zwei Systemen mittels sogenannter Schnittstellen stattfinden. Auf den Datenbanken basieren nachgelagerte Modelle, wie Business Layer, Data-Warehouse oder Data Marts (vgl. Parimala und Pahwa 2008). Diese Modelle greifen mittels ETL auf die Datenbestände der zugrundeliegenden Datenbanken zu (zumindest auf Ebene der einzelnen Zuliefersysteme) oder werden mittels Schnittstellen-Exporten

aus Datenbanken beliefert. Die genannten Modelle wie das Data Warehouse bestehen oftmals aus aggregierten Werten, d. h. sie haben eine hohe Aggregationsstufe, und lassen nur schwer einen Rückschluss auf die zugrundeliegenden Datensätze zu (Messaoud et al. 2006). Daher wird zur Vereinfachung von einer Speicherung in heute üblichen relationalen Datenbanken ausgegangen.

Eine relationale Datenbank besteht aus einer oder mehreren Tabellen, in welchen die Datensätze gespeichert werden. Dabei stellt jede Zeile einer Tabelle einen Datensatz dar – dieser Datensatz wird im Kontext der relationalen Algebra auch als Tupel bezeichnet. Jedes Tupel enthält ein oder mehrere Datenfelder, wobei gleichartige Datenfelder spaltenweise angeordnet werden und als Attribute gekennzeichnet werden. Jedes Attribut kann mit einem Attributnamen versehen werden. Die konkrete Wertebelegung der einzelnen Attribute wird als Attributsausprägung bezeichnet. Alternativ werden die Einträge in Tabellen häufig als Entitäten, ihre Spalten als Merkmal und die Ausprägungen als Merkmalsausprägungen bezeichnet. Eine Erklärung hierfür kann in den Entity-Relationship-Modellen (ERM) gefunden werden. Hier entsprechen die Entitätstypen den Tabellen und die konkreten Instanzen den Entitäten. Im Allgemeinen bezeichnet der Begriff Merkmal die zu beobachtende Eigenschaft eines Objektes und wird auch im spezifischen Kontext als statistische Variable, Untersuchungsmerkmal oder kurz Variable bezeichnet (Fahrmeir et al. 2010, S. 148). Die nachfolgende Definition ist Grundlage für die vorliegende Arbeit:

Definition 2.5 Datensatz: Ein Datensatz ist eine Entität, seine Datenfelder sind Merkmale und die konkreten Belegungen der Datenfelder werden als Merkmalsausprägungen bezeichnet.

2.2.2.2 Kategorisierung der Supply-Chain-Daten

Basierend auf den Veränderungen der Unternehmen (vgl. Abschnitt 2.2.2) nehmen Umfang und Komplexität der Datenbestände in SCs fortwährend zu. Hierbei ist es wichtig, die vorhandenen Daten grundlegend zu kategorisieren, da die verschiedenen Methoden der Wissensentdeckung unterschiedliche Anforderungen an die zugrundeliegende Datenbasis stellen. Dementsprechend sind Daten und die damit verbundenen Kategorisierungssysteme im Allgemeinen nicht für spezifische SC-Typen ausgelegt, sodass der Begriff der Datenkategorisierung deutlich von der SC-Typologie abgegrenzt werden muss (vgl. Abschnitt 2.2.1).

Datenbestände können im Allgemeinen nach verschiedenen Kriterien kategorisiert werden. Eine Vielzahl von Autoren hat unterschiedliche Kategorisierungssysteme vorgeschlagen, wie beispielsweise die Unterteilung in analog und digital (Dworatschek 1989). Dabei ist es heutzutage nicht zweckmäßig, auf diesen Betrachtungsebenen anzusetzen, da die Datenbasis einer globalen SC regelmässig in digitaler Form vorliegt. Für die problemorientierte Sichtweise wird eine Kategorisierung