

# MASTERARBEIT

## **Vorgehensweise zur Aufbereitung von Eingangs- und Ergebnisdaten einer ereignisdiskreten Simulation eines Logistiknetzwerks des Werkstoffhandels zur glaubwürdigen Messbarkeit von komplexen Data-Warehouse-Kennzahlen**

bearbeitet von: Andrea Vennemann

Studiengang: Masterstudiengang Logistik  
Matrikel-Nr.: 135964

Ausgegeben am: 01.10.2014

Eingereicht am: 15.04.2015

Prüfer: Univ.-Prof. Dr.-Ing. Markus Rabe  
Betreuer: Felix Dross, M.Sc. Software Engineering

# Inhaltsverzeichnis

<b>Inhaltsverzeichnis.....</b>	<b>I</b>
<b>Abbildungsverzeichnis .....</b>	<b>III</b>
<b>1 Einleitung.....</b>	<b>1</b>
<b>2 Grundlagen zur ereignisdiskreten Simulation .....</b>	<b>4</b>
2.1 Definitionen .....	5
2.2 Simulationsdaten .....	8
2.2.1 Simulations-Eingangsdaten .....	8
2.2.2 Vorgehensweisen zur Aufbereitung der Eingangsdaten .....	11
2.2.3 Simulations-Ergebnisdaten .....	21
2.2.4 Vorgehensweisen zur Aufbereitung der Ergebnisdaten .....	23
2.3 Simulationsvorgehensmodell mit V&V .....	25
2.4 Simulationswerkzeug Sim Chain .....	28
<b>3 DWH-Kennzahlen .....</b>	<b>32</b>
3.1 Definitionen .....	32
3.1.1 Data Warehouse .....	32
3.1.2 Kennzahl.....	34
3.1.3 Kennzahlensystem .....	35
3.2 Architektur von DWH-Systemen .....	36
3.3 Datenmodell innerhalb eines Data Warehouse .....	39
3.4 Kennzahlen und Kennzahlensysteme im Data Warehouse .....	43
<b>4 Aufbereitung von Simulationsdaten zur glaubwürdigen Messbarkeit von DWH-Kennzahlen .....</b>	<b>46</b>
4.1 Vorgehensweise zur Aufbereitung von Eingangsdaten.....	47
4.1.1 Zieldefinition bezüglich der notwendigen Daten zur Messung der DWH-Kennzahlen.....	49
4.1.2 Bestimmung der Abweichungen zwischen den notwendigen Ergebnisdaten und den notwendigen Zieldaten .....	57
4.1.3 Vorbereitung der Erhebung der realen Zieldaten .....	59
4.1.4 Erhebung der realen Zieldaten.....	60
4.1.5 Zieldefinition bezüglich der notwendigen Eingangsdaten .....	61
4.1.6 Informationsidentifikation.....	62
4.1.7 Erhebungsplanung.....	64

---

4.1.8	Erhebung und Datenerfassung .....	65
4.1.9	Datenstrukturierung .....	65
4.1.1	Statistische Datenanalyse .....	66
4.1.2	Datennutzbarkeitsprüfung .....	66
4.2	Vorgehensweise zur Aufbereitung der Ergebnisdaten .....	68
4.2.1	Datenstrukturierung .....	69
4.2.2	Statistische Datenaufbereitung .....	71
4.2.3	Ergänzung fehlender Attribute .....	72
4.2.4	Aggregation .....	74
4.2.5	Formatierung.....	75
4.3	Einbettung der entwickelten Vorgehensweise in das Simulationsvorgehensmodell zur kontinuierlichen V&V .....	76
4.4	Automatisierter Prozess.....	79
<b>5</b>	<b>Anwendung der Vorgehensweise am Bsp. bestehender DWH-Kennzahlen eines Werkstoffhändlers.....</b>	<b>85</b>
5.1	Charakterisierung des Fallbeispiels.....	85
5.1.1	SAP NetWeaver Business Warehouse .....	86
5.1.2	Simulationsmodell .....	87
5.2	Anwendung der entwickelten Vorgehensweise .....	88
5.2.1	Aufbereitung der Eingangsdaten zur glaubwürdigen Messbarkeit der komplexen DWH-Kennzahlen RDLB und OTIF .....	88
5.2.2	Aufbereitung der Ergebnisdaten zur glaubwürdigen Messbarkeit der komplexen DWH- Kennzahl RDLB .....	98
5.3	Fazit .....	101
<b>6</b>	<b>Zusammenfassung und Ausblick .....</b>	<b>103</b>
	<b>Literaturverzeichnis.....</b>	<b>I</b>

## Abbildungsverzeichnis

Abbildung 1:	Begriffsbestimmung Eingangsdaten .....	9
Abbildung 2:	Vorgehensweise für das Eingangs-Datenmanagement in der ereignisdiskreten Simulation [SkJo08].....	14
Abbildung 3:	Integration der Informationsgewinnung in das Simulationsvorgehensmodell [BJH+07, S.6] .....	17
Abbildung 4:	Prozessorientiertes Vorgehensmodell zur Informationsgewinnung [BJH+07, S.7] .....	18
Abbildung 5:	Vorgehensmodell bei der Simulation mit V&V [RSW08a, S.5] .....	26
Abbildung 6:	SimChain-Architektur, Ausschnitt aus [Fech15].....	29
Abbildung 7:	SimChain-Basistabellen und ihre Abhängigkeiten [FeGu14, S.10] ...	30
Abbildung 8:	Szenarienaufbau durch die Verknüpfung verschiedener Konfigurationen [FeGu14, S.13] .....	31
Abbildung 9:	Screenshot - gefüllte SimChain-Ergebnistabelle stat_summary_location.....	31
Abbildung 10:	Architektur eines DWH-Systems [Fark11, S.67].....	37
Abbildung 11:	Klassifikationsschemata – links: einfacher Aufbau, Dimension Zeit; rechts: parallel Aufbau, Dimension Produkt (i.A.a [Baue03, S.27; Holt99, S.137] ) .....	41
Abbildung 12:	multidimensionaler Datenwürfel .....	42
Abbildung 13:	Erweiterung des IGP's zur Vorgehensweise zur Aufbereitung von Eingangsdaten zur glaubwürdigen Messbarkeit von DWH-Kennzahlen ( i.A.a. [BJH+07, S.8]) .....	49
Abbildung 14:	überschneidende Daten zweier Kennzahlensysteme.....	51
Abbildung 15:	Unterteilung der notwendigen Daten zur Messung der DWH Kennzahlen .....	52
Abbildung 16:	Ermittlung der Eingangscubes im Schatten-DWH-System und der zugehörigen Ergebnistabellen in SimChain .....	54
Abbildung 17:	Überschneidung zweier Kennzahlenberechnungen im DWH-System	56
Abbildung 18:	Abweichungsmöglichkeiten zwischen den notwendigen Ergebnisdaten der Simulation und den notwendigen Zieldaten zum Eingang in das Schatten-DWH-System .....	58
Abbildung 19:	Prozessorientiertes Vorgehensmodell zur Aufbereitung von Ergebnisdaten zur glaubwürdigen Messbarkeit von komplexen DWH-Kennzahlen. ....	68
Abbildung 20:	Integration der entwickelten Vorgehensweise in das Simulationsvorgehensmodell von Rabe et al. (i.A.a [JKW09, S.5]) ..	77

---

Abbildung 21: Automatisiertes System zur glaubwürdigen Messung von DWH-Kennzahlen .....	80
Abbildung 22: Architektur des SAP Business Warehouse [SAP15] .....	86
Abbildung 23: Screenshot - Teilausschnitt des Simulationsmodells .....	88
Abbildung 24: Kennzahlenbaum für die Kennzahl RDLB.....	89
Abbildung 25: Kennzahlenbaum für die Kennzahl OTIF .....	90
Abbildung 26: Screenshot – Spalten der ausgewählten Ergebnistabelle stat_summary_skus zur Messung des RDLBs.....	91
Abbildung 27: Berechnung der Umschlagshäufigkeit in SimChain [FeGu14, S.197]92	
Abbildung 28: Screenshot - Aufbereitung der Ergebnisdaten in den Hilfstabellen des Excel-Tools.....	100

# 1 Einleitung

Unternehmen des Werkstoffhandels sind durch hohen Wettbewerbsdruck, die Verankerung des Wertorientierungsgedankens sowie steigendes Umweltbewusstsein geprägt. Hieraus ergeben sich veränderte Anforderungen an Logistik-Netzwerke der Unternehmen. Die Dynamik und Komplexität des Wandels der Anforderungen werden beispielsweise durch kürzer werdende Produktlebens- und Innovationszyklen und den damit einhergehenden zunehmenden zeitlichen Druck für die Unternehmen deutlich. Die Logistik ist in ihrer Querschnittsfunktion besonders von einer Vielzahl der veränderten Rahmenbedingungen betroffen. Sie umfasst die wirtschaftlichen Grundfunktionen wie die Beschaffung, die Produktion, die Distribution und die Entsorgung sowie die Koordination innerhalb der Supply Chain und weist somit in allen Bereichen Berührungspunkte mit den sich verändernden Rahmenbedingungen auf [HoNo09, S.1 ff.]. Zum Erhalt ihrer Wettbewerbsfähigkeit müssen die Unternehmen den sich aus dem Wandel ergebenden Anforderungen gerecht werden. Aus der notwendigen Anpassung der Unternehmen an die veränderten Rahmenbedingungen ergeben sich komplexe Logistiknetzwerke. Diese zeichnen sich durch ihre Globalität, Mehrstufigkeit, Multimodalität und zahlreichen internen Abhängigkeiten aus.

Vor dem Hintergrund komplexer Logistikstrukturen und der internen Wirkungszusammenhänge der Beeinflussungsgrößen innerhalb der Logistik, ist die Steuerung der Logistik in den Unternehmen eine komplexe Aufgabe. Zur Lenkung und Kontrolle der Betriebs- und Logistikprozesse in den Unternehmen und der dafür notwendigen transparenten Darstellung ihrer Betriebsabläufe und Logistikstrukturen sowie deren Interdependenzen werden Steuerungsinstrumente eingesetzt. Betriebswirtschaftliche Kennzahlen stellen ein solches Steuerungsinstrument dar. Diese werden insbesondere in Unternehmen des Großhandels verwendet. Sie enthalten verdichtete Informationen über erfolgsentscheidende Sachverhalte und Zusammenhänge im Unternehmen und können somit zur Kontrolle und Problemerkennung auch bei komplexen Unternehmensstrukturen eingesetzt werden. Eine notwendige Voraussetzung für die Kennzahlenberechnung ist die Analysierbarkeit der Daten. Diese wird durch den Einsatz eines Data-Warehouse-Systems (DWH-Systems) ermöglicht. In einem DWH-System werden die Daten aus den verschiedenen operativen Systemen vereint gespeichert und für Analysezwecke aufbereitet. Die dafür durchzuführenden, teilweise sehr komplizierten und vielschichtigen Berechnungs-, Filterungs- und Vereinheitlichungsoperationen sowie Aggregationen können manuell nur schwer berechnet werden. DWH-Kennzahlen haben somit, vor dem Hintergrund der heutzutage vorliegenden komplexen Logistikstrukturen, eine große Bedeutung. Der Handlungsbedarf, der sich aus den Kennzahlenwerten ableiten lässt, wird durch die

Höhe der Abweichung des Kennzahlenwertes vom zuvor definierten Soll-Wert bestimmt.

Neben dem Einsatz von Kennzahlen als Steuerungsinstrument in Unternehmen hat sich die ereignisdiskrete Simulation als eine wichtige Analysemethode in Produktion und Logistik etabliert. Die ereignisdiskrete Simulation ermöglicht, Vorhersagen für Systeme zu treffen, die Variabilität, Vernetzung und Komplexität aufweisen [Robinson 2003, S.7]. Dies ist in der manuellen Berechnung zu aufwendig und auf Grund dynamischer Prozesse wie beispielsweise einer schwankenden Kundennachfrage teilweise nicht realisierbar. So ist es durch die Möglichkeit der Zeitraffung und der leichten Modifizierbarkeit des Simulationsmodells möglich, die Auswirkungen verschiedener durchgeführter Maßnahmen auf die klassischen Kennzahlen direkt zu ermitteln. Dabei steht oftmals die Glaubwürdigkeit des Simulationsmodells und damit einhergehend die Glaubwürdigkeit der Ergebnisse im Fokus. Ein Simulationsmodell gilt als glaubwürdig, wenn es vom Auftraggeber als hinreichend genau akzeptiert wird, um als Entscheidungsunterstützung zu dienen [RSW08a, S.2].

Dieser Arbeit liegt ein Forschungsvorhaben zugrunde, welches die Messung von Maßnahmenauswirkungen auf DWH-Kennzahlen thematisiert. Hintergrund des Forschungsvorhabens ist die Unsicherheit über die Auswirkungen einzelner Maßnahmen auf die verschiedenen Kennzahlensysteme. Maßnahmen, die zur Verbesserung des Wertes einer DWH-Kennzahl dienen, wirken sich gleichzeitig teilweise auch auf andere DWH-Kennzahlen aus. Diese Interdependenzen lassen sich in der Theorie nur schwer detailliert abbilden. Mit Hilfe der ereignisdiskreten Simulation ist es möglich, die Auswirkungen der Maßnahmen auf die einzelnen DWH-Kennzahlen aufzuzeigen. Voraussetzung zur Messung der Auswirkungen ist, dass die Kennzahl glaubwürdig am Simulationsmodell gemessen werden kann. Hier ordnet sich die vorliegende Arbeit ein.

Aufgrund der komplexen Logistiknetzwerke und Prozesse in den Unternehmen haben sich auch die Kennzahlen hinsichtlich ihrer Komplexität gesteigert. Für den Einsatz der ereignisdiskreten Simulation, beispielsweise zur Analyse der möglichen Auswirkungen von Maßnahmen auf solche komplexen Kennzahlen, ist es notwendig, die DWH-Kennzahlen am Simulationsmodell glaubwürdig messen zu können. Anders als bei einfachen Kennzahlen, für die die Aufbereitung der notwendigen Eingangsdaten in der Literatur hinreichend beschriebene Schritte umfasst, bedarf es für komplexe DWH-Kennzahlen einer umfassenderen Aufbereitung nicht nur der Eingangs- sondern auch der Ergebnisdaten. Da es sich hierbei um umfangreiche und komplexe Bearbeitungsschritte handelt, ist hierfür einer genaueren Vorgehensweise erforderlich.

Das Ziel dieser Arbeit ist die Entwicklung der zuvor beschriebenen Vorgehensweise zur Aufbereitung der Eingangs- und Ergebnisdaten einer ereignisdiskreten Simulation,

sodass komplexe DWH-Kennzahlen glaubwürdig am Simulationsmodell gemessen werden können. Die Simulation konzentriert sich dabei auf das System eines Logistiknetzwerks des Werkstoffhandels. Aus der oben beschriebenen Problematik und der Aufgabenstellung ergeben sich damit folgende zu beachtende Punkte:

- Die Vorgehensweise soll die einzelnen Schritte zur Aufbereitung der Eingangsdaten ausführlich darstellen. So soll genau vorgegeben werden, wie man die zur Messung der Kennzahl notwendigen Daten identifiziert und diese in der passenden Form in das Simulationsmodell einpflegt.
- Ebenso soll detailliert erläutert werden, wie die Ergebnisdaten der Simulation bearbeitet werden müssen, um diese an geeigneter Stelle in die Berechnungsstruktur des Data Warehouse einzupflegen.
- In der gesamten Vorgehensweise soll der Aspekt der Glaubwürdigkeit berücksichtigt werden.

Zur Erreichung der beschriebenen Zielsetzung wird ein zuvor definiertes methodisches Vorgehen gewählt. Zu Beginn erfolgt eine Literaturrecherche, um ein umfassendes Grundlagenverständnis auf den Gebieten der Simulationsdaten und der DWH-Kennzahlen zu erhalten. Der Fokus der Recherche liegt insbesondere auf dem derzeitigen Stand der Technik bezüglich bereits bestehender Vorgehensweisen zur Aufbereitung von Eingangs- und Ergebnisdaten sowie der Messung von DWH-Kennzahlen mit Hilfe der ereignisdiskreten Simulation.

Im nächsten Schritt erfolgt die Bewertung des Standes der Technik hinsichtlich der Anforderungen, die an die zu entwickelnde Vorgehensweise gestellt werden. Die Bewertung wird durch den Abgleich der vorhandenen Vorgehensweisen mit den Anforderungen an die zu entwickelnde Vorgehensweise realisiert. Hieraus wird der zu erfüllende Entwicklungsbedarf abgeleitet.

Anschließend wird auf Grundlage der bereits bestehenden Vorgehensweisen eine Vorgehensweise zur Aufbereitung von Eingangs- und Ergebnisdaten zur glaubwürdigen Messbarkeit von komplexen DWH-Kennzahlen entwickelt. Das Ziel besteht darin den zuvor ermittelten Entwicklungsbedarf abzudecken.

Das entstandene Grundgerüst der entwickelten Vorgehensweise wird durch die Kenntnisse über DWH-Kennzahlen sowie der Simulationsdaten anschließend detailliert ausgearbeitet. Außerdem wird die entwickelte Vorgehensweise in das Vorgehensmodell zur kontinuierlichen V&V für die Simulation von *Rabe et al.* eingebettet [RSW08a, S.5].

Um das so erstellte Vorgehensmodell zu evaluieren, wird die entwickelte Vorgehensweise auf ein bestehendes DWH-Kennzahlensystem in einem Logistiknetzwerk des Werkstoffhandels und ein derzeit entstehendes Simulationsmodell angewandt.

## 2 Grundlagen zur ereignisdiskreten Simulation

Die Simulation ist eine Methode zur Analyse von Systemen. Das zu untersuchende System wird in einer abstrahierten Form in einem Modell abgebildet. Abstraktion bedeutet in diesem Fall eine vereinfachte Nachbildung der Strukturen und des Verhaltens des Systems. Die Simulation selbst beinhaltet die Durchführung von Experimenten an diesem Modell. Aus den Simulationsergebnissen werden Rückschlüsse für das reale System gezogen. [RoMä11, S.13]

Simulationsmodelle können auf drei Arten charakterisiert werden. Hinsichtlich der zeitlichen Abhängigkeit wird zwischen statischen und dynamischen Modellen unterschieden. Die Klassifizierung nach der Zufälligkeit beinhaltet die Unterteilung in deterministische und stochastische Formen der Simulation. Eine dritte Gliederungsmöglichkeit liegt bezüglich der vorliegenden Art der Variablen vor. Hierbei wird nach kontinuierlichen und diskreten Variablen differenziert. [LaKe00, S.5 f.]

In den Bereichen von Produktion und Logistik kommt in der Regel die ereignisdiskrete Simulation zum Einsatz. Die ereignisdiskrete Simulation zeigt ausschließlich Zeitpunkte im System auf, in denen das System eine Veränderung erfährt. Dabei wird mit Hilfe der Simulation eine Serie von Ereignissen modelliert, in deren Momenten eine Zustandsänderung erscheint. Die ereignisdiskrete Simulation ordnet sich wie folgt in die zuvor beschriebene Klassifizierung der Simulation ein. Produktions- und Logistiksysteme weisen ein dynamisches Verhalten auf, das mit Hilfe von stochastischen Komponenten dargestellt wird. Um das dynamische Verhalten angemessen abbilden zu können, wird das zeitliche Verhalten des Systems am Modell simuliert. Durch die Verwendung von stochastischen Zufallsgrößen werden die zufälligen, nicht vorhersehbaren Ereignisse innerhalb eines Logistiknetzwerkes, wie beispielsweise die Kundennachfrage, abgebildet. Innerhalb von Logistiksystemen ergeben sich Zustandsänderungen nur zu bestimmten diskreten Zeitpunkten. Die Zustandsänderungen resultieren durch den Eintritt von Ereignissen. Ein typisches Ereignis in Produktions- und Logistiksystemen ist das Eingehen eines neuen Kundenauftrages. [LaKe00, S.5 f.; Robinson 2003, S.15 f.; RoMä11, S.14]

Die Grundbausteine der ereignisdiskreten Simulation lassen sich nach *Pidd* in zwei Gruppen untergliedern. Dabei wird zwischen Objekten aus denen das System zusammengesetzt wird und Operationen, in die die Objekte eingebunden sind differenziert. Zu den Objekten des Systems zählen sowohl Entitäten wie beispielsweise Maschinen in einer Fabrik als auch Ressourcen wie beispielsweise die Anzahl der Passagiere an einem Bussteig. Beide stellen individuelle Elemente des Systems dar und unterscheiden sich durch ihren individuellen Modellierungsgrad. Während für jede Entität das Verhalten detailliert beobachtet wird, bestehen die Ressourcen aus zählbaren identischen Objekten deren individueller Zustand nicht analysiert wird. Entitäten

werden unter Beibehaltung ihrer Individualität teilweise in Gruppen zusammengefasst. Hierdurch ist beispielsweise der Bezug auf gleiche oder ähnliche Entitäten in einem Schritt möglich. Werden gleiche oder ähnliche Entitäten dauerhaft in einer Gruppe zusammengefasst, so wird diese als Klasse bezeichnet. Ein Set umfasst hingegen Entitäten, die nur für einen bestimmten Zeitraum eine Gruppe abbilden. Diese verändert sich in Abhängigkeit vom Zustand der Entitäten. Entitäten werden durch Attribute beschrieben. Während der Simulation finden verschiedene Operationen statt, welche den Zustand der Objekte verändern. Es wird zwischen Ereignissen, Aktivitäten und Prozessen unterschieden. Ein Ereignis zeigt den Zeitpunkt auf, an dem das System, nach Meinung des Analytikers, eine entscheidende Zustandsänderung aufweist. Die Zustandsänderungen der Entitäten resultieren aus Aktivitäten. Diese werden wiederum durch das Eintreffen eines bestimmten Ereignisses initialisiert. Die chronologische Abfolge von Ereignissen wird als Prozess bezeichnet. Mit diesem kann der gesamte oder ein Teil des Lebenszyklus einer vorübergehenden Entität dargestellt werden. Die Simulationsuhr gibt den Zeitpunkt an, der in der aktuell simulierten Zeit in der Simulation erreicht wurde. [BCN+14, S.90; Pidd04, S.63 ff.]

## 2.1 Definitionen

Um eine gemeinsame Begriffswelt mit einheitlichen Fachtermini zu schaffen, folgen für diese Arbeit relevante Definitionen. Diese stehen im Zusammenhang mit den Begrifflichkeiten der ereignisdiskreten Simulation sowie der Glaubwürdigkeit einer Simulationsstudie. Ein Teil der Definitionen bezieht sich speziell auf den Bereich Materialfluss, Logistik und Produktion.

### Simulation

Im Rahmen dieser Arbeit wird der Begriff der Simulation entsprechend VDI-Richtlinie 3633 für den Bereich Materialfluss, Logistik und Produktion wie folgt definiert:

*„Simulation ist die Nachbildung eines Systems mit seinen dynamischen Prozessen in einem Modell, um zu Erkenntnissen zu gelangen, die auf die Wirklichkeit übertragbar sind.“ [Vere14, S.3]*

Wie in der Definition beschrieben, nutzt die Simulation ein Modell eines Systems. In diesem wird das Verhalten des Systems mit einem geringeren Detaillierungsgrad abgebildet. Nach VDI-Richtlinie 3633 ist ein Modell definiert als:

*„Vereinfachte Nachbildung eines geplanten oder existierenden Systems mit seinen Prozessen in einem anderen begrifflichen oder gegenständlichen System.“ [Vere14, S.3]*

In dieser Arbeit handelt es sich dabei um das System eines Logistiknetzwerks. Logistiksysteme zeichnen sich durch komplexe Kunden-Lieferanten-Beziehungen aus.

Den verschiedenen Akteuren in einem Logistiknetzwerk wie beispielsweise dem Lieferanten, Hersteller oder Empfänger sind unterschiedliche Standorte zugeordnet. Hierzu zählen beispielsweise Produktionsstätten, Lager und Hubs. Diese sind wiederum durch den Güter-, Informations- und Finanzfluss miteinander verknüpft. [Arno08, S.934]

Allgemein wird ein System im Umfeld von Materialfluss, Logistik und Produktion folgendermaßen definiert:

*„Eine von ihrer Umwelt abgegrenzte Menge von Elementen, die miteinander in Beziehung stehen.“ [Vere14, S.4]*

Um mittels Simulation zu Erkenntnissen zu gelangen, müssen Simulationsexperimente durchgeführt werden. Zur Definition von Simulationsexperimenten bedarf es zunächst der Begriffsbestimmung eines Simulationslaufes. In einem Simulationslauf wird das Systemverhalten über einen definierten Zeitraum genau abgebildet. Das bedeutet, dass das Simulationsmodell innerhalb eines Simulationsdurchlaufes einmal über den zuvor festgelegten Zeitraum ausgeführt wird. In der VDI-Richtlinie 3633 ist ein Simulationslauf wie folgt definiert:

*„Nachbildung des Verhaltens eines Systems mit einem spezifizierten ablauffähigen Modell über einen bestimmten (Simulations-)Zeitraum (Simulationszeit) wobei gleichzeitig die Werte untersuchungsrelevanter Zustandsgrößen erfasst und gegebenenfalls statistisch ausgewertet werden.“ [Vere14, S.4]*

Um aus den Ergebnisdaten des Simulationslaufes Rückschlüsse auf das reale durch Variabilität gekennzeichnete System ziehen zu können, müssen diese statistisch abgesichert werden. Hierzu bedarf es einer mehrfachen Wiederholung eines Simulationsdurchlaufes mit identischen Eingangs-Parametern. Diese werden als Replikationen bezeichnet. Lediglich die Startwerte für die Zufallsgeneratoren müssen sich hierbei unterscheiden.

Parameter sind auf der einen Seite Eingangsgrößen des Simulationsmodells, die beim Aufruf des Simulationsmodells von dem aufrufenden Programm, hier dem eigentlichen Simulator (vgl. 2.4), übergeben werden. Auf der anderen Seite stellen auch Ausgangsgrößen des Simulationsmodells, die an den Simulator zurückgegeben werden, Parameter dar. Die Eingangsgrößen werden als inputs und die Ausgangsgrößen als outputs bezeichnet. Inputs können sowohl schwankende Variablen, wie beispielsweise die Zwischenankunftszeiten von Aufträgen, als auch Strategien oder Entscheidungsparameter sein. Die outputs stellen Größen dar, auf Basis derer die Leistung des Systems gemessen werden kann. [BCN+14, S.35; Spril5a]

Ein Simulationsexperiment ist nach der VDI-Richtlinie folgendermaßen definiert:

*„Gezielte empirische Untersuchung des Verhaltens eines Modells durch wiederholte Simulationsläufe mit systematischer Parameter- oder Strukturvariation“ [Vere14, S.3]*

Da die VDI-Definition keinen Unterschied zwischen Replikationen und Wiederholungen mit unterschiedlichen Parametern berücksichtigt, wird ein Experiment von *Rabe et al.* in Anlehnung an die oben aufgezeigte VDI-Definition als „...eine Reihe von Läufen (bzw. Replikationen) mit unterschiedlichen Parametern [RSW08a, S.13]“ bezeichnet.

### **Glaubwürdigkeit**

Ein Simulationsmodell und damit einhergehend die daraus resultierenden Ergebnisse gelten als glaubwürdig, wenn sie vom Auftraggeber als korrekt angesehen werden [LaKe00, S.265]. Die Glaubwürdigkeit des Modells kann durch Maßnahmen zur Überprüfung des Simulationsmodells erhöht werden. Diese werden innerhalb der Verifikation und Validierung durchgeführt [RSW08a, S.17].

Im Gegensatz zur VDI-Richtlinie 3633, in der mit der Verifikation der Korrektheitsnachweis des Simulationsmodells gefordert wird, wird von *Rabe et al.* im Rahmen des Vorgehensmodells zur kontinuierlichen V&V für die Simulation folgende Definition für die Verifikation verwendet:

*„Verifikation ist die Überprüfung, ob ein Modell von einer Beschreibungsart in eine andere Beschreibungsart korrekt transformiert wurde.“ [RSW08a, S.14]*

Der Definition liegt folgende Überlegung zugrunde. Der für die Durchführung der Verifikation erforderliche formale Nachweis der Korrektheit, kann aufgrund hoher Systemkomplexität nicht vollständig realisiert werden. Daher wird von einer nur hinreichenden Genauigkeit gesprochen. Weiterhin wird die VDI-Definition dahingehend abgeändert, dass sich die Überprüfung nicht auf das korrekte Modell selbst, sondern auf die Transformation von einem in das andere Modell bezieht.

Laut der VDI-Richtlinie 3633 ist die Validierung „die Prüfung dieser hinreichenden Übereinstimmung von Modell und Original“ [Vere14, S.37]. Darauf aufbauend wird im Vorgehensmodell zur kontinuierlichen V&V für die Simulation folgende Definition genutzt:

*„Validierung ist die kontinuierliche Überprüfung, ob die Modelle das Verhalten des abgebildeten Systems hinreichend genau wiedergeben.“ [RSW08a, S.15]*

Hinreichend genau bedeutet in diesem Zusammenhang, dass das Modell das für die Zielerreichung relevante Verhalten des betrachteten Systems widerspiegelt. Analog zur

Verifikation werden die Validierungs-Elemente während der gesamten Simulationsstudie durchgeführt. So werden beispielsweise die bereits erfassten Daten zur Erstellung des Simulationsmodells einer Validierung unterzogen. Im Unterschied zur Verifikation, bei der die Frage nach der Korrektheit der Modelltransformation im Vordergrund steht, liegt der Fokus der Validierung auf der Korrektheit des Modells. [RSW08a, S.15 f.]

## 2.2 Simulationsdaten

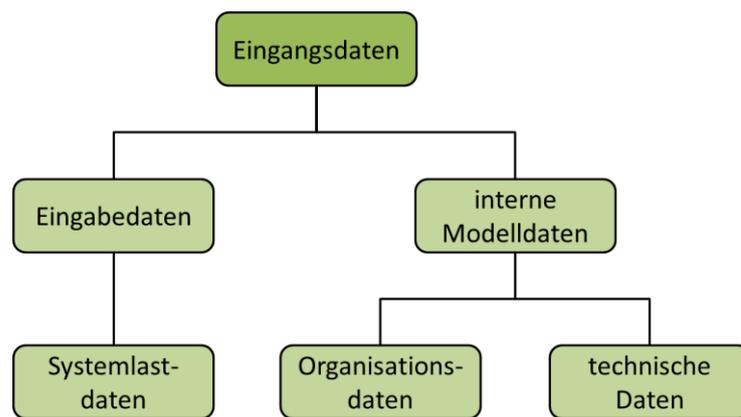
Simulationsdaten umfassen alle Daten und Informationen, die hinsichtlich des Modells vor, nach und während des Simulationsexperiments von Bedeutung sind. Sie bilden die Basis für den Aufbau des Simulationsmodells, zur Durchführung von Simulationsexperimenten sowie der Ergebnisinterpretation. Die Simulationsdaten unterteilen sich nach VDI-Richtlinie 3633 in vier verschiedenen Datenarten. Diese sind Eingabe-, und Ergebnisdaten sowie interne Modell- und Experimentdaten [Vere14, S.15]. Sowohl die Eingabedaten als auch die internen Modelldaten dienen zur Beschreibung des zu untersuchenden technischen Systems. Während die internen Modelldaten „bei der Modellentwicklung festgelegte Konstanten sowie während des Simulationslaufs berechnete Variablen“ [Vere14, S.15] sind, werden die Eingabedaten „vom Nutzer des Modells bereitgestellt“ [Vere14, S.15]. Die Eingabedaten umfassen die Daten, die zur Initialisierung eines Modells und der davon abhängigen Durchführbarkeit von Simulationsexperimenten notwendig sind. Die Ergebnisdaten werden während sowie am Ende des Experiments am Simulationsmodell ausgegeben und stellen Rückmeldungen über das Verhalten des simulierten Systems dar. Alle Daten, die im Zusammenhang mit dem Simulationsexperiment selbst stehen, werden unter den Experimentdaten zusammengefasst. Diese beinhalten beispielsweise die Simulationsdauer und die Anzahl der Simulationsdurchläufe. Die aufgezeigten Datenarten der Simulation werden in den folgenden Unterkapiteln ausführlich beschrieben. [Schu12, S.168; Vere14, S.15 ff.]

### 2.2.1 Simulations-Eingangsdaten

In vielen Werken zur Simulationsliteratur wird keine direkte Unterscheidung zwischen Eingabedaten und internen Modelldaten getroffen, stattdessen wird der Begriff Eingangsdaten geprägt. So sprechen sowohl *Banks* als auch *Rabe et al.* von input data oder Eingangsdaten [BCN+14, S.15; RSW08a, S.63]. *Law und Kelton* verweisen allgemein auf Daten, die zur Darstellung von Eingangswahrscheinlichkeitsverteilungen und zur Spezifizierung von Modellparametern verwendet werden [LaKe00, S.84 f.]. In der Vorgehensweise zum Eingangsdaten-Management von *Skoogh und Johansson* und dem Informationsgewinnungsprozess (IGP) von *Bernhard und Wenzel* wird ebenfalls der Begriff Eingangsdaten verwendet.

Außerdem wird bei der in dieser Arbeit eingesetzten Simulationssoftware SimChain keine Differenzierung zwischen Eingabedaten und internen Modelldaten vorgenommen.

Im Rahmen der Arbeit werden unter dem Begriff Eingangsdaten die Eingabedaten und die internen Modelldaten zusammengefasst (vgl. Abbildung 1). Durch die Eingangsdaten werden die oben beschriebenen Eingangs-Parameter näher spezifiziert. Allgemein stellen die Eingangsdaten die zu erfassenden Daten für die Entwicklung eines Simulationsmodells dar. Der Umfang und der Erfassungszeitraum der Eingangsdaten werden durch die zu analysierende Zielsetzung und die Komplexität des Systems bestimmt.



**Abbildung 1: Begriffsbestimmung Eingangsdaten**

Die erforderlichen Daten zur Beschreibung von Materialfluss-, Produktion- und Logistiksystemen für die Simulation werden laut VDI-Richtlinie 3633 übergeordnet in Systemlastdaten, Organisationsdaten und technische Daten gegliedert [Vere14, S.34]. Die Systemlastdaten beziehen sich auf abzuarbeitende Aufträge sowie auf zu produzierende Produkte eines Systems. Die Organisationsdaten umfassen sowohl Arbeitszeitmodelle und -regelungen, die Ressourcenzuordnung von Mitarbeitern und Maschinen, als auch Strategien und Notfallpläne für den Ablauf im betrachteten System. Die klassischen innerbetrieblichen Fabrikstruktur-, Fertigungs-, Materialfluss- und Stördaten werden unter den technischen Daten zusammengefasst.

Die drei aufgezeigten Datentypen lassen sich in die oben beschriebene Begriffsbestimmung der Eingangsdaten einordnen (vgl. Abbildung 1). Da die technischen und organisatorischen Daten oft statische Größen darstellen, die zur Erstellung des Modells benötigt werden, bilden diese meist interne Modelldaten ab. Hierzu zählen somit auch Daten, durch welche die innere Modellogik abgebildet wird. Die Systemlastdaten sind hingegen als Eingabedaten einzuordnen. Diese sind meist veränderlich und werden in der Experimentierphase vom Nutzer in das Simulationsmodell eingepflegt. [Berg13, S.20; Vere14, S.15]

Die zum Aufbau des Simulationsmodells notwendige Datenbeschaffung der Eingangsdaten zählt zu den wichtigsten und zeitaufwändigsten Schritten einer Simulationsstudie. Die Daten stehen häufig nicht in der benötigten Form, in der notwendigen Qualität oder nur unvollständig zur Verfügung. Der Aufwand zur Datenbeschaffung liegt laut [BDKR01, S.132] für den Auftraggeber bei 48% und für den Dienstleister bei 13% des Gesamtaufwandes. [BCN+14, S.16; BDKR01, S.132; CBW08]

In Produktions- und Logistiksystemen existieren viele Situationen in denen die Vorgänge innerhalb des Systems nicht vollständig vorhergesagt werden können. Das Nachfrageverhalten der Kunden oder die Reparaturzeit für eine Maschine lassen sich beispielsweise nicht genau prognostizieren. Einer der Hauptvorteile der Simulation ist die Möglichkeit diese Wandelbarkeit darzustellen. Es gibt drei Methoden die Volatilität in ein Simulationsmodell zu integrieren. Voraussetzung zur Anwendung dieser Methoden ist, dass die entsprechenden historischen Daten vorhanden sein müssen.

- Die gesammelten, realen Daten werden direkt während der Simulation eingelesen. Diese Form der Simulation wird auch trace-driven-Simulation genannt. Der Nachteil dieser Simulationsmethode besteht darin, dass sie sich nur zur Abbildung vergangenheitsbezogener Abläufe eignet. Diese Form der Simulation wird häufig zur Modell-Validierung verwendet. Hierbei findet ein Vergleich der Ergebnisdaten der Simulation mit den realen Output-Daten des Systems statt.
- Aus den gesammelten, realen Daten wird eine empirische Verteilung gebildet. Stichproben aus dieser Verteilung gehen in die Simulation ein. Durch die entwickelte Verteilung können kontinuierlich Daten zwischen dem Minimum und Maximum der realen Daten erzeugt werden.
- Die Parameter einer theoretischen Verteilung werden an die gesammelten realen Daten angepasst. Stichprobenartige Werte dieser Verteilung gehen in die Simulation ein. Zur Bestimmung einer passenden theoretischen Verteilung sind drei Schritte durchzuführen. Die grundlegende Verteilungsfunktion kann auf Basis der bekannten Eigenschaften des Verlaufs der realen Daten bestimmt werden. Das Histogramm aus den gesammelten Daten zeigt hierfür eine hilfreiche Darstellung auf. Im zweiten Schritt werden die Parameter der gewählten Verteilung an die realen Daten angepasst. Abschließend wird durch einen Anpassungstest die Anpassungsgüte der entwickelten Verteilung an die realen Daten überprüft.

[BCN+14, S.331 ff.; LaKe00, S.296 ff.; Robinson 2003, S.100 ff.]

Die Verfügbarkeit der benötigten Daten zum Aufbau und zur Durchführung der Simulation ist unterschiedlich. So ist ein Teil der erforderlichen Daten bereits

vorhanden, für einen anderen Teil müssen diese zunächst erfasst werden. Für den Fall, dass Prozesse noch nicht existieren oder die Beschaffung der Daten durch Gesetze oder Regeln verhindert wird, ist keine Erhebung historischer Daten möglich. In solchen Fällen werden Expertenmeinungen oder adäquate Abschätzungen verwendet. *Robinson* führt hierzu eine entsprechende Klassifizierung der Daten entsprechend ihrer Verfügbarkeit und Erfassungsmöglichkeit ein. Die Daten der Kategorie A sind verfügbar, da sie entweder bereits bekannt sind oder zuvor bereits erfasst wurden. B-Daten sind nicht verfügbar, können aber durch entsprechende Erhebungsmethoden erfasst werden. Die C-Daten sind weder verfügbar noch können sie gesammelt werden. [BCN+14, S.171; Robinson 2003, S.97 ff.]

### 2.2.2 Vorgehensweisen zur Aufbereitung der Eingangsdaten

Während der Durchführung einer Simulationsstudie besteht, analog zu anderen modellgestützten Analyseverfahren, oftmals Unsicherheit und Unkenntnis über ein zweckmäßiges Vorgehen in Bezug auf die Ermittlung der Eingangsdaten. Da das Management der Eingangsdaten einen großen Teil des zeitlichen Gesamtaufwandes des Simulationsprojektes ausmacht (vgl. 2.2.1), bedarf es einer strukturierten Vorgehensweise für die Handhabung der Eingangsdaten. Mit Hilfe der Vorgehensweise soll die Qualität und die Performanz des Eingangsdaten-Managements erhöht werden. [JKW09, S.1; SkJo08]

Nach *Skoogh und Johansson* wird die Eingangsdaten-Management-Phase als gesamter Prozess zur Erstellung qualitätsgesichteter und an die Simulationssoftware angepasster relevanter Eingangsdaten für das Simulationsmodell definiert [SkJo08]. Grundsätzlich unterteilt sich dieser Prozess in die Elemente Datenbeschaffung und Datenaufbereitung. Nach *Rabe et al.* umfasst die Datenbeschaffung neben der Datenerhebung auch die Identifikation der Art und des Umfangs der erforderlichen Daten [RSW08a, S.51]. Innerhalb der Datenaufbereitung werden die Rohdaten durch verschiedene Schritte in eine passende Form für den Eingang in die Simulation überführt. Hierzu zählen beispielsweise die Filterung erforderlicher Daten, die Transformation der Daten in eine passende Form oder die Abbildung der Daten durch eine statistische Verteilung. Weiterhin erfolgt die Prüfung der Daten zur Eignung für die gegebene Aufgabenstellung. [RSW08a, S.51 ff.]

Die meisten Werke zur ereignisdiskreten Simulation behandeln die Thematik des Eingangsdaten-Managements nur kurz. So erwähnen *Law und Kelton* die Datenbeschaffung lediglich als Teilschritt der Simulationsstudie [LaKe00, S.84 f.]. Im *Handbook of Simulation* von *Banks* beschränken sich die Ausführungen für die Datenbeschaffung einerseits auf das Aufzeigen möglicher Fehler der bereits vorhandenen Daten und andererseits auf Ratschläge zur Datenbeschaffung, wie beispielsweise über die Anzahl der Beobachtungen [Vinc98, S.59 f.]. *Robinson*

entwickelt eine Kategorisierung der Daten (vgl. 2.2.1) entsprechend ihrer Verfügbarkeit und Erfassbarkeit und beschreibt den Umgang mit nicht verfügbaren Daten. Zusätzlich nimmt er Bezug auf die Genauigkeit und das Format der Daten [Robinson 2003, S.95 ff.]. Bei *Wenzel et al.* werden vor dem Hintergrund des qualitätskonformen Vorgehens in einer Simulationsstudie und dem grundlegenden Qualitätskriterium der konsequenten Dokumentation, Checklisten und Datenstrukturen unter anderem für die Datenbeschaffung und Datenaufbereitung definiert [Wenz08]. Diese können an entsprechender Stelle in das in Unterkapitel 0 vorgestellte Simulationsvorgehensmodell von *Rabe et al.* eingeordnet werden [RSW08a, S.13]. Weiterhin werden allgemeine Hinweise zur Datenbeschaffung und Datenaufbereitung wie beispielsweise zur Erhebung von Informationen der Daten gegeben [Wenz08, S.119 ff.]. Allgemein enthält die klassische Literatur zum Aufbau von Simulationsmodellen viele Informationen bezüglich der Darstellung umfangreicher Rohdaten. Hierzu zählt beispielsweise die statistische Aufbereitung von unvorhersehbaren Schwankungsdaten durch die Bestimmung einer passenden statistischen oder empirischen Verteilung. (vgl. [BCN+14, S.331 ff.; LaKe00, S.235 ff.; Robinson 2003, S.100 ff.]) *Pidd* thematisiert ausschließlich die Erzeugung von Zufallszahlen zur Abbildung von stochastischen Elementen in einem Simulationsmodell [Pidd04, S.175 ff.].

In der modernen Literatur hingegen existieren einige Ansätze und teilweise vollständige Vorgehensweisen zum Vorgehen innerhalb der Eingangsdaten-Management-Phase (vgl. [BSJ+09; LeSe97; HiOn14; SkJo08; CBW08]). Diese unterscheiden sich durch ihren Schwerpunkt und dem damit verbundenen Umfang. Außerdem werden Verifikations- & Validierungsaspekte unterschiedlich berücksichtigt. Zunächst werden einige Ansätze zum Vorgehen innerhalb der Eingangsdaten-Management-Phase beschrieben. Anschließend erfolgt die Erläuterung von zwei vollständigen Vorgehensweisen.

Mit ihrer Methode *A methodology for data gathering and analysis in a logistics simulation project* haben *Lehtonen und Seppälä* 1997 einen der ersten Ansätze über die zu dem Zeitpunkt unerforschten Bereichen der Problemdefinition, Datenbeschaffung und Datenanalyse entwickelt. Ziel dieser Methode ist es potenzielle Verbesserungsbereiche innerhalb eines Systems aufzuzeigen. [LeSe97]

Einen aktuelleren Ansatz zeigen *Hill und Onggo* in ihrer Vorgehensweise zur Datenidentifizierung und Aufbereitung auf. Diese beruht auf einer Fallstudie, in der das Vorgehen von Mitarbeitern einer Management-Berater-Firma in der Eingangsdaten-Management-Phase einer Simulationsstudie beobachtet wurde. Die Vorgehensweise unterscheidet sich insofern von den anderen Ansätzen, als dass hier der Auftraggeber mit einbezogen wird. [HiOn12; HiOn14]

Die Forschung im Bereich der Datensammlung wird vom automatisierten Datensammelungsprozess dominiert [HiOn14]. Dieser besteht aus der Kombination

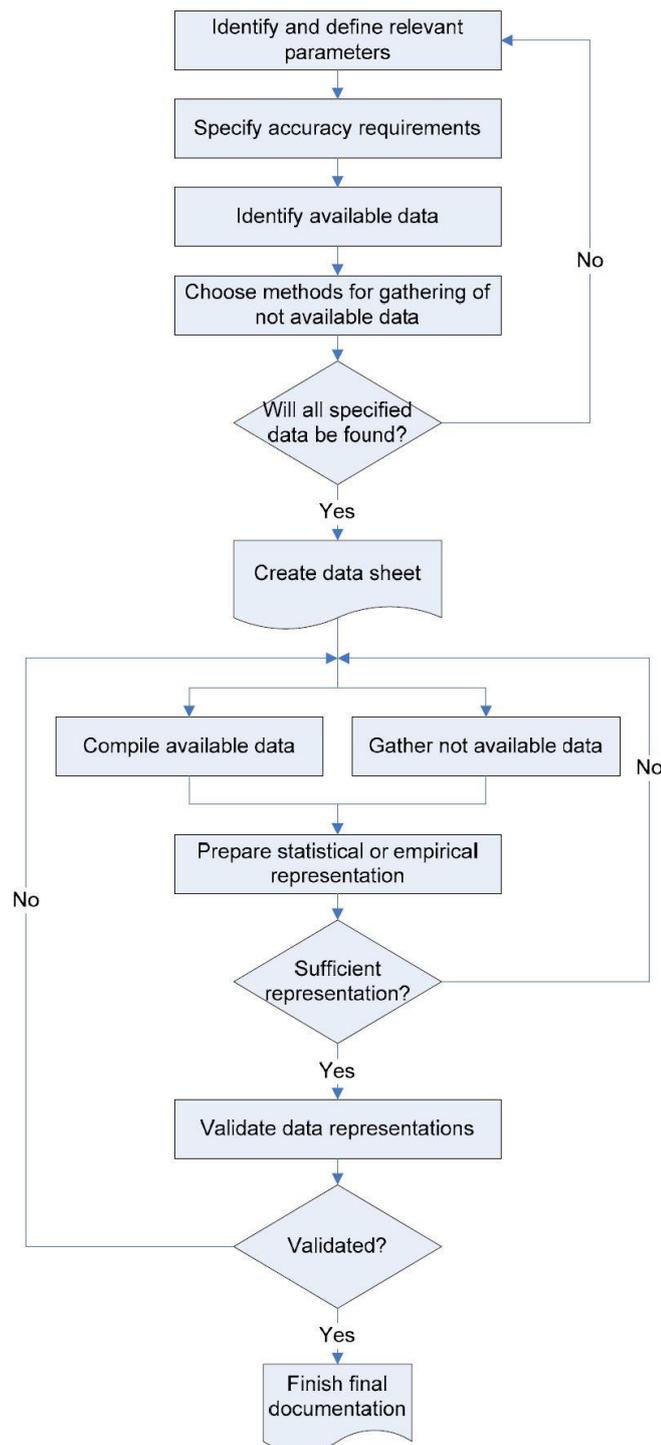
einer Datenerfassungsmethode und einer Enterprise Resource Planning Software. *Robertson und Perera* beschreiben in ihrem Artikel *Automated data collection for simulation?* wie das Corporate Business System als Datenquelle für die Simulation genutzt werden kann und somit eine schnellere Datensammlung gegenüber dem heutzutage üblichen manuellen Prozess möglich ist. [RoPe02]

Zur Identifikation der benötigten Daten für eine Simulationsstudie haben *Baron et al.* Checklisten entwickelt [BDKR01]. Diese geben Auskunft darüber, welche Daten für die Darstellung bestimmter Prozesskettenelemente und damit zusammenhängender Kategorien notwendig sind. Die Checklisten bilden unter anderem die Grundlage für die *Methodik zur systematischen Informationsgewinnung für Simulationsstudien* von *Csanady et al.* [CBW08]. Hierbei handelt es sich um eine checklistenbasierte Methodik zur Bestimmung der für ein bestimmtes Simulationsziel benötigten Datenbasis. Mittels einer Kreuztabelle erfolgt die Zuordnung von bestimmten Simulationszielen zu den dafür benötigten Informationen. In einem weiteren Schritt werden die entsprechend notwendigen Erhebungs- und Auswertungsmethoden zugeordnet.

### **2.2.2.1 A methodology for input data management in discrete event simulation projects**

Die von *Skoogh und Johansson* entwickelte Vorgehensweise umfasst die vier Schritte Identifikation, Erfassung, Analyse und Speicherung der Daten. In Abbildung 2 sind folgend die 13 Teilschritte der Vorgehensweise dargestellt. Die erstellte Vorgehensweise lässt sich in die Vorgehensmodelle für ein gesamtes Simulationsprojekt einordnen, wie es beispielsweise von *Law und Kelton* vorgestellt wird [LaKe00, S.84 f.]. Innerhalb dieser spezifiziert sie den Teilschritt Eingangsdaten-Management.

Die Vorgehensweise beginnt mit der Identifikation und Definition der relevanten Parameter für das Simulationsmodell. Es wird dokumentiert, wie diese ermittelt und im Modell dargestellt werden. Im zweiten Schritt erfolgt die Spezifikation der Genauigkeitsanforderungen für jeden einzelnen Parameter. In Abhängigkeit von der Bedeutung des Parameters innerhalb des Simulationsmodells wird jeweils eine eigene Genauigkeitsanforderung erstellt, wodurch bei der Erhebung ein effektives Vorgehen gewählt werden kann. Bevor die Methoden zur Datenerfassung ausgewählt werden, wird zunächst analysiert, welche Daten im Unternehmen bereits vorhanden sind. Diese sind direkt verfügbar und bilden nach der oben beschriebenen Kategorisierung von *Robinson A*-Daten ab. Die vorhandenen Daten müssen auf ihre Zweckmäßigkeit bezüglich ihrer Verwendung im Simulationsprojekt überprüft werden. Für die Parameter der Kategorie B, die nicht verfügbar aber erfassbar sind, erfolgt eine Auswahl möglicher Erfassungsmethoden. Für den Fall, dass keine entsprechenden Daten erfassbar sind, werden entsprechende Schätzverfahren bestimmt.



**Abbildung 2: Vorgehensweise für das Eingangs-Datenmanagement in der ereignisdiskreten Simulation [SkJo08]**

Ist es für manche Parameter nicht möglich die benötigten Daten in einem ausreichenden Umfang sowie einer ausreichenden Datenqualität und Datengenauigkeit zu bestimmen, so muss zunächst die Genauigkeitsanforderung oder die Relevanz der Parameter angepasst werden. Im nächsten Schritt wird ein Datenblatt zur Dokumentation der Rohdaten und der aufbereiteten Daten erstellt. Hierdurch wird die Datenkohärenz in der Datensammlung unterstützt.

Es folgt die eigentliche Phase der Datenbeschaffung. Während die verfügbaren Daten aus den Datenquellen extrahiert und zusammengetragen werden, folgt die Ermittlung der nicht verfügbaren Daten mit Hilfe der zuvor bestimmten Erfassungsmethoden. Die nicht erfassbaren Daten werden durch Expertenwissen oder historische Daten ähnlicher Prozesse bestimmt. Bei den A-Daten kann es sich um Rohdaten oder bereits analysierte, für die Simulation aufbereitete Daten handeln. Für die Rohdaten wird gemäß der zuvor festgelegten Genauigkeitsanforderungen eine ausreichende Menge an Datenpunkten aus den Datenquellen extrahiert. Diese werden in einem ersten Schritt mit Hilfe verschiedener Berechnungsschritte in eine analysefähige Form gebracht und im Datenblatt dokumentiert.

Im Anschluss an die Bereitstellung der für die Analyse aufbereiteten Daten werden diese innerhalb der Datenaufbereitungsphase in eine für das Simulationsmodell passende Form transformiert. Die Darstellungen von konstanten Daten erfordert dabei wesentlich weniger Aufwand als die von schwankenden Daten. Zur Abbildung der Variabilität vgl. 2.2.1. Anschließend erfolgt die Überprüfung, ob die Datendarstellungen den Genauigkeitsanforderungen genügen. Hierzu werden Anpassungstests oder der graphische Vergleich mit den Original-Daten genutzt. Werden die Anforderungen nicht erfüllt, bedarf es einer zusätzlichen Datensammlung und Analyse.

Bevor die Daten in das Simulationsmodell eingepflegt werden, erfolgt eine Validierung der aufbereiteten Daten. Kann die Validierung nicht für alle Parameter erfolgreich durchgeführt werden, so werden die Ursache und Probleme ermittelt.

Das sogenannte Abschlussdokument besteht aus dem Datenblatt, das sowohl die identifizierten Parameter, die Rohdaten als auch die aufbereiteten Daten in Form von Verteilungen enthält und einem Datenbericht. In diesem sind die Erhebungsmethoden, Datenquellen, Validierungsergebnisse und Annahmen bei der Eingabe enthalten. [SkJo08]

Die beschriebene Vorgehensweise von *Skoogh und Johansson* wird von *Bengtsson et al.* um den Aspekt der einheitlichen Speicherung erweitert [HiOn14]. Hierdurch ergibt sich eine einfachere Verarbeitungsmöglichkeit der Daten für weitere ereignisdiskrete Simulationsprojekte. Im Anschluss an die vorgestellte Vorgehensweise werden die Daten zur Identifizierung und Sammlung mit Hilfe des Generic Data Management - Tools zeitnah aus den Quellen extrahiert, strukturiert und in das Core Manufacturing Simulation Format (CMSD) umgewandelt. Das Ziel ist es, dass dieses Format zu einem Standard innerhalb von ereignisdiskreten Simulationsprojekten wird, sodass eine problemlose Verwendung der Daten innerhalb verschiedener Simulationssoftware-Programme erfolgen kann. [BSJ+09; SkJo08]

### 2.2.2.2 Vorgehensmodell zur Informationsgewinnung

Der Informationsgewinnungsprozess (IGP) nach *Bernhard et al.* wurde im Rahmen des Projektes des Sonderforschungsbereichs 559 Modellierung großer Netze in der Logistik der Universität Dortmund entwickelt. Der Prozess ist Bestandteil eines Methodennutzungsmodells, das ein übergreifendes Konzept für die Gewinnung von Eingangsdaten für große Netze der Logistik (GNL) darstellt. Hierbei handelt es sich um ein Dreiebenen-Konzept, in dem der Informationsgewinnungsprozess die oberste Ebene abbildet. Die darin geforderte Methodenauswahl erfolgt über die in der zweiten Ebene angeordneten Methodentaxonomien. Diese stellen Kriterien zur Klassifizierung der einzusetzenden Methoden bereit. Die zur Verfügung stehenden Methoden sind in der untersten Ebene aufgezeigt. Die drei Schichten des Methodennutzungsmodells sind durch Metainformationen miteinander verbunden. Für eine detaillierte Beschreibung des Methodennutzungsmodells vgl. [JKW09, S.2 ff.].

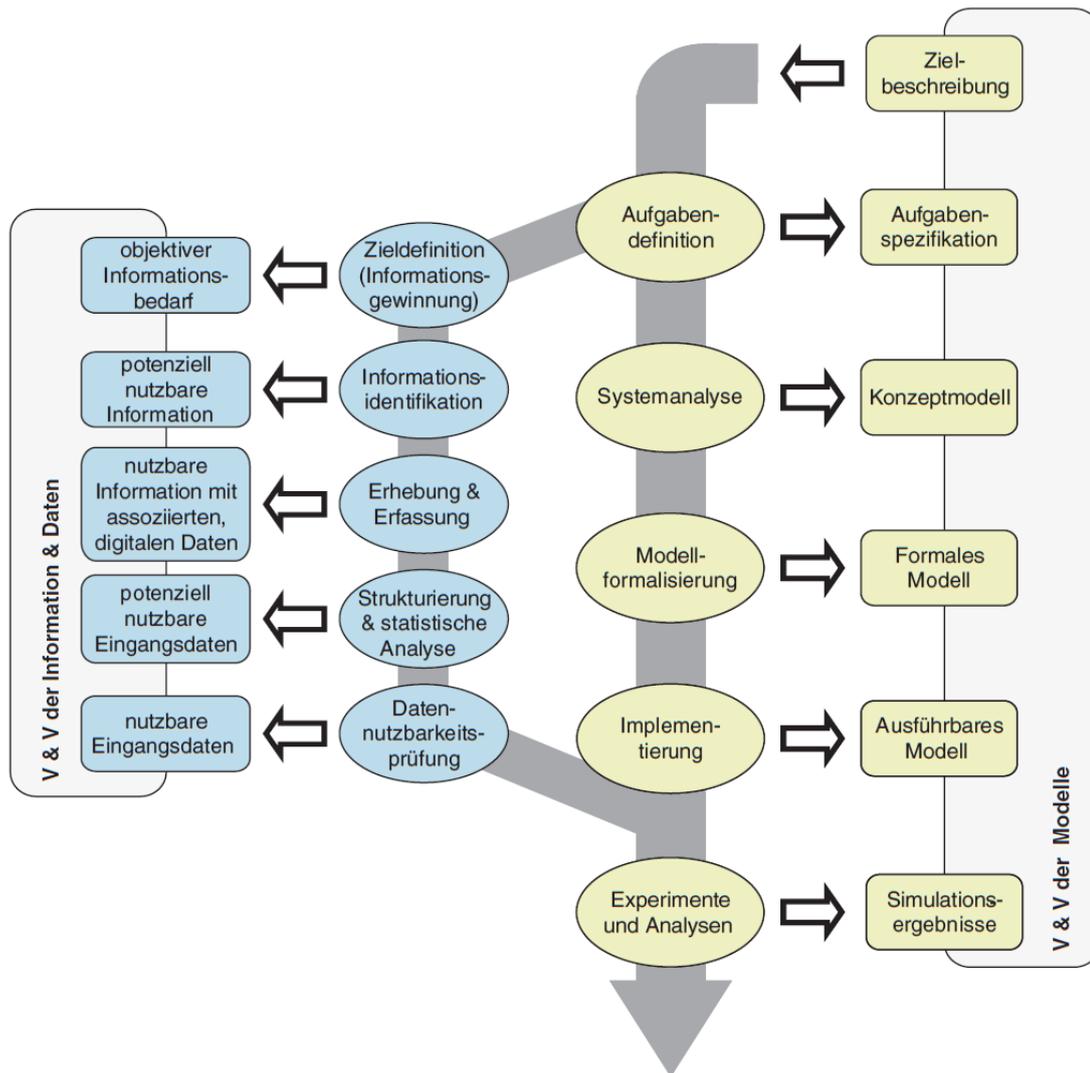
Im Vorgehensmodell zur Informationsgewinnung stehen, im Gegensatz zum heute üblichen Vorgehen, nicht die Daten selbst, sondern die erforderliche, richtige und gültige Information hinsichtlich der Aufgabenstellung im Vordergrund. Es erfolgt dabei eine Separation zwischen dem Informationsbedarf und der jeweiligen Codierung der Informationen in Daten. Information und Daten werden dazu folgendermaßen begrifflich voneinander abgegrenzt:

*„Daten bestehen aus analogen oder digitalen Signalen oder Zeichen (Syntax) und dienen der Repräsentation und Darstellung von Information zum Zwecke der Vorbereitung.“ [WeBe03]*

*„Information umfasst Aussagen und Beschreibungen zu Eigenschaften und Strukturen bestimmter Gegenstände oder Sachverhalte (Semantik); sie wird mitgeteilt und entgegengenommen und ist aufgrund dieser Sender-Empfänger-Relation zweckgebunden. Information wird durch Daten codiert, ist kommuniziertes und formalisiertes Wissen und dient der Wissensvermehrung.“ [WeBe03]*

Durch die Verbindung von Informations- und Datensicht in einem Prozessmodell (vgl. Abbildung 4) ergibt sich die Möglichkeit, den Handlungsablauf angefangen vom notwendigen Informationsbedarf, der sich aus der Zieldefinition ableiten lässt, bis zu den tatsächlich nutzbaren Eingangsdaten aufzuzeigen. [JKW09, S.4 ff.]

Der IGP kann allgemein in modellgestützten Analyseverfahren verwendet werden. Beispielhaft wurde dieser vor dem Hintergrund der Einordnung in den Kontext der ereignisdiskreten Simulation in das Vorgehensmodell zur kontinuierlichen V&V in der Simulation von *Rabe et al.* integriert [RSW08a, S.13].

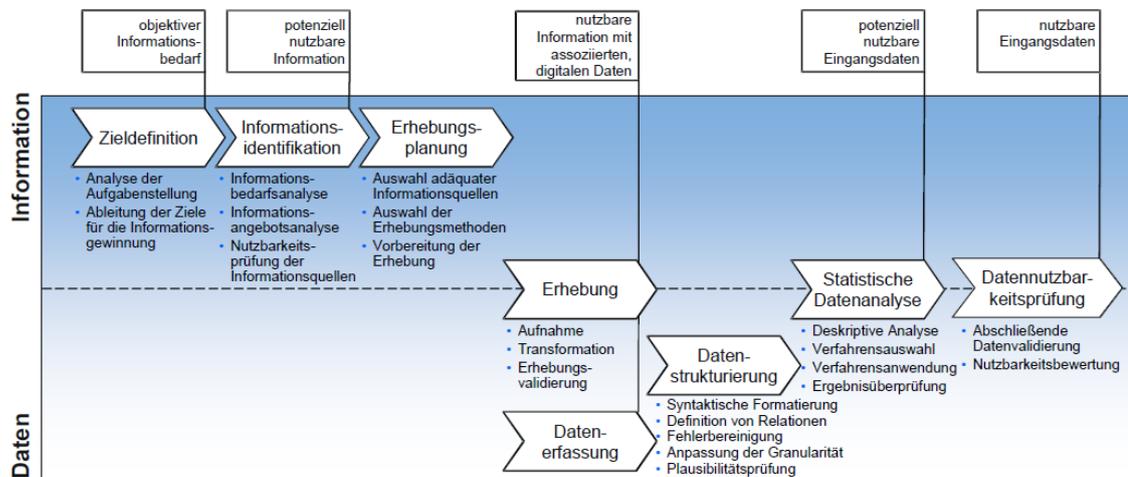


**Abbildung 3: Integration der Informationsgewinnung in das Simulationsvorgehensmodell [BJH+07, S.6]**

Wichtig für die Einordnung des IGP in das Simulationsvorgehensmodell ist die gesonderte Behandlung von Datengewinnung und Modellbildung. Durch den Informationsgewinnungsprozess erfolgt eine Detaillierung der beiden Phasen Datenerhebung und Datenaufbereitung des Vorgehensmodells. In Abbildung 3 wird das erweiterte Simulationsvorgehensmodell, indem der Informationsgewinnungsprozess in das Vorgehensmodell zur kontinuierlichen V&V integriert ist, vorgestellt. Sowohl der Informationsgewinnungsprozess selbst, als auch das übergeordnete Vorgehensmodell zur kontinuierlichen V&V sind durch die V&V aller Zwischenergebnisse der einzelnen Phasen geprägt.[BJH+07, S.5]

In Abbildung 4 wird eine detaillierte Übersicht des prozessorientierten Vorgehensmodells zur Informationsgewinnung dargestellt. Die Prozessschritte selbst werden durch sechsköpfige Pfeile abgebildet. Die aufgelisteten Unterpunkte der Prozessschritte zeigen die inhaltliche Ausrichtung und die viereckigen Kästchen an den

Prozessschritten das zugehörige Phasenergebnis auf. Die einzelnen Prozessschritte sind je nach ihrem inhaltlichen Fokus auf der Informations- oder Datenebene eingeordnet. [JKW09, S.6]



**Abbildung 4: Prozessorientiertes Vorgehensmodell zur Informationsgewinnung**  
[BJH+07, S.7]

Das Vorgehensmodell beginnt mit der Zieldefinition. Hierbei soll durch die Analyse der Aufgabenstellung und der drauf aufbauenden Ableitung der Ziele für die Informationsgewinnung der objektive Informationsbedarf ermittelt werden. Die Analyse der Aufgabenstellung umfasst eine Detaillierung der Aufgabenstellung hinsichtlich der geforderten Granularität, der Abbildungsform und der Bedeutung der vorhandenen Informationen. Aus den hieraus resultierenden Informationen und unter Berücksichtigung der Simulation als Analysemethode sowie der Systembeschreibung werden die Ziele der Informationsgewinnung abgeleitet. Die Zieldefinition dient zusammen mit der Informationsidentifikation zur Vorbereitung der Anwendung der Methoden zur Informationsgewinnung.

Die Informationsidentifikation umfasst drei Schritte. Diese sind die Informationsbedarfsanalyse, die Informationsangebotsanalyse und –bewertung sowie die Nutzbarkeitsprüfung der Informationsquellen. Innerhalb der Informationsbedarfsanalyse leitet der Planer durch eine methodische und strukturierte Analyse des objektiven Bedarfs seinen subjektiven Informationsbedarf her. Dies erfolgt unter Beachtung einer problembezogenen Detaillierung und Konkretisierung der Aufgabenstellung sowie der Zieldefinition. Die Informationsangebotsanalyse beinhaltet eine Bewertung aller verfügbaren Informationsquellen im Hinblick auf die Verfügbarkeit, die Zuverlässigkeit, den Erhebungsaufwand und der Qualität sowohl der Quelle selbst als auch der eigentlichen Daten. Die Nutzbarkeitsprüfung der Informationsquellen bildet den dritten Teil der Informationsidentifikation ab. Die Prüfung umfasst den Abgleich der Informationsangebots durch die Informationsquellen

mit dem aufgezeigten subjektiven Informationsbedarf. Ergebnis der gesamten Informationsidentifikation ist die Festlegung der potenziell nutzbaren Informationen.

Anschließend folgt die Erhebungsplanung. Hier werden adäquate Informationsquellen und Erhebungsmethoden ausgewählt. Außerdem erfolgt die Vorbereitung der Erhebung. Die Erhebungsplanung verfolgt das Ziel, dass der Anwender in der Lage ist, erforderliche Daten ohne zusätzliches Expertenwissen zu erheben. Im Anschluss an die Auswahl einer geeigneten Informationsquelle durch die Gegenüberstellung des Erhebungsaufwandes zu den damit gewonnenen Ergebnissen, erfolgt die Auswahl der zu verwendenden Erhebungsmethode. Die Art der Quelle bestimmt, ob es sich um eine Primär- oder um eine Sekundärerhebung handelt. In Abhängigkeit von verschiedenen Kriterien, wie beispielsweise dem geschätzten Zeitverbrauch oder dem Aufwand für Personal und Kosten, erfolgt die Auswahl der Erhebungsmethoden für die verschiedenen Informationsquellen. Zur Vorbereitung der Erhebung werden hierfür notwendige Erhebungsunterlagen erstellt und Genehmigungen eingeholt. Außerdem werden die erstellten Erhebungsinstrumente durch Voruntersuchungen auf gültige und zuverlässige Erhebungsergebnisse untersucht.

Innerhalb der Datenerhebung werden die in der Informationsbedarfsanalyse aufgezeigten erforderlichen Informationen und Daten entsprechend der Ergebnisse der Erhebungsplanung aus den Datenquellen extrahiert. Anschließend werden die erhobenen Daten erfasst und in digitaler Form für die nachfolgenden Prozesse zur Verfügung gestellt. Weiterhin bedarf es für den Fall, dass die Daten in keiner standardisierten Form vorliegen, einer entsprechenden Transformation. Zur Validierung der erhobenen Daten werden diese mit dem aufgezeigten subjektiven Informationsbedarf abgeglichen. Außerdem wird der Erhebungsvorgang auf Qualität und Vollständigkeit hin überprüft.

Durch die Datenstrukturierung werden die nach der Erhebung digital vorliegenden Daten für die nachfolgende statistische Datenanalyse aufbereitet. Hierzu erfolgen die syntaktische Formatierung, die Definition von Relationen, die Fehlerbereinigung, die Anpassung der Granularität und die Plausibilitätsprüfung. Durch die syntaktische Formatierung werden die Daten in eine analysefähige Form gebracht. Hierzu zählt beispielsweise die Konvertierung in ein einheitliches Datenformat. Abschließend erfolgt eine Überprüfung der Formatierung. Durch die Definition von Relationen werden die Beziehungen zwischen den verschiedenen Variablen oder ganzen Tabellenteilen aufgezeigt. Die Beziehungen können verschiedener Natur sein. Hierzu zählen beispielsweise Abhängigkeiten. So sind manche Variablen nur bei einem bestimmten Wert einer anderen Variablen erkennbar. Ebenso lassen sich hierarchische Beziehungen aufzeigen. Variablen stehen ebenfalls zueinander in Beziehung, wenn diese eine gemeinsame Dimension aufweisen. Die Fehlerbereinigung beinhaltet die Überprüfung

der Daten in Bezug auf die Eigenschaften wie Typ und Wertebereiche der zugehörigen Variablen. Außerdem erfolgt in der Fehlerbereinigung das Auffinden und Untersuchen von Ausreißern. Im Anschluss an die Identifikation der fehlerhaften Einträge in der Datenbank werden diese gelöscht und als fehlender Wert klassifiziert oder falls möglich, durch den Vergleich mit übrigen Daten korrigiert. In einem nächsten Schritt wird für jede Variable bestimmt in welcher Einheit und mit welcher Genauigkeit die Ausprägungen erfasst werden sollen. Maßgeblich entscheidend hierfür ist der im subjektiven Informationsbedarf erschlossene Detaillierungsgrad. Die Phase der Datenstrukturierung wird mit einer Plausibilitätsprüfung abgeschlossen. Diese beinhaltet die Überprüfung der Datenbank auf Vollständigkeit und Widerspruchsfreiheit. Ebenso wird geprüft, ob durch die Einträge in der Datenbank der subjektive Informationsbedarf abgedeckt wird. An dieser Stelle werden Instrumente der Datenverifikation und Datenvalidierung genutzt. Zunächst wird durch die Datenverifikation nachgewiesen, ob die Daten inhaltlich richtig und widerspruchsfrei sind. Dies erfolgt beispielsweise durch den Nachweis von Eigenschaften der Daten, die durch theoretisches Wissen oder experimentelle Ergebnisse bekannt sind. Im nächsten Schritt wird durch die Datenvalidierung überprüft, ob der subjektive Informationsbedarf mit den vorliegenden Daten befriedigt werden kann. Hierzu sind ein entsprechender Umfang der Datenbasis sowie die Repräsentativität der Daten hinsichtlich des Untersuchungsgegenstandes zu überprüfen. Kann keine Verifikation und Validierung im erforderlichen Umfang durchgeführt werden, so muss zwingend eine neue Erhebung oder Nacherhebung erfolgen.

Basis der statistischen Datenanalyse sind die durch die Datenstrukturierung aufbereiteten Daten. Diese werden innerhalb der statistischen Datenanalyse mit Hilfe verschiedener Prozesse zu potenziell nutzbaren Eingangsdaten für die Simulation erstellt. Zu Beginn steht eine deskriptive Analyse. Diese führt zu einer Beschreibung der Beobachtungen in der Datenbank in Hinblick auf die allgemeinen statistischen Eigenschaften der Daten. Die deskriptive Analyse kann beispielsweise durch die Verwendung von Verteilungskennzahlen, wie z.B. Lage- und Steuerungsparameter oder die grafische Veranschaulichung der empirischen Verteilungen z.B. durch Histogramme erfolgen. Im nächsten Schritt werden die zu verwendenden Verfahren ausgewählt, mit denen sichergestellt werden kann, dass die Anforderungen, die sich durch den subjektiven Informationsbedarf ergeben, abgedeckt werden. Es folgt die Anwendung der ausgewählten statistischen Verfahren. Im Anschluss an die Überprüfung auf Vollständigkeit werden die Ergebnisse erneut verifiziert. Durch die Datenverifikation soll auf der einen Seite die inhaltlich logische Richtigkeit durch Plausibilitätsüberprüfungen und auf der anderen Seite die statistische Richtigkeit durch die Überprüfung der Gültigkeit der jeweiligen Modellannahmen erfolgen. Führt die Datenverifikation zu negativen Ergebnissen, so kann dies in Abhängigkeit von der

Ursache zu einer erneuten Durchführung der Erhebungsprozesse sowie der statistischen Datenanalyse führen.

Die Ergebnisse der statistischen Datenanalyse liefern die potenziell nutzbaren Eingangsdaten. Aus diesen werden innerhalb der Datennutzbarkeitsprüfung, dem letzten Schritt des Informationsgewinnungsprozesses, die nutzbaren Eingangsdaten abgeleitet. Nachdem innerhalb der statistischen Datenanalyse der Fokus auf der Verifikation der Daten lag, erfolgt in der Datennutzbarkeitsprüfung eine abschließende Datenvalidierung. Hierzu werden die vorliegenden potenziell nutzbaren Eingangsdaten mit den anfangs identifizierten subjektiven Informationsbedarf abgeglichen und somit die Zweckdienlichkeit der vorliegenden Daten überprüft. Die Daten erweisen sich als zweckdienlich, wenn sie den Anforderungen des subjektiven Informationsbedarfs, wie der Plausibilität, Vollständigkeit, Glaubwürdigkeit und Richtigkeit gerecht werden. Im Anschluss an die Datenvalidierung wird die Nutzbarkeitsprüfung durchgeführt. Hierbei werden die potenziell nutzbaren Eingangsdaten im Zusammenhang mit dem vorliegenden Simulationsmodell betrachtet. An dieser Stelle ergibt sich die Schnittstelle zum übergeordneten Simulationsvorgehensmodell mit V&V. Hier werden die Daten unter Beachtung des Simulationsmodells auf Korrektheit und Angemessenheit überprüft. Die Angemessenheit bezieht sich auf die hinreichende Abbildung des Systems durch das Simulationsmodell einschließlich der Daten. Ergebnisse der IGPs sind validierte Eingangsdaten, die im Zusammenhang mit einem Simulationsmodell genutzt werden können. [BJH+07; BeWe05; JKW09]

### **2.2.3 Simulations-Ergebnisdaten**

Das Ergebnis eines Simulationslaufes besteht aus Datensätzen, die sowohl parallel als auch am Ende der Experimente am Simulationsmodell gewonnen werden. Diese sogenannten Trace-Daten werden an bestimmten Messpunkten im Modell erfasst und spiegeln die Zustandsänderungen der ortsfesten und dynamischen Modellelemente des Systems wider. Neben der unmittelbaren Aufzeichnung der Ereignisse in Form von Trace-Daten in einer Datei oder Datenbank, können diese auch direkt in eine entsprechende Zählvariable verbucht und somit verdichtet werden. Die zugehörigen Trace-Daten werden dabei nicht gespeichert. Die erzielbare Qualität der Ergebnisdaten der Simulation hängt von der Güte der genutzten Informationen zur Bestimmung der Eingangsdaten ab. [DaCh10, S.153; Vere14, S.15; Vere97, S.3 ff.]

Damit die Ergebnisse einer Simulation Rückschlüsse auf das Verhalten des realen Systems ermöglichen, müssen diese entsprechend aufbereitet, verdichtet und interpretiert werden. Der damit verbundene Aufwand ist von der Variabilität des Systems abhängig. In einem deterministischen System werden aufgrund der konstanten Eingangsdaten für alle Replikationen die gleichen Ergebnisse erzeugt. Daher ist für ein deterministisches System ein Durchlauf ausreichend um die Leistung des Systems zu

bestimmen. Im Gegensatz dazu ist für die Ermittlung des Verhaltens in einem stochastischen System ein größerer Aufwand notwendig. So weisen die Ergebnisdaten, bedingt durch die stochastischen Eingangsgrößen, Zufallsvarianzen auf. Das Ergebnis eines Simulationslaufes ist daher nur eines von vielen möglichen Ergebnissen und somit nicht aussagekräftig, da dies von den realen Werten abweichen kann. Erst durch die Auswertung mehrerer Simulationsläufe ergibt sich ein aussagekräftiges Ergebnis. Zur Ermittlung der Performanz des Systems sind daher mehrere Simulationsläufe notwendig, deren Ergebnisse durch verschiedene statistische Analysemethoden aufbereitet werden. [BCN+14, S.415 f.; DaCh10, S.152 f.; LaKe00, S.496 ff.]

Ergebnisdaten können weiterhin in vorübergehende und andauernde Daten unterteilt werden. Die Unterscheidung ergibt sich aus der Gliederung der Simulationsmodelle anhand ihres Zeithorizontes in terminierende und nicht-terminierende Simulationen. In der terminierenden Simulation determiniert ein natürlicher Endzeitpunkt die Länge des Durchlaufes. Da hier die Verteilung der Ergebnisdaten permanent schwankt, werden die Ergebnisdaten als vorübergehend bezeichnet. Ein Beispiel hierfür ist die Simulation der Abläufe in einer Bank. Durch die festgelegten Öffnungszeiten wird hier der Endzeitpunkt des Simulationslaufes bestimmt. Über den gesamten Simulationslauf schwankt das Verhalten in Abhängigkeit von unterschiedlichen Zeitpunkten, dabei startet und endet das System leer. Die nicht-terminierenden Simulationen laufen hingegen kontinuierlich oder über einen sehr langen Zeitraum. Die Länge eines Simulationslaufes wird dabei durch den Anwender festgesetzt. Die Daten werden in diesem Fall als dauerhaft bezeichnet, da die Outputs entsprechend einer festgelegten Verteilung (Steady-State) variieren. Da die simulierten Systeme in der Realität zu keinem Zeitpunkt leer sind, muss für die nicht-terminierende Simulation eine Warmlauf-Phase definiert werden. Erst wenn das System den Steady-State erreicht hat, ergeben die Ergebnisdaten aussagekräftige Ergebnisse. Zur Feststellung, ob der Steady-State erreicht ist und somit auch zur Definition der Warmlaufphase, existieren verschiedene analytische und graphische Verfahren (vgl. [Robinson 2003, S.143 ff.]). Das Verhalten einer Produktionsanlage stellt exemplarisch eine solche Steady-State Verteilung dar. [BCN+14, S.416 ff.; LaKe00, S.502 ff.; Robinson 2003, S.138 ff.]

Aufgrund von durchgeführten Replikationen sowie verschiedener Messpunkte innerhalb eines Durchlaufes ergeben sich für die Ergebnis-Parameter mehrere Werte. Für den Output  $Y$ , der z.B. den Bestand im Lager angibt, werden die zugehörigen Messwerte für  $n$  Durchläufe mit jeweils  $m$  Werten pro Lauf wie folgt in einer Matrix dargestellt:

$$Y (\textit{Bestand}) = \begin{pmatrix} Y_{11} & \cdots & Y_{1m} \\ \vdots & \ddots & \vdots \\ Y_{n1} & \cdots & Y_{nm} \end{pmatrix}$$

[LaKe00, S.496 f.]

Die Ergebnisdaten werden auf unterschiedliche Arten dargestellt. Während der Simulation werden die dynamischen Zustandsänderungen durch Animationen sowie die Veränderung der Zustandsgröße durch Monitoring in graphischer oder textueller Form aufgezeigt. Die Ergebnisdaten am Ende der Simulation werden mit Hilfe von Statistiken präsentiert. Diese beinhalten neben den durch das Simulationsmodell generierten Daten teilweise auch bereits bestimmte Kenngrößen. Diese werden durch statistische Verfahren ermittelt, welche in den Standardausgaben der Simulatoren integriert sind. [DaCh10, S.154 f.; RiWi98; Vere14, S.17 f.]

### 2.2.4 Vorgehensweisen zur Aufbereitung der Ergebnisdaten

Grundsätzlich können zwei Ziele der Ergebnisdatenauswertung unterschieden werden. Zum einen wird die Leistung eines Systems für eine bestimmte Parameterkombination ermittelt, zum anderen werden verschiedene Systemvarianten miteinander verglichen [LaKe00, S.86].

Die Aufbereitung der Ergebnisdaten stellt nach VDI Richtlinie 3633 Blatt 3 einen der drei Hauptschritte zur Auswertung der Simulationsergebnisse dar [Vere97]. Aufbauend auf den aufbereiteten Rohdaten erfolgt die Interpretation und Bewertung dieser. Innerhalb der Interpretation werden die Ergebnisdaten in Beziehung zu den im Zuge der Simulation getätigten Annahmen gesetzt, sodass eventuell vorliegende Abhängigkeiten analysiert werden können. Die getätigten Annahmen werden durch die Eingangsdaten dargestellt. Bei der Bewertung erfolgt ein Vergleich der verschiedenen Varianten. [RSW08a, S.82; Vere97, S.3]

Im Gegensatz zu den Eingangsdaten, für deren Aufbereitung in der Literatur mehrere Ansätze existieren, ist dies für die Ergebnisdaten nicht der Fall. So liegt in vielen Werken zur Simulation der Schwerpunkt auf der statistischen Aufbereitung der Ergebnisdaten. (vgl. [BCN+14, S.421 ff.; LaKe00, S.505 ff.]

Im Rahmen der Datenaufbereitung nach VDI Richtlinie 3633 Blatt 3 werden in einem ersten Schritt die häufig unzugänglichen oder zumindest unverständlichen Datensätze für den Anwender handhabbar gemacht. Die entsprechend genutzten Aufbereitungsschritte ergeben sich aus der jeweiligen Fragestellung. Nachfolgend werden mögliche Aufbereitungsschritte aufgezählt.

- Selektion bestimmter Daten aus dem Datenbestand
- Sortierung der Daten nach einzelnen Kriterien
- Umrechnung der Daten
- Formatierung der Dateien

Das Sortieren beinhaltet die Umsortierung der Daten von der ursprünglich zeitlichen Sichtweise in eine zur Auswertung kompatiblen Sichtweise. Hierbei wird zwischen auftragsorientierter und elementorientierter Sichtweise unterschieden. Bei der

auftragsorientierten Sichtweise werden die Ergebnisdaten nach dem zugehörigen Auftrag selektiert und sortiert. Die elementorientierte Sichtweise beinhaltet eine Auswertung der Ergebnisdaten aus Sicht der Modellelemente. Hier werden die Daten beispielsweise in Abhängigkeit von der Art der Elemente, an denen sie erhoben wurden differenziert. [Vere97, S.3 ff.]

Im nächsten Schritt folgt die statistische Aufbereitung der Ergebnisdaten. Aufgrund der stochastischen Eingangsgrößen gelten die Ergebnisdaten der Simulation statistisch als unsicher. Durch die Aufbereitung der Ergebnisdaten über die Gesamtheit der Simulationsdurchläufe mit Hilfe von statistischen Verfahren können Ergebnisse ermittelt werden, die unter einer vorgegebenen Wahrscheinlichkeit als korrekt angenommen werden können. Grundsätzlich gilt die Aussage, dass je mehr unabhängige Stichproben erfolgt sind, desto zuverlässiger sind die Aussagen, welche mit Hilfe statistischer Verfahren gewonnen werden. Die unabhängigen Stichproben resultieren aus den Replikationen. Die Unabhängigkeit wird dabei durch die Verwendung unterschiedlicher Initialwerte garantiert. [Vere97, S.9 f.; LaKe00, S.247]

Zur Bestimmung der notwendigen Anzahl von Durchläufen zur Gewährleistung der statistischen Sicherheit sind verschiedene Verfahren vorhanden (vgl. [Robinson 2003, S.152 ff.]). Alternativ zur Durchführung mehrerer Simulationsdurchläufe existiert für nicht-terminierende Simulationen die Möglichkeit, die Ergebnisdaten auf Basis eines langen Simulationsdurchlaufes zu ermitteln. Statt der Replikationen des Systems wird der längere Lauf in mehrere Abschnitte (Batches) unterteilt und deren Mittelwerte als Stichprobenmittelwert verwendet. Dieses Vorgehen wird als Batch-Methode bezeichnet. Dabei wird die Warmlauf-Phase bei der Einteilung der Batches nicht berücksichtigt oder die Simulation wird mit Initial-Werten begonnen, sodass die Warmlauf-Phase entfällt. Die Anzahl sowie die Länge der Simulationsdurchläufe zählen zu den Experimentdaten. [Vere97, S.9; Robinson 2003, S.152 ff.; CuCh13]

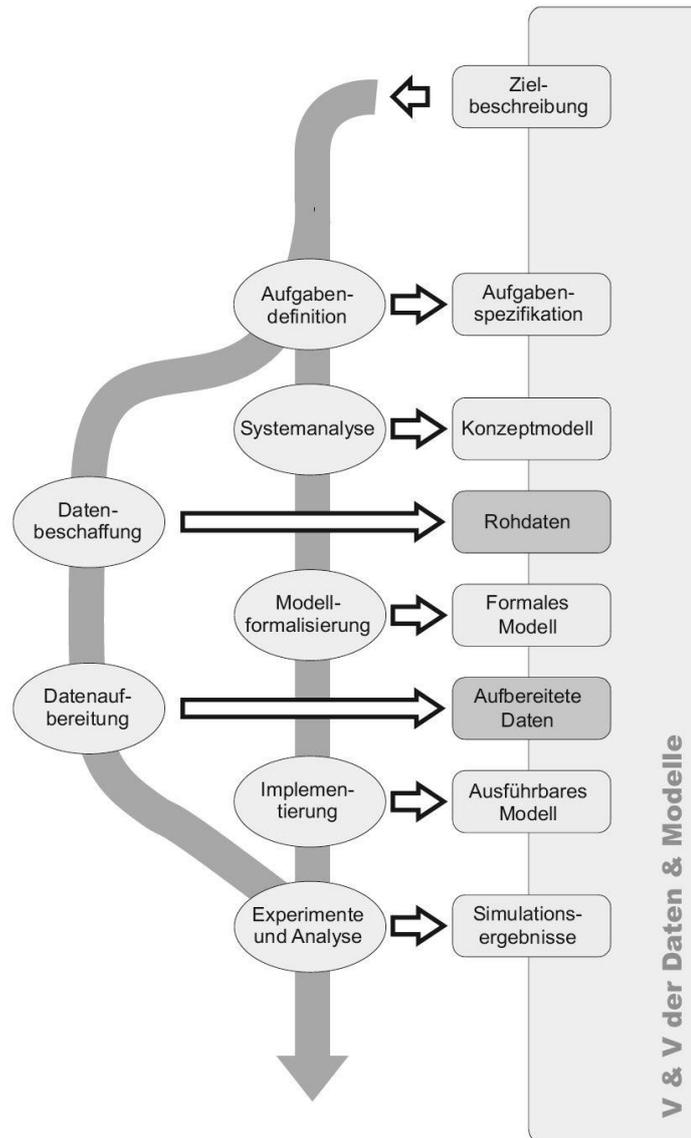
Mit Hilfe statistischer Verfahren können sowohl die Daten eines einzelnen Laufes, als auch die Daten über Replikationen verdichtet werden. Grundlage der statistischen Auswertung stellen statistische Kennwerte dar. Ein oft genutzter statistischer Kennwert ist das arithmetische Mittel. Durch die Berechnung des Konfidenzintervalls wird aufgezeigt, wie genau der gewogene Mittelwert einer Größe geschätzt wurde. Für die verwendeten Daten müssen dabei folgende Voraussetzungen erfüllt sein. Zum einen muss die Unabhängigkeit der Stichproben gewährleistet sein. Dies wird durch die Verwendung von unterschiedlichen Startwerten zur Erzeugung der Eingangs-Zufallsgrößen gewährleistet. Zum anderen müssen die Daten gleiche Mittelwerte und Varianzen aufweisen, was sich durch den Einsatz desselben Simulationsmodells ergibt. Außerdem muss gewährleistet sein, dass die untersuchten Ergebnisse normalverteilt sind. Oftmals wird die statistische Untersuchung durch die Annahme, dass die

Ereignisse einer Normalverteilung folgen, vereinfacht. Die Standardabweichung ist ein Streuungsmaß und gibt Auskunft darüber, wie sehr die Daten vom Mittelwert abweichen. Neben dem arithmetischen Mittelwert wird häufig auch der Median als Mittelwert verwendet. Dieser ist im Gegensatz zum arithmetischen Mittelwert robust gegen Ausreißer, da er den mittleren Wert einer größenmäßig sortierten Datenreihe darstellt. Weitere mögliche Kenngrößen sind Minimum und Maximum. Außerdem ist es möglich die Ergebnisdaten grafisch auszuwerten. Hier werden beispielsweise Histogramme oder Gantt-Diagramme verwendet. Zur näheren Beschreibung der Vorgehensweise in der statistischen Auswertung wird auf die VDI Richtlinie 3633 verwiesen.[Robinson 2003, S.279 ff.; Vere97, S.9 ff.]

### 2.3 Simulationsvorgehensmodell mit V&V

Das primäre Ziel von Verifikation und Validierung ist die Vermeidung von fehlerhaften Ergebnissen der Simulationsstudie aus denen falsche Rückschlüsse auf und gegebenenfalls schwerwiegende Fehlentscheidungen für die reale Welt gezogen werden. Zur detaillierten Definition von V&V vgl. 2.1. Zusammenfassend bezieht sich die Verifikation auf die Prüfung der Richtigkeit und die Validierung auf die Prüfung der Eignung des Modells sowie des jeweiligen Phasenergebnisses. Hierdurch soll ein glaubwürdiges System geschaffen werden (vgl. 2.1). Ein wichtiger Bestandteil und Grundlage für die Verifikation und Validierung ist eine detaillierte Dokumentation. [RSW08a, S.2; RSW09, S.2509; Vere14, S.37; Wenz08, S.26]

Damit Fehler der Simulationsstudie von Anfang an vermieden werden, muss die V&V in den Prozess der Modellbildung integriert werden. Dies haben *Rabe et al.* in ihrem Vorgehensmodell der Simulation mit V&V beschrieben [RSW08a, S.4 ff.]. In der Literatur existieren verschiedene Vorgehensweisen zur Durchführung einer Simulationsstudie. Das Ziel aller Vorgehensmodelle ist es, mit Hilfe eines systematischen Vorgehens zur Modellentwicklung, Fehler so früh wie möglich zu erkennen und somit im Vorhinein vermeiden zu können. Das hier zugrunde liegende Simulationsvorgehensmodell von *Rabe et al.* beschreibt die wichtigsten zu bearbeitenden Phasen innerhalb einer Simulationsstudie [RSW08a, S.4 ff.]. Außerdem ist es verglichen mit anderen Vorgehensmodellen durch die Einführung von Phasenergebnissen geprägt. Die Ergebnisse werden innerhalb der jeweiligen Phasen dokumentiert. In Abbildung 5 ist das Simulationsvorgehensmodell aufgezeigt. Die einzelnen Phasen werden durch Rechtecke mit runden Kanten und die Phasenergebnisse als Ellipsen dargestellt. Innerhalb der aufgezeigten Abfolge der Phasen sind Rücksprünge und Iterationen möglich. Ein weiteres wichtiges Merkmal des Vorgehensmodells besteht in der Trennung von Modell und Daten. Die Handhabung der Daten erfolgt zeitlich parallel und ist eng vernetzt mit der Modellbildung. [Wenz08, S.5 ff.; RSW08a, S.46]



**Abbildung 5: Vorgehensmodell bei der Simulation mit V&V [RSW08a, S.5]**

Wie Abbildung 5 aufgezeigt ist, werden die V&V-Aktivitäten in Bezug auf die Phasenergebnisse durchgeführt. Da im Folgenden der Fokus auf der Verifikation und Validierung innerhalb einer Simulationsstudie liegt, stehen nachfolgend die Phasenergebnisse und nicht die Durchführung der Phasen selbst im Vordergrund.

Ausgangspunkt für das Vorgehen innerhalb einer Simulationsstudie ist die vorliegende Zielbeschreibung. Auf Grundlage dieser wird innerhalb der Aufgabendefinition eine Aufgabenspezifikation verfasst, die nach Auffassung aller Projektbeteiligten eine detaillierte Aufgabenbeschreibung enthält. [RSW08a, S.46 f.]

Im Anschluss an die Konkretisierung der Aufgabenstellung folgt der Prozess der Modellbildung. Dieser umfasst die Phasen Systemanalyse, Modellformalisierung und Implementierung. Das Ergebnis der Systemanalyse ist ein Konzeptmodell, das die Beschreibung der abzubildenden Realität enthält. Die Beschreibung umfasst die

relevanten zu modellierenden Elemente, sowie den jeweils notwendigen Detaillierungsgrad. Dieser wird in Abhängigkeit von der Wichtigkeit des jeweiligen Elements festgelegt. Zusätzlich enthält das Konzeptmodell weiterführende Informationen in Bezug auf Eingaben und Ausgaben sowie Annahmen und Vereinfachungen, welche für das Modell getroffen wurden. Zur Erstellung des Konzeptmodells erfolgt eine Zergliederung des Systems in seine Elemente. Die Zergliederung kann auf zwei verschiedene Weisen durchgeführt werden. Beim Top-down-Ansatz erfolgt die Detaillierung ausgehend vom Ganzen hin zum Detail. Im Gegensatz dazu wird beim Bottom-up-Ansatz vom Detail ausgehend das Ganze schrittweise synthetisiert. Innerhalb der Modellformalisierung entsteht das formale Modell, welches die im Konzeptmodell beschriebenen Elemente und Abläufe konkretisiert und formalisiert. Hiermit soll garantiert werden, dass die im Konzeptmodell beschriebenen Systemelemente mit ihren Wechselwirkungen implementierbar sind. Erst durch die Implementierung wird das Ziel der Modellbildung, die Erstellung eines ausführbaren Modells (Simulationsmodell), erreicht. Hierzu wird das formale Modell unter Beachtung der Eigenschaften des vorliegenden Simulationswerkzeuges umgesetzt. [RSW08a, S.47 f.; RSW08b, S.1720]

Zeitgleich zur Modellbildung findet die Beschaffung und Aufbereitung der Daten statt. Innerhalb der Datenbeschaffung erfolgt die Bereitstellung der für die Simulation notwendigen Daten. Voraussetzung hierfür ist die Kenntnis über Art und Umfang der bereitzustellenden Daten. Diese Angaben ergeben sich aus der Aufgabenspezifikation und dem Konzeptmodell. Ergebnis dieser Phase sind Rohdaten, die direkt aus den Datenquellen stammen und daher hinsichtlich Struktur und Format unverändert sind. Diese werden in der Phase der Datenaufbereitung von Simulationsfachleuten in eine passende Form zum Eingang in das Simulationsmodell umgewandelt. [RSW08a, S.51 f.]

Innerhalb der Phase Experimente und Analyse folgt die Zusammenführung der aufbereiteten Daten und des Simulationsmodells. Das Ergebnis dieser Phase sind die Simulationsergebnisse, die aus der Durchführung von Experimenten resultieren. [RSW08a, S.50 f.]

Die Phasenergebnisse werden nicht nur in Bezug auf sich selbst, sondern auch in Bezug auf die Ergebnisse der anderen Phasen untersucht. Um den damit verbundenen permanent iterativen Prüfprozess der V&V handhaben zu können, bedarf es eines systematischen Ansatzes, durch den die Glaubwürdigkeit unterstützt wird und eine angemessene Dokumentation erfolgt. Nur durch einen solchen Ansatz können die einzelnen V&V-Techniken innerhalb der Simulationsstudie strukturiert angewendet werden. Zur Beschreibung der V&V-Techniken wird auf *Rabe et al.* verwiesen [RSW08a, S.93 ff.]. Das Vorgehensmodell zur V&V für Simulation in Produktion und Logistik nach *Rabe et al.* stellt einen solchen Ansatz zu Strukturierung der einzelnen

V&V-Aktivitäten dar [RSW08a, S.119]. Grundlage hierfür bildet das in Abbildung 5 aufgezeigte Simulationsvorgehensmodell. [RSW08a, S.2; RSW09, S.2509; Wenz08, S.34]

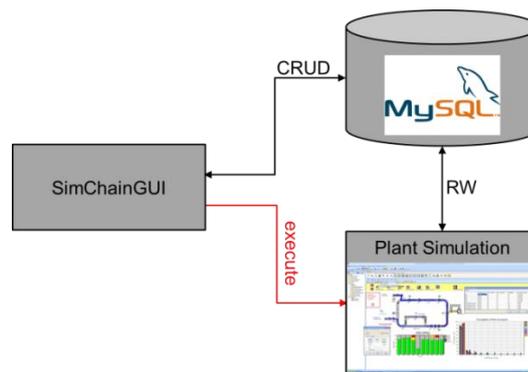
## 2.4 Simulationswerkzeug Sim Chain

Simulationswerkzeuge ermöglichen die softwaretechnische Nachbildung eines Systems in einem Modell. Häufig wird der Begriff Simulator als Synonym für Simulationswerkzeuge verwendet. Nach VDI-Richtlinie 3633 bestehen Simulationswerkzeuge im Normalfall aus den vier Komponenten Simulatorkern, Datenverwaltung, Benutzungsoberfläche und Schnittstelle zu externen Programmen. Der Simulatorkern stellt mit der Bereitstellung und Ausführung der Modellelemente den Mittelpunkt des Simulationswerkzeuges dar. Die Instrumente zur Simulation weisen einen unterschiedlichen Spezialisierungsgrad auf. Reine Programmiersprachen, die universell einsetzbar sind, aber keine anwendungsspezifischen Bausteine enthalten, bieten eine hohe Flexibilität und Allgemeingültigkeit. Jedoch weisen diese nur einen geringen Anwendungsbezug auf. Den höchsten Spezialisierungsgrad weisen Simulatoren auf, welche speziell auf Anwendungsbereiche eines Teilgebietes konzipiert sind. In dieser Arbeit werden zur Abgrenzung von Programmiersprachen lediglich Programmpakete als Simulator bezeichnet. Da diese nur für einen begrenzten Anwendungsfall eingesetzt werden können, aber vorgefertigte Bausteine und Methoden zur Verfügung stellen können, weisen sie im Gegensatz zu reinen Programmiersprachen eine geringe Allgemeingültigkeit und einen höheren Anwendungsbezug auf. Die Anwendung in dieser Arbeit bezieht sich auf die Bereiche Produktion und Logistik [Arno08, S.83 ff.; Eley 2012, S.10 f.; Vere14, S.4 ff.]

Das hier gewählte Simulationswerkzeug SimChain [GuAl04; SimP15a] stellt einen objektorientierten Bausteinkasten dar, der auf dem Simulationssystem Plant Simulation [SimP15b] der Firma Siemens PLM Software basiert. Plant Simulation stellt eine Bibliothek von Objekten zur Abbildung des Material- und Informationsflusses in Produktions- und Logistiksystemen dar. Der Bausteinkasten SimChain erweitert die Bibliothek um Objekte für den Bereich Modellierung und Analyse von Liefernetzwerken. In Analogie zur zuvor beschriebenen Klassifizierung der Simulationswerkzeuge ist Plant Simulation mit einem mittlerem Spezialisierungsgrad einzuordnen und weist somit einen Ausgleich zwischen Einsatzbreite und Implementierungsunterstützung auf. [Eley 2012, S.11; GuAl04, S.6; SimP15b; VöSc10, S.94]

In Abbildung 6 ist die Software-Architektur von SimChain dargestellt. Diese weist die oben benannten vier charakteristischen Komponenten eines Simulationswerkzeuges auf. Mit Hilfe der Benutzeroberfläche SimChainGUI (Graphical User Interface) kann der Anwender ein Modell anlegen und verschiedene Modellszenarien verwalten. Dabei

greift er mittels grundlegender Datenbankoperationen auf eine MySQL-Datenbank zu. Durch den Start des Modells über die Benutzeroberfläche werden die Daten aus der Datenbank über eine Schnittstelle in den Simulatorkern importiert und automatisch zu einem lauffähigen Modell zusammengesetzt. Es erfolgt somit nicht nur die Versorgung des Modells, sondern auch die Konfiguration des Modells zum Beginn der Laufzeit durch die Datenbank. Modelle, die erst zur Laufzeit zusammengesetzt werden, werden als datenbankgetriebene Simulationsmodelle bezeichnet. Die Initialisierung und Parametrierung des Simulationsmodells erfolgt dabei über die Tabellen in der Datenbank, in denen die notwendigen Daten zur Beschreibung des gesamten Simulationsmodells vorgehalten werden. Die Erstellung, Veränderung und Pflege des Simulationsmodells wird daher ausschließlich über die Datenbasis ausgeführt. Die Ergebnisdaten der Simulation werden in die Datenbank zurückgeschrieben und können von dort über die Benutzeroberfläche gelesen werden. Weiterhin bietet SimChain durch Schnittstellen die Möglichkeit zur Verknüpfung mit externen Programmen. So ist beispielsweise die Verknüpfung mit Excel über eine ODBC-Schnittstelle möglich. [Fech15; Gute13, S.1; Witt99, S.22]

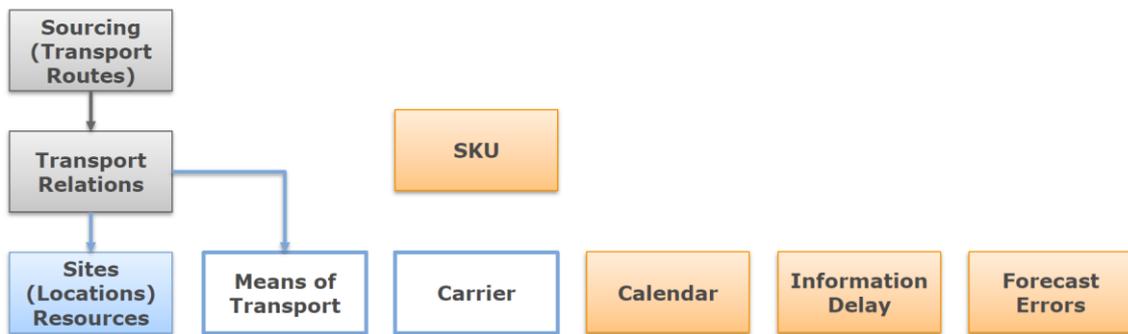


**Abbildung 6: SimChain-Architektur, Ausschnitt aus [Fech15]**

Das Datenbankmanagement-System MySQL verwaltet relationale Datenbanken. Diese zeichnen sich durch ihre tabellarische Darstellungsform aus. Die zweidimensionalen Tabellen weisen eine feste Anzahl an Spalten auf, in denen die Attribute der Relation dargestellt sind. Die Zeilen der Tabellen werden als Datensätze bezeichnet. Um die Beziehung zwischen zwei Relationen abzubilden, werden Fremdschlüssel verwendet. Durch den sogenannten Primärschlüssel, der sich durch ein Attribut oder eine Kombination von Attributen ergibt, wird ein Datensatz eindeutig identifiziert. Wird dieser in einer anderen Tabelle referenziert, so wird er in dieser als Fremdschlüssel bezeichnet. [MBK+12, S.47 ff.; StHa05, S.172 f.]

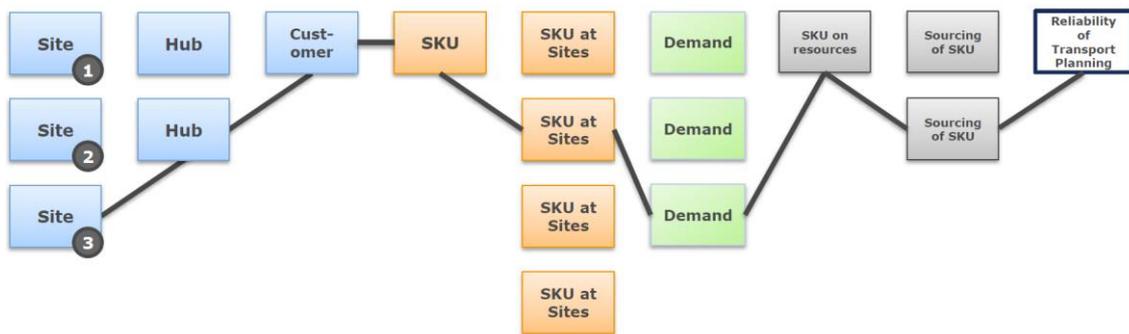
Das Datenmodell von SimChain umfasst alle Informationen zur Abbildung von Lieferketten und Logistiknetzwerken. Grundsätzlich werden die SimChain-Tabellen in Basis- und Konfigurationstabellen unterschieden. In den Basistabellen wird die grundlegende Struktur des Systems beschrieben. Diese enthalten unter anderem die

Informationen, die für alle Szenarien in einer Simulationsstudie identisch sind. Hierzu zählt beispielsweise die geographische Lage der Standorte. In der Tabelle locations werden die customer, sites, hubs und plan supplier angelegt. Abbildung 7 zeigt eine Übersicht über die Basistabellen und ihre Beziehungen zueinander. Die Abhängigkeiten verdeutlichen eine Reihenfolge. Entsprechend dieser Reihenfolge sind die Tabellen zu füllen. So können beispielsweise die Transportrelationen erst angelegt werden, wenn die Tabellen über die Standorte und die Transportmittel gefüllt sind. [FeGu14, S.9 ff.]



**Abbildung 7: SimChain-Basistabellen und ihre Abhängigkeiten [FeGu14, S.10]**

Die eigentliche Erstellung des Simulationsmodells erfolgt über die Befüllung der Konfigurationstabellen. In diesen werden die Basis-Elemente des Simulationsmodells näher spezifiziert. So werden beispielsweise in der Konfigurationstabelle customer die Kunden näher beschrieben und in der Tabelle sku\_demand die Nachfrage für die skus (stock keeping unit) je customer festgelegt. Innerhalb der Konfigurationstabellen können verschiedene Varianten für das in der entsprechenden Tabelle betrachtete Element angelegt werden. Die unterschiedlichen Varianten dienen zur Definition von verschiedenen Szenarien im Simulationsmodell. Innerhalb der Tabelle customer werden die Kunden beispielsweise durch die Zuordnung eines Arbeitskalenders spezifiziert. Zur Abbildung verschiedener Szenarien, in denen beispielsweise die Auswirkungen von unterschiedlichen Arbeitskalendern der customer auf die System-Performanz ermittelt werden sollen, werden in der Tabelle customer für jeden Kunden unterschiedliche Arbeitskalender zugeordnet. Die Beschreibung der jeweiligen Variante erfolgt innerhalb einer zugehörigen Tabelle, mit dem gleichen Tabellennamen erweitert um die Endung \_conf. Diese zugehörigen Tabellen weisen einen einheitlichen Aufbau für jede Konfigurationstabelle auf. Diese bestehen aus zwei Spalten, in denen der Beschreibung der Konfiguration eine entsprechende ID zugeordnet ist. Analog werden für alle Sachverhalte entsprechende Varianten in der Konfigurationstabellen sowie den zugehörigen Tabellen festgelegt. Durch die unterschiedliche Kombination der verschiedenen Konfigurationen innerhalb der einzelnen Konfigurationstabellen können verschieden Szenarien abgebildet werden. In Abbildung 8 ist der Szenarienaufbau durch die Verknüpfung verschiedener Konfigurationen beschrieben.[Gute13, S.84; FeGu14, S.11]



**Abbildung 8: Szenarienaufbau durch die Verknüpfung verschiedener Konfigurationen [FeGu14, S.13]**

Die Ergebnisdaten der Simulation werden in die Statistiktabellen der Datenbank zurückgeschrieben. Diese sind durch die Bezeichnung `stat_` am Anfang des Tabellennamens gekennzeichnet. In der Tabelle `stat_summary_location` sind beispielsweise die Ergebnisse für die verschiedenen locations zusammengefasst. Die Ergebnisdaten stellen sowohl Trace-Daten, in Zählvariablen vorverdichtete Trace-Daten sowie bereits bestimmte Kennzahlen dar. Für die vorverdichteten Ergebnisdaten werden die zugehörigen Trace-Daten nicht zusätzlich gespeichert (vgl. 2.2.3). Der jeweilige Simulationsdurchlauf wird in SimChain unter der `IdRun` sowie das jeweilige Szenario, durch das Attribut `ScenarioMan_Id_Scenario` innerhalb der Ergebnistabelle bestimmt. Letzteres bezieht sich auf die Spalte `Id_Scenario` der Tabelle `Scenario_manager`, in der die verschiedenen Szenarien durch die Verknüpfung der verschiedenen Konfigurationen definiert werden (vgl. Abbildung 8). Die restlichen Spalten enthalten in Abhängigkeit von der jeweiligen Ergebnistabelle verschiedene Ergebnisdaten in Form von Trace-Daten, Zählvariable und Kennzahlen. Außerdem werden die entsprechend zugehörigen Attribute dargestellt. [FeGu14, S.84 ff.]

So werden für die locations beispielsweise die Anzahl der Artikel, welche in die locations eingegangen sind in Form einer Zählvariablen gespeichert. Außerdem wird das Servicelevel im Betrachtungszeitraum als Kennzahl aufgezeigt. In Abbildung 9 ist der zugehörige Ausschnitt aus der Ergebnistabelle `stat_summary_location` aufgezeigt. Die Tabellen können zur Auswertung über die ODBC-Schnittstelle in Excel exportiert werden [Gute13, S.42].

<code>_Id_Scenario</code>	<code>IdRun</code>	<code>Location</code>	<code>Type</code>	<code>NumItems_Inbound</code>	<code>B_ServiceLevel_Inbound</code>
1	1	<code>loc_id_customer_300197</code>	Plain customer	3	0
1	1	<code>loc_id_customer_299761</code>	Plain customer	3	0
1	1	<code>loc_id_customer_373379</code>	Plain customer	1	0
1	1	<code>loc_id_customer_303304</code>	Plain customer	3	0
1	1	<code>loc_id_customer_297093</code>	Plain customer	1	0
1	1	<code>loc_id_customer_296642</code>	Plain customer	1	0
1	1	<code>loc_id_customer_292866</code>	Plain customer	4	0
1	1	<code>loc_id_customer_292415</code>	Plain customer	4	0
1	1	<code>loc_id_customer_180837</code>	Plain customer	5	0
1	1	<code>loc_id_customer_301029</code>	Plain customer	10	40
1	1	<code>loc_id_customer_297875</code>	Plain customer	6	66.67

**Abbildung 9: Screenshot - gefüllte SimChain-Ergebnistabelle `stat_summary_location`**

## 3 DWH-Kennzahlen

Eine DWH-Kennzahl ist eine Kennzahl, die in einem Data Warehouse berechnet wird. Neben ihrem Informationszweck steht in dieser Arbeit der Steuerungsaspekt von Kennzahlen sowie Kennzahlen-Systemen im Vordergrund. Sie stellen durch ihren Vorgabecharakter ein Steuerungsinstrument dar, an dem Entscheidungen und Handlungen ausgerichtet werden [Küpp95, S.320]. Die Ermittlung einer DWH-Kennzahl wird gegenüber einer einfachen Kennzahl durch ihre vielzähligen Vereinheitlichungs-, Filterungs-, und Rechenschritte sowie diversen Aggregationen geprägt. Der Aufbau eines DWH-Systems sowie der vorliegende Datenfluss, beginnend bei den operativen Systemen, bis hin zum Kennzahlbericht, bilden die Architektur eines DWH-Systems. Die eigentliche Berechnung einer Kennzahl erfolgt dabei aus den Daten, die in der Datenbank des DWH-Systems gespeichert sind. Die Datenbank eines DWH-Systems, realisiert ein aus Daten-Würfeln bestehendes, multidimensionales Datenmodell. Neben den für die Ermittlung der Kennzahlen notwendigen Schritten zur Filterung und Aggregation der Daten erfolgt die eigentliche Berechnung von Kennzahlen im multidimensionalen Datenmodell über die Verknüpfung zweier oder mehrerer Datenwürfel.

### 3.1 Definitionen

Zur Schaffung eines einheitlichen Fachterminus werden nachfolgend die wichtigsten Begriffe im Bereich DWH-Kennzahlen erläutert.

#### 3.1.1 Data Warehouse

Es existiert keine Standard-Definition des Data Warehouse. Daher wird zur Beschreibung und Abgrenzung zu operativen Datenbanksystemen nachfolgend die wohl am häufigsten verwendete DWH-Definition von *Inmon* verwendet. [Zeh03]

*“A data warehouse is a subject-oriented, integrated, nonvolatile, and time-variant a collection of data in support of management’s decisions”*  
[Inmo05, S.29]

*W.H. (Bill) Inmon* gilt als der Begründer des DWH-Konzeptes und legt somit die Grundlage vieler weiterer Definitionen. Laut *Inmons* Definition besteht ein Data Warehouse aus einer Sammlung von Daten, die zur Unterstützung von Managemententscheidungen eingesetzt werden kann. Dabei weisen die Daten in einem Data Warehouse vier spezifische Merkmale auf: [Fark11, S.5 f.; JuWi00, S.5]

Themenorientierung: Die Datenstrukturen eines Data Warehouse sind nach den zu analysierenden Themengebieten, welche die wichtigen Sachverhalte in einem Unternehmen widerspiegeln, organisiert. Beispiele für solche Themenschwerpunkte

sind z.B. Produkte und Kunden. Somit haben operative Daten, die ausschließlich für Funktionen und Geschäftsprozesse von Bedeutung sind, im Gegensatz zu operativen Systemen für das Data Warehouse keine Relevanz. [Goek06, S.17 f.]

**Integration:** In einem Data Warehouse werden unternehmensweite Daten und Schemata aus unterschiedlichen inkompatiblen Systemen in einer Datenbank vereinheitlicht gespeichert. [Fark11, S.6; Holt99, S.74]

**Beständigkeit:** Die Daten in einem Data Warehouse sind nicht volatil, das heißt sie werden persistent gespeichert. Im Unterschied zu Transaktionsdaten, welche in einem operativen Datenbanksystem sowohl gelöscht als auch verändert werden können, ist dies für die in das Data Warehouse geladenen Daten nicht möglich. [Holt99, S.75; JuWi00, S.5]

**Zeitliche Varianz:** In operativen Datenbanken spiegeln die Daten den aktuellen Status zum Moment der Eingabe wider. Die Dateneinheiten im Data Warehouse bilden hingegen einen Schnappschuss zu einem bestimmten Zeitpunkt ab. Da die Daten in einem Data Warehouse nicht verändert bzw. gelöscht, sondern stattdessen ergänzt werden, ergibt sich durch die zeitbezogenen abgelegten „Schnappschüsse“ eine Historisierung der Daten. Hierdurch entsteht die Möglichkeit einer Zeitreihenanalyse, beispielsweise für die Abbildung der Entwicklung des Unternehmens über einen bestimmten Zeitraum. [Fark11, S.6; Goek06, S.19 f.; Holt99, S.73]

Obwohl *Inmon* als der Vater des Data Warehouse gilt, wird seine Definition von 1992 von einigen Autoren beispielsweise hinsichtlich ihrer Zweckmäßigkeit, Aussagekraft, und Anwendbarkeit in Theorie und Praxis kritisiert. [BaGü13, S.8; Zeh03]

Aus dieser kritischen Betrachtung heraus leiteten *Bauer und Günzel* eine neue und erweiterte Definition ab, die speziell auf die Analysefunktion des Data Warehouse ausgerichtet ist.

„Ein Data Warehouse ist eine physische Datenbank, die eine integrierte Sicht auf beliebige Daten zu Analysezwecken ermöglicht.“ [BaGü13, S.8]

Diese Definition stellt die Analysefunktion in den Fokus und hebt somit die Aufgabe des Data Warehouse, die für die Analysen des Anwenders notwendigen Daten zu verwalten und aufzubereiten, hervor. Kennzahlen, die innerhalb des Data Warehouse berechnet werden, stellen folglich ein Analysewerkzeug dar. In diesem Sinne ist die Definition von *Bauer und Günzel* passend für diese Arbeit und wird im Folgenden verwendet. [BaGü13, S.59]

In der Literatur wird der Begriff Data Warehouse unterschiedlich verwendet. Es folgt daher eine genaue Begriffsabgrenzung, um eine gemeinsame Begriffswelt mit einheitlichen Fachtermini zu schaffen.

Ein Teil der Autoren, beispielsweise *Jung und Winter* [JuWi00, S.10 f.] charakterisieren nicht ausschließlich die zugrunde liegende Datenbank, welche die für die Analysen der Anwender notwendigen Daten enthält, als Data Warehouse. Stattdessen definieren sie das gesamte System inklusive der zur Datenbank gehörenden Systemkomponenten als Data Warehouse. Zur Abgrenzung wird im Rahmen dieser Begriffsauslegung die Datenbank als Kern-DWH bezeichnet. Der Großteil der Autoren wie beispielsweise *Bauer und Günzel*, *Lehner* sowie *Mucksch und Behme* (vgl. [BaGü13, S.59; Lehn03, S.9 f.; MuBe00, S.14 f.]) verstehen unter dem Begriff Data Warehouse ausschließlich die Datenbank, in der die für die Analyse aufbereiteten Daten verwaltet werden. Die Gesamtheit der Datenbank und der zugehörigen Komponenten wird als DWH-System oder teilweise auch als DWH-Konzept definiert.

In dieser Arbeit wird letzteres, auch in der Literatur häufiger verwendetes Begriffsverständnis, angenommen. Der Begriff DWH-System umfasst somit im Folgenden die Datenbank selbst, sowie alle zugehörigen Komponenten. Das Data Warehouse bezeichnet die Datenbank. Wie in 3.2 noch genauer beschrieben wird, existieren zwei Datenbanken, die Basisdatenbank und das Data Warehouse. *Bauer und Günzel* beschreiben die Bezeichnungen dieser Datenbanken in Abhängigkeit von der jeweiligen Betrachtungsweise. Die als Top-Down bezeichnete Betrachtungsweise wird im Folgenden als gegeben angenommen, da diese den in der Arbeit betrachteten Analyse-Aspekt besonders hervorhebt. Demnach enthält die sogenannte Basisdatenbank eine integrierte, flexible und nicht auswertungsorientierte Datenbasis. Die Datenbank des eigentlichen Data Warehouse ist hingegen auf Analysezwecke fokussiert, dabei wird eine einzelne und von anderen unabhängige Anwenderschicht dargestellt. [BaGü13, S.37]

### 3.1.2 Kennzahl

Der Kennzahlenbegriff hat sich im Laufe der Zeit stark verändert. Mitte der 70er Jahre setzte sich folgende von *Reichmann* geprägte Begriffsdefinition durch:

„Kennzahlen werden als jene Zahlen betrachtet, die quantitativ erfassbare Sachverhalte in konzentrierter Form erfassen.“ [Reic11, S.24]

Mit der Definition benennt *Reichmann* die drei wichtigsten Elemente einer Kennzahl. Der Informationscharakter schafft die Voraussetzung für die Urteilsbildung über einen wichtigen Sachverhalt. Die Quantifizierbarkeit ermöglicht die Messung von Sachverhalten auf einem metrischen Skalenniveau und gewährleistet somit relativ präzise Aussagen. Durch die spezifische Form der Information bieten Kennzahlen die Möglichkeit zur übersichtlichen Darstellung komplizierter Sachverhalte. Betriebswirtschaftliche Kennzahlen stellen somit einen wesentlichen Inhalt von

analyseorientierten Anwendungen dar. Heutzutage wird im deutschsprachigen Raum auch der Begriff Key Performance Indicators (KPI) verwendet. [BaGü13, S.20; Reic11, S.24]

Es werden grundsätzlich absolute und relative Kennzahlen (Verhältniskennzahlen) unterschieden. Die absoluten Kennzahlen lassen sich untergliedern in Bestands- und Bewegungsgrößen. Die Bestandsgrößen spiegeln unternehmerische Zustände, wie beispielsweise den Lagerbestand, zu einem bestimmten Zeitpunkt wider und weisen somit einen statischen Charakter auf. Im Gegensatz dazu erfassen Bewegungsgrößen die kumulierten Ereignisse zwischen zwei Zeitpunkten und verfügen daher über einen dynamischen Charakter. Relative Kennzahlen besitzen gegenüber den absoluten Kennzahlen eine größere Aussagekraft, da durch sie verschiedene Sachverhalte miteinander in Beziehung gesetzt werden. Innerhalb der relativen oder auch Verhältnis-Kennzahlen unterscheidet man Gliederungs-, Beziehungs- und Indexzahlen. Gliederungszahlen geben das Verhältnis von einer Teil- zu einer Gesamtmenge an. Beziehungszahlen setzen verschiedenartige Mengen, die einen logischen/sachlichen Zusammenhang aufweisen, in Beziehung zu einander. Indexzahlen stellen Beziehungen von gleichartigen Größen zu verschiedenen Zeitpunkten dar. [Holt99, S.27; Küpp95, S.317; Prei08, S.12 ff.]

Kennzahlen stellen den Output eines Managementinformationssystems dar. Im Rahmen dieser Arbeit liegt der Fokus auf Kennzahlen, die sich als Output eines DWH-Systems ergeben. [Kütz11, S.42]

### 3.1.3 Kennzahlensystem

Da eine einzelne Kennzahl in der Regel nicht alle notwendigen Informationen beinhaltet, um einen Sachverhalt sinnvoll zu interpretieren, werden für die Analyse z.B. komplexer Logistikstrukturen mehrere Kennzahlen eingesetzt. Um Verwirrungen und widersprüchliche Aussagen bei der Interpretation mehrerer Kennzahlen zu vermeiden, ist es notwendig die Menge an Kennzahlen in eine Ordnung zu bringen. Dies erfolgt durch die integrative Erfassung der Kennzahlen in einem Kennzahlensystem. In diesem sollen die Beziehungen zwischen den einzelnen Kennzahlen abgebildet werden. [Küpp95, S.318; Reic11, S.25 ff.; Wien02, S.36]

Für diese Arbeit ist ein Kennzahlensystem wie folgt definiert:

*„Kennzahlensysteme führen Kennzahlen, die sachlich und sinnvoll zueinander in Beziehung stehen, meist in hierarchischer Form zusammen, die in einer Spitzenkennzahl gipfelt. Sie ermöglichen damit eine zusammenhängende Betrachtung von Funktionen oder Prozessen innerhalb- und außerhalb des Unternehmens.“ [BaGü13, S.20]*

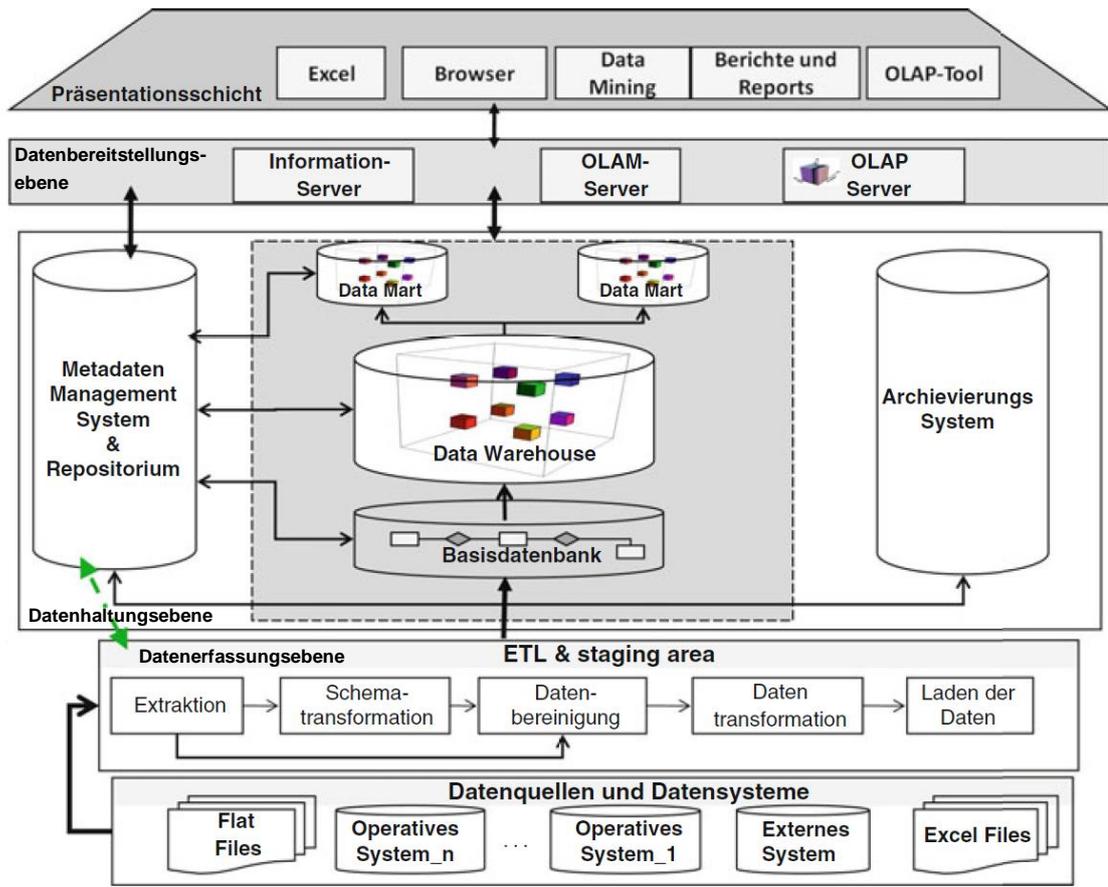
Es existieren verschiedene Beziehungen zwischen den Kennzahlen innerhalb eines Kennzahlensystems. Eine mögliche Klassifikation der Kennzahlensysteme erfolgt nach dem jeweiligen Beziehungstypen. Dementsprechend wird zwischen Ordnungssystemen, Rechensystemen und Mischformen unterschieden. In Kennzahlensystemen, die zu den Ordnungssystemen zählen, weisen die Kennzahlen eine sachlogische Verknüpfung auf. Die Kennzahlen sind in Gruppen angeordnet, die durch betriebswirtschaftliche Sachzusammenhänge miteinander verbunden sind. Innerhalb der Rechensysteme werden Kennzahlen mathematisch miteinander verknüpft. Das Zustandekommen der Spitzenkennzahl wird durch die Zerlegung der zugehörigen Berechnungsformel dargestellt. Durch die Aufschlüsselung der Berechnungsformeln und der damit einhergehenden Zerlegung der einzelnen Kennzahlen ergibt sich eine pyramidenförmige Struktur. Durch sogenannte Hilfskennzahlen werden nicht mathematische Beziehungen innerhalb des Kennzahlensystems überbrückt. Die mathematische Zerlegung hängt vom jeweiligen Kennzahlentypen ab. Kennzahlensysteme, die sowohl Merkmale eines Rechen- als auch eines Ordnungssystems aufweisen, werden als Mischform bezeichnet. [Prei08, S.17 ff.]

In Bezug auf Logistiknetzwerke des Werkstoffhandels liegt in dieser Arbeit der Fokus auf Kennzahlen im Handel und der Logistik. Kennzahlen im Handel beschreiben handelsbetriebliche Tätigkeiten und knüpfen somit an die Produktionsfaktoren Ware, Person, Betriebsmittel und besetzter Raum an. Hierzu zählen beispielsweise der Waren-, Personal-, oder Betriebsmitteleinsatz sowie Kennzahlen zur Kostenkontrolle. Logistikkennzahlen informieren über alle mit den Zielen der Logistik in Verbindung stehenden Sachverhalte, wie der Erreichung von Leistungsmengen, Servicegraden und Logistikkosten. Durch die Erfassung und Aggregation der Ist-Leistungsdaten und die Relativierung zur möglichen Leistung ergeben sich Kennzahlen über die Logistikleistungen. Ein Beispiel hierfür ist das Verhältnis von Ist-Einsatzstunden zu möglichen Einsatzstunden. [Spri15b]

## 3.2 Architektur von DWH-Systemen

Die Architektur eines DWH-Systems spiegelt die Existenz, die Struktur sowie die Beziehungen zwischen den einzelnen Komponenten des Systems wider. In der Literatur werden verschiedene Architekturen in Abhängigkeit von ihrer Ebeneneinteilung unterschieden. [Goek06, S.26 f.; JuWi00, S.10 f.]

Die nachfolgende Beschreibung der Architektur eines DWH-Systems von *Farkisch* (vgl. Abbildung 10) orientiert sich an dem Referenzmodell für die Architektur von DWH-Systemen von *Bauer und Günzel*. Neben der Beschreibung der statischen Architektur wird zusätzlich die dynamische Sicht in Bezug auf den Datentransport innerhalb des DWH-Systems aufgezeigt. [BaGü13, S.36 ff.].



**Abbildung 10: Architektur eines DWH-Systems [Fark11, S.67]**

In der gewählten Architektur werden fünf Schichten beschrieben. Die unterste Schicht und damit die Basis des DWH-Systems bilden die verschiedenen operativen und externen Datenquellen für das Data Warehouse. [JuWi00, S.10]

In der zweiten Ebene, der Datenerfassungsebene, findet der ETL-Prozess statt. Die Bezeichnung ETL resultiert aus den drei Phasen, die dieser Prozess umfasst: Extraktion, Transformation und Laden. Im ersten Schritt werden die für das Data Warehousing relevanten Daten aus den Datenquellen in den sogenannten Arbeitsbereich extrahiert. Der Arbeitsbereich wird dabei nur temporär genutzt, bis die Daten in die Datenbank geladen sind. Nachfolgend beginnt der Transformationsschritt. Die Schemaintegration erfolgt nicht bei jedem Ladevorgang erneut, sondern wird nur bei Hinzunahme von neuen Quellsystemen wiederholt. Im nächsten Schritt folgt die Transformation der heterogenen Daten in die für das Data Warehouse syntaktisch passende Form durch. Dieser Vorgang wird als Homogenisierung bezeichnet. Neben der Vereinheitlichung der Daten hinsichtlich ihres Formates zählt auch die Datenbereinigung von fehlerhaften, redundanten, veralteten oder fehlenden Werten zum Transformationsschritt. Der Transformationsschritt beinhaltet folgende Prozesse, die für jeden Datensatz durchgeführt werden. Die extrahierten Daten werden in ihre Bestandteile zerlegt und hinsichtlich ihrer unterschiedlichen Formate und Schreibweisen in einer Standardform

abgebildet. Letzteres umfasst beispielsweise die Auflösung von Abkürzungen, die Umrechnung von Maßeinheiten sowie Konvertierung von Datumsangaben in ein vorab festgelegtes Standardformat. Die dafür notwendigen Informationen erhält das Analysewerkzeug durch die Verknüpfung über den Kontroll-Datenfluss mit dem Metadatenmanagement bzw. dem Repositorium. Der letzte Schritt umfasst das eigentliche Laden der Daten, wodurch die Basisdatenbank und damit einhergehend auch das Data Warehouse mit Daten befüllt werden. [BaGü13, S.93 ff.; Fark11, S.56 ff.]

Die dritte Ebene wird als Datenhaltungsebene bezeichnet. Diese umfasst neben der Basisdatenbank und dem Data Warehouse die Komponenten zur Verwaltung von Metadaten. Metadaten enthalten betriebswirtschaftliche und für die Datenverarbeitung erforderliche Informationen über den Datenbestand in der Basisdatenbank und dem Data Warehouse. Zudem erhalten sie prozessbezogene Informationen über die Verarbeitung der Daten z.B. im ETL-Prozess. Die Verwaltung der Metadaten findet durch das Metadaten Management System mittels eines Repositorium statt. Die Basisdatenbank meldet die jeweiligen Ladevorgänge an das Metadaten Management System und bekommt im Gegenzug von diesem Steuerungs- und Kontrollinformationen. Weiterhin ist das Archivierungssystem in die Datenhaltungsebene einzuordnen. Dieses zeichnet sich durch eine effiziente Speicherung, Verarbeitung und Archivierung großer Datenmengen aus. Sowohl die Basisdatenbank als auch das Data Warehouse selbst sind physisch und technisch von den Datenquellen getrennt und somit von diesen unabhängig. Der Fokus eines Data Warehouse liegt auf dem Auswertungsaspekt. Die entsprechend benötigte auswertungsorientierte Sicht wird durch ein multidimensionales Schema erreicht, welches aus mehreren abgeleiteten Datenwürfeln mit aggregierten Daten besteht. Auswertungsflexibilität und Konsistenz werden durch die Basisdatenbank gewährt, in der die Detaildaten vorgehalten werden. [BaGü13, S.72 f.; Fark11, S.57 ff.; MuBe00, S.27]

Eine weitere Komponente der Datenhaltungsebene sind die Data Marts. Hierbei handelt es sich um multidimensional modellierte und auswertungsorientierte Datenbanken, die auf dem Data Warehouse aufbauen. Die Grundidee des Data-Mart-Konzeptes ist es eine inhaltlich beschränkte Sichtweise des Unternehmens oder einer Abteilung, beispielsweise in Bezug auf eine Region oder eine Produktgruppe als Teilsicht eines Data Warehouse abzubilden. Das semantische Datenmodell der Data Marts entspricht dem des Data Warehouse. Ziel der Verwendung von Data Marts ist es, die Performanz von Auswertungsvorgängen zu steigern und eine verbesserte Überschaubarkeit der Daten zu bieten. Unterschieden wird zwischen abhängigen und unabhängigen Data Marts. Bei der abhängigen Form bilden die Daten innerhalb der Data Marts eine echte Teilmenge des Data Warehouse ab. Im Gegensatz dazu beziehen die unabhängigen Data

Marts ihre Daten direkt aus den operativen Systemen. [BaGü13, S.61 ff.; MuBe00, S.19 f.]

Die Datenbereitstellungsebene, die vierte Schicht in der Architektur eines DWH-Systems, beinhaltet Informations-, Online Analytical Data Mining (OLAM)- und Online Analytical Processing (OLAP)-Server. Diese dienen innerhalb der Analysephase zur Daten- und Informationsaufbereitung. Innerhalb des Online Analytical Data Minings werden bisher unentdeckte Zusammenhänge zwischen den Daten identifiziert. Das Online Analytical Processing stellt eine interaktive Datenanalyse dar, bei der der Anwender mit verschiedenen Operatoren durch die Daten navigieren kann. Hierdurch können entscheidungsunterstützende Informationen aus dem Data Warehouse gewonnen und aufbereitet werden. Zur Durchführung des Online Analytical Processing über die Endanwendungen und Präsentationswerkzeuge sind multidimensionale Strukturen notwendig. Diese werden durch den OLAP-Server zur Verfügung gestellt.

Den Methoden OLAP und Data Mining liegen verschiedene Analyserichtungen zugrunde. Beim OLAP-Ansatz gilt die Annahme, dass klar ist wonach man sucht. Voraussetzung hierfür ist die genaue Kenntnis über das Ziel der Analyse. Daher müssen die entsprechenden Dimensionen bekannt sein. Die OLAP-Suche wird daher als Top-Down orientiert beschrieben. Im Gegensatz dazu ist der Suchprozess beim Data Mining nach unbekanntem Zusammenhängen datengetrieben. Die Dimensionen sind nicht bekannt. Dieser Prozess wird als Bottom-Up bezeichnet. [BaGü13, S.66 ff.; Gada12, S.288 f.; SBU99; JuWi00, S.10 ff.].

Die oberste Schicht der DWH-System-Architektur, die Präsentationsschicht, umfasst Anwendungsprogramme wie z.B. Excel, Browser, Data Mining, Berichte und Reports sowie OLAP-Tools, mit Hilfe derer der Anwender auf die aufbereiteten Daten in der Datenbereitstellungsebene zugreifen kann. [Fark11, S.66 f.; Goek06, S.26 ff.; SBU99]

Entsprechend der verschiedenen Strukturierungsmöglichkeiten eines DWH-Systems existieren unterschiedliche Umsetzungsalternativen der Komponenten des DWH-Systems. Bei der Festlegung des Data Warehouse und der Analysekomponenten ergeben sich somit unterschiedliche Aufteilungsmöglichkeiten der Funktionalitäten. Hierdurch wird bestimmt, welche Komponente der Kennzahlenberechnung im Datenbanksystem des Data Warehouse und welche Komponente im Analyseprozess durchgeführt werden. Zur genaueren Beschreibung der verschiedenen Gestaltungsmöglichkeiten wird auf *Bauer und Günzel* verwiesen. [BaGü13, S.60 ff.]

### **3.3 Datenmodell innerhalb eines Data Warehouse**

Das DWH ist im Gegensatz zur Basisdatenbank, welches Detaildaten enthält, auf Analysezwecke fokussiert. Zu diesem Zweck werden die für die Analysen notwendigen Daten im Data Warehouse persistent gespeichert und den Analyseprozessen in

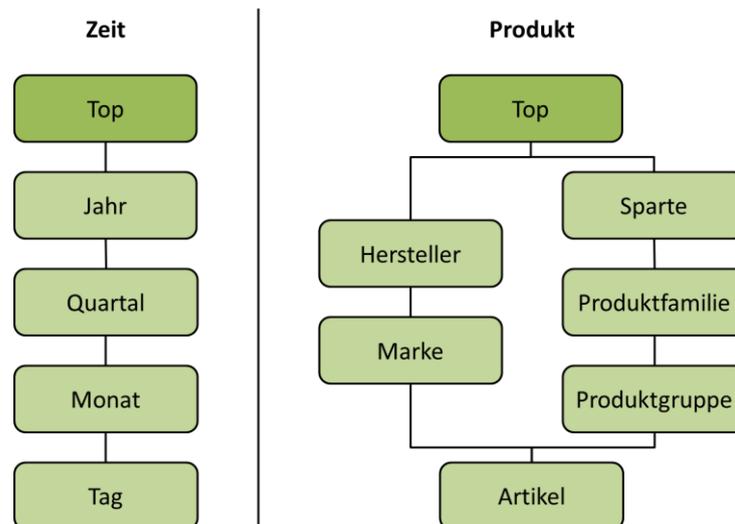
geeigneter Form zur Verfügung gestellt. Bei der Wahl eines geeigneten logischen Schemas hat sich für viele Problemstellungen das multidimensionale Datenmodell als besonders geeignet erwiesen. Die zugrunde liegende Struktur aus multidimensionalen Würfeln ermöglicht eine zielgerichtete Auswertung der Daten nach unterschiedlichen Kriterien und ist somit auf die Unterstützung der Analyse ausgerichtet. Die Umsetzung des logischen Modells des Data Warehouse kann sowohl durch ein multidimensionales als auch ein relationales Speichermodell in physische Datenstrukturen umgesetzt werden. Das weit verbreitete relationale Speichermodell erfordert eine zusätzliche Umsetzung des multidimensionalen Datenmodells in die relationalen Datenstrukturen in Form von Tabellen. Diese Relationen werden durch Primär- und Fremdschlüssel in Beziehung zu einander gesetzt. Beim multidimensionalen Speichermodell erfolgt eine direkte Umsetzung des multidimensionalen Modells in eine physische Struktur. In Bezug auf die OLAP-Produkte spricht man beim relationalen Speichermodell von relationalem OLAP (ROLAP) und beim multidimensionalen Speichermodell von multidimensionalem OLAP (MOLAP). Neben dem Zugriff auf die verwalteten Daten gewährt das Data Warehouse ebenfalls die Möglichkeit zur Verarbeitung dieser. [BaGü13, S.59 und 213 ff.]

Für das multidimensionale Datenmodell existiert keine allgemein anerkannte Formalisierung. Im Folgenden wird die aus Dimensionen und Würfeln bestehende Datenstruktur des multidimensionalen Datenmodells beschrieben. [BaGü13, S.188]

### **Dimension**

Die zu analysierenden Zahlen besitzen beschreibende Attribute, nach denen sie innerhalb des multidimensionalen Datenmodells organisiert werden. Diese Attribute werden als Dimensionen bezeichnet. Mit Hilfe einer ausgewählten Dimension wird eine Auswertungssicht eines Anwendungsbereiches auf die betrachtete Kennzahl festgelegt. Die das multidimensionale Datenmodell prägenden Datenwürfel werden durch Dimensionen aufgespannt und somit wird eine eindeutige und orthogonale Strukturierung des Datenraums erreicht. [BaGü13, S.189 ff.; Baue03, S.26 ff.]

Eine Dimension besteht aus einer Menge von Dimensionselementen wie beispielsweise einzelne Artikel, Standorte oder Tage. Die Gesamtheit aller Dimensionselemente einer Dimension stellen die Blätter des sogenannten Klassifikationsbaumes dar. In diesem ist die hierarchische Struktur einer Dimension abgebildet. Die einzelnen Ebenen innerhalb des Baumes bilden die verschiedenen Verdichtungsstufen einer Dimension ab. Nur die Ausprägungen der niedrigsten, nicht verdichteten, Stufe der Klassifikationshierarchie werden als Dimensionselement verstanden. Die Beziehungen und funktionalen Abhängigkeiten zwischen den Verdichtungsstufen, auch Klassifikationsstufen genannt, werden durch das Klassifikationsschema beschrieben.



**Abbildung 11: Klassifikationsschemata – links: einfacher Aufbau, Dimension Zeit; rechts: parallel Aufbau, Dimension Produkt (i.A.a [Baue03, S.27; Holt99, S.137] )**

In Abbildung 11 sind Klassifikationsschemata für die Dimensionen Zeit und Produkt dargestellt. Die Ausprägungen auf den Verdichtungsstufen werden als Klassifikationsknoten (KK) bezeichnet. Innerhalb der Klassifikationshierarchie sind die Klassifikationsknoten aufgrund ihrer funktionalen Abhängigkeit untereinander angeordnet. [BaGü13, S.189 ff.; Baue03, S.26 ff.]

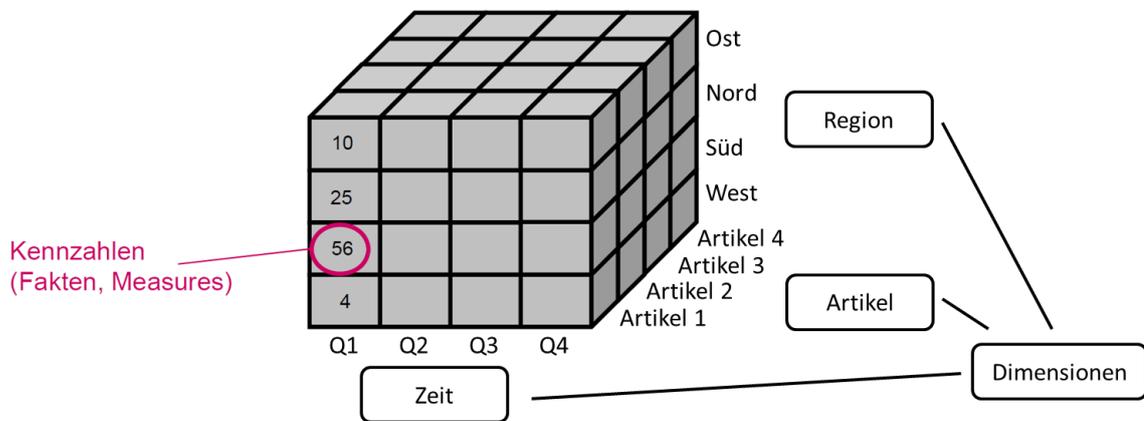
Dimensionen können zwei verschiedene Klassifikationshierarchien beinhalten. Es wird zwischen einem einfachen und einem parallelen Aufbau unterschieden (vgl. Abbildung 11). Die einfache Klassifikationshierarchie weist eine hierarchische Beziehung zwischen den Klassifikationsstufen auf. So existiert innerhalb der entsprechenden Dimension nur ein Pfad vom untersten Dimensionselement zum so genannten TOP-Gesamtknoten. Hier liegt der Unterschied zur parallelen Klassifikationshierarchie. In dieser sind mehrere mögliche Pfade vorhanden. Als Beispiel für eine einfache Klassifikationshierarchie wird in Abbildung 11 die Zeitdimension beschrieben. Der einzige mögliche Weg lautet: Tag → Monat → Quartal → Jahr → Top. Im Gegensatz dazu stellt die Produktdimension mit zwei möglichen verschiedenen Pfaden vom Artikel zum TOP-Knoten eine parallele Klassifikationshierarchie dar. [Baue03, S.26 ff.; BaGü13, S.189 ff.]

### Würfel

Die Daten werden als Elemente eines mehrdimensionalen Würfels angeordnet und durch die auf den Kanten des Datenwürfels aufliegenden Dimensionen beschrieben. Abbildung 12 stellt einen Datenwürfel dar, dem man beispielsweise entnehmen kann, welcher Artikel, in einem bestimmten Quartal in einer bestimmten Region verkauft wurde. Die Dimensionselemente und Klassifikationsknoten stellen dabei die Würfelkoordinaten dar. Die Anzahl der zu qualifizierenden Dimensionen der Daten

bestimmt die Dimensionalität des Datenwürfels. Die Kennzahlen selbst sind in den Zellen des Würfels enthalten. [Kimb96, S.xxi]

Entsprechend der interaktiven Datenanalyse mit Hilfe des Online Analytical Processing wird der Datenwürfel in der Literatur oftmals auch als OLAP-Würfel bezeichnet.



**Abbildung 12: multidimensionaler Datenwürfel**

### Operatoren

Im multidimensionalen Datenmodell sind verschiedene Operatoren verankert, welche die Möglichkeit bieten durch die multidimensionalen Datenstrukturen zu navigieren. Bei der Pivotierung erfolgt eine Drehung des Würfels. Hierbei werden die Dimensionen des Würfels vertauscht. Dies ermöglicht dem Anwender die Daten aus verschiedenen Perspektiven zu betrachten. Mit Hilfe der Operationen Roll-Up, Drill-Down und Drill-Across ist ein Wechsel zwischen den Verdichtungsstufen der Dimensionen möglich. Beim Roll-Up handelt es sich um eine Aggregation der Daten entlang der Klassifikationshierarchie. So kann für die in Abbildung 11 beschriebene Zeitdimension von Tages- bis hin zu Jahreswerte navigiert werden. Die Drill-down Funktion beschreibt die komplementäre Operation. Dementsprechend erfolgt hier die Navigation von den verdichteten Daten zurück zu den Detaildaten. Beim Drill-Across erfolgt kein Wechsel der Dimensionsebene, sondern zwischen den Ausprägungen der Dimensionselemente. So wird beispielsweise eine andere Region oder ein anderer Monat betrachtet. Beim Slicing wird anschaulich eine Scheibe aus dem Datenwürfel ausgeschnitten. Am Beispiel des in Abbildung 12 dargestellten Datenwürfel werden beispielsweise nicht alle vier Quartale, sondern nur ein Quartal betrachtet. Hierbei wird eine Ebene senkrecht zur Zeit-Achse herausgeschnitten. Erfolgen mehrere Slicing-Vorgänge gleichzeitig, so ergibt sich ein Teilwürfel, der eine konkrete Kombination von Klassifikationsknoten aufzeigt. Dieser Prozess wird als Dicing bezeichnet. [Baue03, S.27; BaGü13, S.113; Holt99, S.44 f.; Kimb96, S.38]

### 3.4 Kennzahlen und Kennzahlensysteme im Data Warehouse

Da betriebswirtschaftliche Kennzahlen einen wesentlichen Inhalt von analyseorientierten Anwendungen abbilden, stellen diese auch den eigentlichen Untersuchungsgegenstand des DWH-Systems dar. Die Kennzahlen werden durch dimensionale Informationen definiert, welche auf den Dimensionsachsen der Datenwürfel des multidimensionalen Datenmodells aufgetragen sind. Die Kennzahlen bilden dabei den Inhalt der Datenzellen ab. Dies bedeutet, dass die Dimensionen den qualifizierenden und die Kennzahlen den quantifizierenden Anteil des multidimensionalen Datenmodells repräsentieren. Aufgrund ihrer quantifizierenden Informationen besitzen Kennzahlen einen numerischen Datentyp. [Baue03, S.79 ff.]

Kennzahlen werden in Basiskennzahlen und abgeleitete Kennzahlen differenziert. Basiskennzahlen werden aus den Datenquellen extrahiert und als Fakt bezeichnet. Eine allgemeine Kennzahl wird unter Verwendung von Aggregationsfunktionen wie beispielsweise  $\text{Sum}()$ ,  $\text{Avg}()$ ,  $\text{Max}()$ ,  $\text{Min}()$ ,  $\text{Count}()$ ,  $\text{Variance}()$ ,  $\text{Covariance}()$ , arithmetischer Operationen (+, -, \*, /, mod) oder ordnungsbasierter Funktionen, wie beispielsweise Kumulation, aus den Basiskennzahlen abgeleitet. Daher wird diese Kennzahl auch als abgeleitete Kennzahl bezeichnet. Durch die Aggregatfunktion erfolgt eine Verdichtung eines Datenwürfels selbst. Die arithmetischen Operationen bieten die Möglichkeit, verschiedene Datenwürfel miteinander zu verbinden. Die zu ermittelnde Kennzahl steuert die Ausprägung der Funktionen. Dieser Vorgang der Ableitung wird durch die Berechnungsvorschrift  $f()$  beschrieben. Beispiele für eine abgeleitete Kennzahl sind Umsatz, Gewinn, Verlust oder Deckungsbeitrag. [Fark11, S.19 f.; Lehn03, S.67 ff.; SSH10, S.628] Ein Fakt und eine abgeleitete Kennzahl werden daher wie folgt definiert.

*„Ein Fakt  $F$  eines dimensionalen Datenschemas wird durch das Tupel  $F = (G, \text{SumTyp})$  spezifiziert, wobei  $\text{SumTyp}$  der Summationstyp und  $G$  die Granularität ist.“ [Lehn03, S.67]*

*„Eine abgeleitete Kennzahl  $M$  ist definiert durch*

- eine Granularität  $G$ ,*
- eine Berechnungsvorschrift  $f()$  über Fakten sowie*
- einem Summationstyp  $\text{SumTyp}$ .*

*Die Berechnung erfolgt über einer nichtleeren Teilmenge der im Schema existierenden Fakten. Das Schema einer Kennzahl ist also eine Drei-Tupel wie folgt:  $M = (G, f(F_1, \dots, F_k), \text{SumTyp})$ " [SSH10, S.628]*

Die Granularität gibt das Aggregationsniveau an, in dem die Kennzahl abgebildet wird. Die Wahl der Aggregationsoperationen, zur Verdichtung eines Würfels, wird durch den vorliegenden Summationstypen der Kennzahlen bestimmt. Voraussetzung für die Anwendung der Aggregationsoperationen ist die Aggregierbarkeit der Kennzahlen. Diese erfordert Vollständigkeit, Disjunktheit und Typverträglichkeit der Fakten. Durch die Vollständigkeit wird garantiert, dass jeder Klassifikationsknoten in der darüber liegenden Ebene mindestens einem Vaterknoten zugeordnet ist. Durch die Forderung nach Disjunktheit wird dies auf maximal einen Vaterknoten begrenzt. Da somit jeder Klassifikationsknoten genau einem Vaterknoten zugeordnet ist, ist die Summe aller Knoten einer Verdichtungsstufe gleich der Gesamtsumme. Die dritte Bedingung für die Aggregierbarkeit von Fakten und abgeleiteten Kennzahlen ist die Typverträglichkeit. Diese besagt, dass die Aggregationsoperation mit dem Summationstypen kompatibel ist. Es werden folgende Summationstypen unterschieden. [Baue03, S.79 ff.]

**FLOW:** Die Kennzahl beinhaltet eine Anzahl von Ereignissen in einem Zeitintervall und ist somit über alle Dimensionen aggregierbar. Der Umsatz stellt beispielsweise eine Kennzahl des Summationstypen FLOW dar.

**STOCK:** Die Kennzahl spiegelt Werte wider, die nur zu einem bestimmten Zeitpunkt gültig sind. Diese können mit Ausnahme der temporalen Dimension aggregiert werden. Beispielsweise ist eine zeitliche Summation über die Kennzahl Lagerbestände nicht erlaubt.

**VALUE-PER-UNIT:** Die Kennzahl beschreibt relative Werte, die sich auf eine bestimmte Bezugsgröße beziehen. Auch hier ist die Summation der Werte nicht zulässig. Es sind nur die Operationen MIN, MAX und AVG erlaubt. Der Preis pro Stück repräsentiert eine Kennzahl des Summationstypen VALUE-PER-UNIT dar.[Baue03, S.79 ff.; Fark11, S.20]

Neben der Berechnungsmöglichkeit einer abgeleiteten Kennzahl durch die Verdichtung innerhalb des jeweiligen Datenwürfels kann dies auch durch den Verbund zweier Datenwürfel erfolgen. Der Verbund ergibt sich durch die Anwendung arithmetischer Operationen (+, -, \*, /, mod). Grundsätzlich ist eine Verbindung von zwei Datenwürfeln nur möglich, wenn diese verbundverträglich sind. Die Verbundverträglichkeit zweier Datenwürfel ist gegeben, wenn diese die gleiche Granularität und Bereichseinschränkungen sowie disjunkte Kennzahlenmengen aufweisen. Disjunkt bedeutet in diesem Zusammenhang, dass es sich um getrennte, sich nicht überschneidende Kennzahlenmengen handelt. Wie im vorherigen Unterkapitel beschrieben, gibt die Granularität die Verdichtungsstufe auf den Dimensionsachsen des Datenwürfels an. Ist die Verbundverträglichkeit der beiden Würfel gegeben, so kann der Verbund durch eine einfache zellenweise Verknüpfung erfolgen. Für den Verbund von zwei nicht verbundverträglichen Datenwürfeln bedarf es eines zusätzlichen Schrittes.

Vor der zellenweisen Verknüpfung müssen die Dimensionen der Würfel z.B. durch Deaggregation auf die gleiche Granularität ausgerichtet werden. [BaGü13, S.194]

Kennzahlensysteme stellen eine Zusammenführung von Kennzahlen zu einer Spitzenkennzahl dar (vgl. 3.1.3). Dies bedeutet, dass ein Kennzahlensystem im Gegensatz zu einer abgeleiteten Kennzahl durch den mehrfachen Verbund zweier Datenwürfel gekennzeichnet ist. Da es sich bei dem Verbund um mathematische Operationen handelt, stellen Kennzahlensysteme im DWH-System ein Rechensystem dar (vgl. 3.1.3). Die im Rahmen dieser Arbeit thematisierten komplexen DWH-Kennzahlen werden als Spitzenkennzahl eines Kennzahlensystems eingeordnet. Die Berechnungsvorschrift einer komplexen DWH-Kennzahl weist daher eine hohe Komplexität auf.

### **Dokumentation**

Einhergehend mit der sich stetig weiterentwickelnden Technik und den daraus resultierenden komplexen Produkten nimmt die technische Dokumentation an Bedeutung zu. Da dieser Gedanke noch nicht in allen Unternehmen den entsprechenden Stellenwert eingenommen hat, liegt oftmals keine oder nur eine unzureichende Dokumentation vor. Neben der Forderung nach einer angemessenen Dokumentation technischer Sachverhalte stellt *Kütz* den Anspruch an eine detaillierte und umfassende Dokumentation von Kennzahlen und Kennzahlensystemen. In der Praxis erfolgt diese oftmals in Form eines Kennzahlensteckbriefes. In diesem sind alle notwendigen Informationen zur Beschreibung der Kennzahl sowie Angaben zur Datenermittlung, Datenaufbereitung und Präsentation der Kennzahl enthalten. Zur Beschreibung der Kennzahl werden z.B. Ziel- und Sollwerte und der Gültigkeitsbereich angegeben. Zur Datenermittlung werden unter anderem Datenquellen und zur Datenaufbereitung der Berechnungsweg dokumentiert. Der Steckbrief für Kennzahlensysteme unterscheidet sich unter anderem durch die Ergänzung der zugehörigen Kennzahlen von der Dokumentation der Kennzahlen. [Koth11, S.1; Kütz11, S.44]

## 4 Aufbereitung von Simulationsdaten zur glaubwürdigen Messbarkeit von DWH-Kennzahlen

In diesem Kapitel wird die im Rahmen dieser Arbeit entwickelte Vorgehensweise beschrieben. Diese umfasst die Aufbereitung von Eingangs- und Ergebnisdaten einer ereignisdiskreten Simulation, sodass komplexe DWH-Kennzahlen an den Ergebnisdaten der Simulation glaubwürdig gemessen werden können (vgl. 1). Die Vorgehensweise ist speziell auf die Simulationssoftware SimChain ausgerichtet (vgl. 2.4).

In der Literatur wird die Kombination einer ereignisdiskreten Simulation mit der DWH-Technologie bereits beschrieben. *Ehmke et al.* und *Vasilakis et al.* stellen beispielsweise den Einsatz von DWH-Technologie für die interaktive und flexible Analyse der detaillierten Ergebnisdaten einer ereignisdiskreten Simulation logistischer Systeme dar. [EGM+11; VEC04]. Der Bezug zur glaubwürdigen Messbarkeit von komplexen DWH-Kennzahlen an einem ereignisdiskreten Simulationsmodell wird in der Literatur jedoch bisher nicht berücksichtigt. *Dross* und *Rabe* beschreiben in ihrem Paper [DrRa14] eine Situation, in der die glaubwürdige Messbarkeit von komplexen DWH-Kennzahlen an einem ereignisdiskreten Simulationsmodell von großer Bedeutung ist. Hierbei handelt es sich um ein logistisches Assistenzsystem, das unter anderem auf einem Simulationsmodell basiert. Das logistische Assistenzsystem empfängt die auf Grundlage der Ergebnisdaten der Simulation gemessenen DWH-Kennzahlen. Das Ziel des Systems ist es mögliche Maßnahmen in Bezug auf die DWH-Kennzahlen im Simulationsmodell zu bewerten.

Der Grundgedanke der entwickelten Vorgehensweise besteht darin, dass die Kennzahlenberechnung auf den Ergebnisdaten der Simulation innerhalb eines sogenannten Schatten- DWH-Systems durchgeführt wird. Dieses beinhaltet eine Kopie des gesamten Datenmodells sowie der Kennzahlenlogik des DWH-Systems, in dem die realen Kennzahlen berechnet werden. Dies wird zur begrifflichen Abgrenzung zum Schatten-DWH-System als bestehendes DWH-System bezeichnet. In Bezug auf das multidimensionale Datenmodell bedeutet dies anschaulich, dass nur die Dimensionen der Datenwürfel in das Schatten-DWH-System kopiert werden. Die einzelnen Zellen der Datenwürfel bleiben zunächst leer, da in diese die Ergebnisdaten der Simulation geladen werden. Die Besonderheit des Schatten DWH-Systems besteht in der Verknüpfung mit dem bestehenden DWH-System. Hierdurch werden die Strukturen des kopierten Schatten-DWH-Systems fortlaufend an mögliche Veränderungen im bestehenden DWH-System angepasst. [DrRa14]

Das grundlegende Vorgehen innerhalb der entwickelten Vorgehensweise ist, dass zunächst die notwendigen Ergebnisdaten der Simulation ermittelt werden. Diese sind zur Messung der Kennzahl im DWH-System unabdingbar. Ausgehend von den notwendigen Ergebnisdaten werden die entsprechenden notwendigen Eingangsdaten für die Simulation bestimmt, welche für die Erzeugung der notwendigen Ergebnisdaten erforderlich sind.

Grundsätzlich werden die Daten der operativen Systeme auf einer detaillierten Ebene in ein DWH-System geladen (vgl. 3.2). Die erzeugten Ergebnisdaten der Simulation weisen hingegen teilweise bereits verdichtete Werte in Form von Zählvariablen und Kennzahlen auf (vgl. 2.4). Diese werden zur Messung der DWH-Kennzahlen im Schatten-DWH-System genutzt. So werden die bereits verdichteten Ergebnisdaten an der entsprechenden Stelle der Kennzahlenberechnung in das Schatten-DWH-System integriert. Hierdurch entfallen die vorherigen Transformations- und Berechnungsschritte innerhalb des Schatten-DWH-Systems. Beispielsweise werden die für den Umsatz relevanten Daten aus den operativen Systemen auf Verkaufsbelebene in das bestehende DWH-System geladen. Die Ergebnisdaten der Simulation stellen hingegen bereits den Umsatz pro Kunde und Monat zur Verfügung. Diese werden an der Stelle ins Schatten-DWH-System integriert, an der im bestehenden DWH-System der Umsatz pro Kunde und Monat auf Grundlage der Werte auf Verkaufsbelebene berechnet wird.

Die entwickelte Vorgehensweise setzt sich aus zwei einzelnen Vorgehensweisen zusammen. Zuerst wird das Vorgehen zur Aufbereitung der Eingangsdaten (vgl. 4.1) und im zweiten Teil das Vorgehen zur Aufbereitung der Ergebnisdaten (vgl. 4.2) der ereignisdiskreten Simulation beschrieben. Dabei stehen die beiden Vorgehensweisen insofern in Verbindung zueinander, als dass zur Aufbereitung der Ergebnisdaten Informationen aus der Aufbereitungsphase der Eingangsdaten benötigt werden. Zur Anwendung der entwickelten Vorgehensweise innerhalb einer Simulationsstudie wird die entwickelte Vorgehensweise in das Simulationsvorgehensmodell zur kontinuierlichen V&V von *Rabe et al.* eingeordnet [RSW08a, S.5]. Anschließend werden in Kapitel 4.4 erste Anhaltspunkte zur Umsetzung der Vorgehensweise in einen automatisierten Prozess aufgezeigt. Die beschriebenen V&V-Techniken wurden *Rabe et al.* entnommen [RSW08a, S.113].

## 4.1 Vorgehensweise zur Aufbereitung von Eingangsdaten

Die entwickelte Vorgehensweise zur Aufbereitung von Eingangsdaten basiert auf dem in Abschnitt 2.2.2.2 vorgestellten IGP von *Bernhard et al.*. Eine weitere Grundlage bildet die Methode zum Eingangsdaten-Management von *Skoogh und Johansson* (vgl. 2.2.2.1). Der IGP wurde aufgrund der im Folgenden aufgezeigten Punkte gewählt. Der Hauptgrund besteht in dem detaillierten Überblick über die Vorgehensweise zur

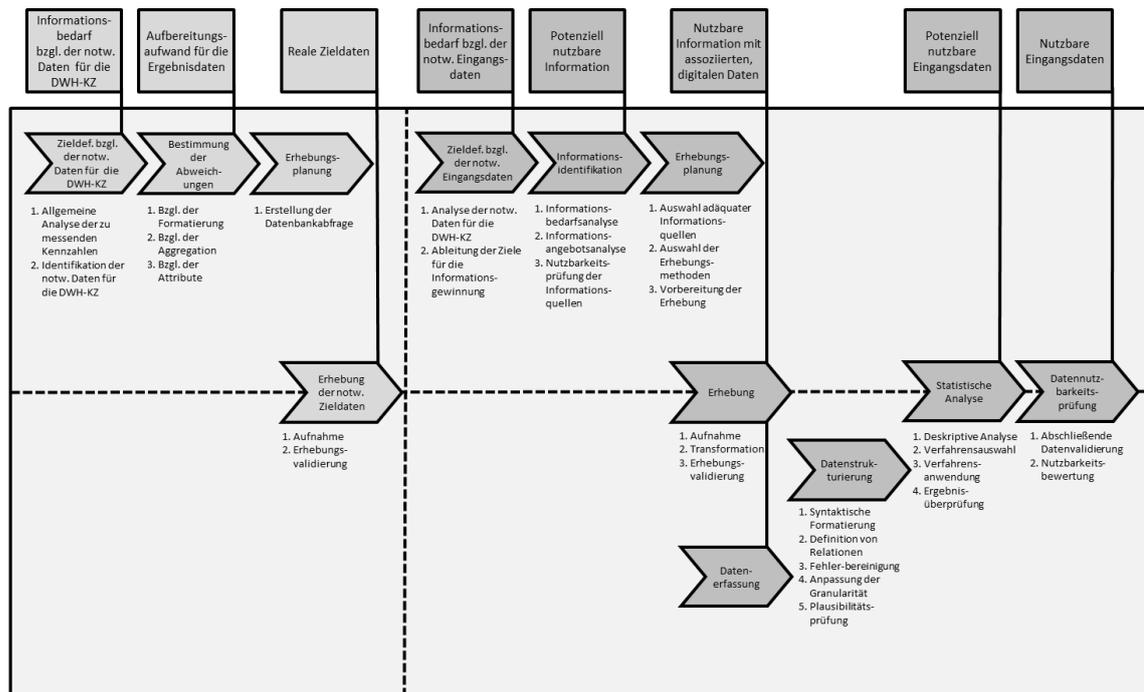
Aufbereitung von Eingangsdaten, der durch den IGP zur Verfügung gestellt wird. Im Gegensatz zur Methode des Eingangsdaten-Managements von *Skoogh und Johansson* erfolgt im IGP eine Differenzierung von Daten und Informationen innerhalb der Prozesse. Hierdurch ist es möglich das Vorgehen in einer detaillierten Übersicht, angefangen von der Zieldefinition über die Ermittlung des Informationsbedarfs bis hin zu den nutzbaren Eingangsdaten, aufzuzeigen. Weiterhin ist der IGP durch eine durchgängige V&V aller Zwischenergebnisse am Ende der einzelnen Prozessschritte gekennzeichnet. Dies bildet eine wichtige Grundlage für den Aspekt der Glaubwürdigkeit der Kennzahlenmessung innerhalb der entwickelten Vorgehensweise ab. Ein weiterer Grund für die Wahl der Vorgehensweise von *Bernhard et al.* ist die Tatsache, dass der IGP in der Literatur bereits exemplarisch in das Simulationsvorgehensmodell zur kontinuierlichen V&V von *Rabe et al.* eingebettet wurde (vgl. Abbildung 3). Ebenso wird die in dieser Arbeit entwickelte Vorgehensweise in dieses Modell integriert (vgl. 4.3). Demnach bietet der IGP einige Anhaltspunkte für die Einbettung. Weiterhin zeigt der IGP ein allgemeingültiges Vorgehen für modellgestützte Analyseverfahren für große Netze der Logistik auf (vgl. 2.2.2.2).

Für die entwickelte Vorgehensweise werden die einzelnen aufgeführten Schritte des IGP in zeitlicher Abfolge (vgl. Abbildung 4) an die gewünschte glaubwürdige Messbarkeit von komplexen DWH-Kennzahlen angepasst. Ebenso wird das Vorgehensmodell um zusätzlich notwendige Schritte erweitert. In Abbildung 13 ist der um den Aspekt der glaubwürdigen Messbarkeit von komplexen DWH-Kennzahlen ergänzte IGP aufgezeigt. Die Einordnung der einzelnen Prozessschritte in Abhängigkeit von ihrem inhaltlichen Fokus in die Daten- oder Informationsebene sowie alle anderen Darstellungsformen innerhalb des IGP (vgl. 2.2.2.2) bleiben erhalten.

In den folgenden Abschnitten werden die einzelnen Schritte des erweiterten Vorgehensmodells erläutert. Dabei liegt der Fokus der Beschreibung für den Teil des IGP auf der Spezifizierung hinsichtlich der Messung von DWH-Kennzahlen. Weiterhin werden die ergänzten Schritte sowie die V&V-Schritte in der gesamten Vorgehensweise detailliert beschrieben. Zur ausführlichen Beschreibung des IGP vgl. Kapitel 2.2.2.2

Im IGP werden die nutzbaren Eingangsdaten für eine gegebene Aufgabenstellung bestimmt. In der entwickelten Vorgehensweise entspricht diese den notwendigen Ergebnisdaten zur Berechnung der DWH-Kennzahlen. Zur Ermittlung der notwendigen Ergebnisdaten wird der IGP um vier Schritte erweitert. Diese sind in Abbildung 13 durch die senkrecht gestrichelte Linie vom anschließenden originalen IGP anschaulich abgegrenzt. Auf Grundlage der notwendigen Ergebnisdaten erfolgt anschließend innerhalb des IGP die Ermittlung der nutzbaren Eingangsdaten für die Simulation zur

Generierung der notwendigen Ergebnisdaten. Die Schritte zur Identifikation der notwendigen Ergebnisdaten basieren ebenfalls auf der Struktur des IGPs.



**Abbildung 13: Erweiterung des IGPs zur Vorgehensweise zur Aufbereitung von Eingangsdaten zur glaubwürdigen Messbarkeit von DWH-Kennzahlen ( i.A.a. [BJH+07, S.8])**

Insgesamt beinhaltet die entwickelte Vorgehensweise für die Eingangsdaten zwei verschiedene Erhebungsschritte. Im ersten Erhebungsschritt werden die zu den benötigten Zieldaten für den Eingang ins Schatten DWH-System zugehörigen realen Daten aus dem DWH-System abgezogen. Diese dienen im weiteren Verlauf der Vorgehensweise zur V&V der simulierten Ergebnisdaten. Der zweite Erhebungsvorgang umfasst die Erhebung der notwendigen Eingangsdaten für die Simulation, sodass die notwendigen Ergebnisdaten generiert werden können.

#### 4.1.1 Zieldefinition bezüglich der notwendigen Daten zur Messung der DWH-Kennzahlen

Der erste Prozessschritt der Vorgehensweise zur Aufbereitung von Eingangsdaten ist die Zieldefinition bezüglich der notwendigen Daten zur Messung der DWH-Kennzahlen. Diese beinhaltet die Bestimmung der Ergebnisdaten und Zieldaten, die zur Berechnung der DWH-Kennzahlen benötigt werden. Hierzu erfolgt zunächst eine allgemeine Analyse der zu messenden Kennzahlen, in der der Fokus auf der Logik der Kennzahlen liegt. Der technische Aspekt der Implementierung im DWH-System wird dabei zunächst vernachlässigt. Aufbauend auf den aus der Analyse gewonnen Erkenntnissen über die Kennzahlen werden die Kennzahlenberechnungen im DWH-System untersucht und die notwendigen Daten bestimmt. Das Ergebnis dieses

Prozessschrittes ist der Informationsbedarf bezüglich der notwendigen Daten zur Messung der DWH-Kennzahl. Dieser beinhaltet die notwendigen Ergebnisdaten der Simulation sowie die notwendigen Zieldaten zum Eingang ins Schatten-DWH-System.

#### **4.1.1.1 Allgemeine Analyse der zu messenden Kennzahlen**

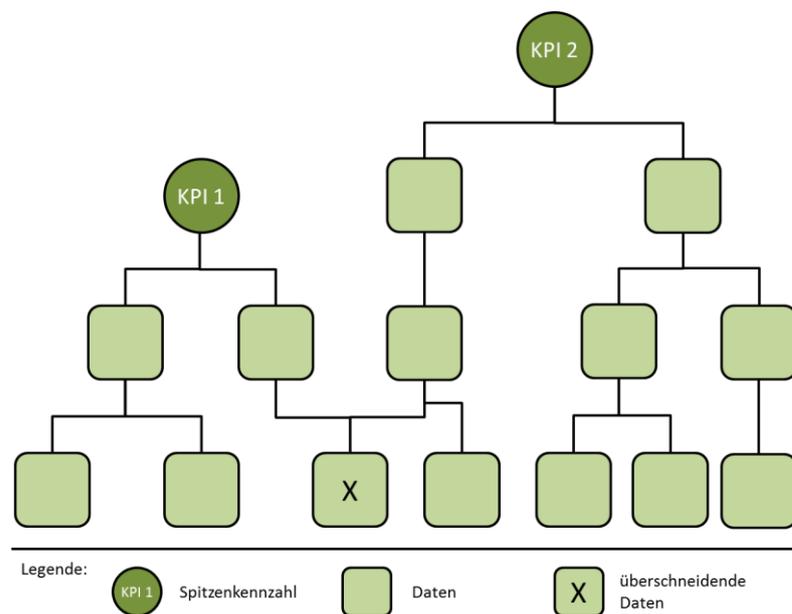
Die Berechnung von komplexen Kennzahlen im DWH-System ist oftmals aufgrund von fehlendem Fachwissen sowie unzureichender technischer Dokumentation (vgl. 3.4) für einen nicht Fach- und IT-Experten zunächst schwer nachzuvollziehen. Daher erfolgt in diesem ersten Schritt zunächst eine allgemeine Analyse der Kennzahlen ohne Technik-Bezug zum DWH-System. Ziel der allgemeinen Analyse ist es einen ersten Überblick über die Kennzahlen zu erhalten. Auf Grundlage dieser Erkenntnisse ist darauf aufbauend ein effektiver Austausch mit Fachexperten über die Implementierung im DWH-System im nächsten Schritt möglich. Der Schwerpunkt der Analyse liegt dabei auf den Attributen der Kennzahlen sowie der Logik, welche den Kennzahlen zugrunde liegt, da diese die wesentlichen Informationen für die weitere Vorgehensweise darstellen.

Im Rahmen der Analyse der Kennzahlen-Logik wird die Zusammensetzung der Kennzahlen untersucht. Das zugehörige Kennzahlensystem zu einer DWH-Kennzahl basiert auf einem Rechensystem (vgl. 3.4). Demnach wird die Zusammensetzung der Kennzahlen durch die nach und nach erfolgende Zerlegung und Unterteilung der Berechnungsvorschrift bestimmt. Hierbei handelt es sich um eine allgemeine Berechnungsvorschrift, die den Aufbau der Kennzahl widerspiegelt. Diese ist von der weitaus komplexeren Berechnung im DWH-System abzugrenzen. Die Art der Zerlegung und Unterteilung ergibt sich in Abhängigkeit von der jeweiligen Kennzahlenart (vgl. 3.1.3). Die so erfolgende Analyse der Kennzahlensysteme spiegelt den Gedanken des Top-Down-Ansatzes wider, der innerhalb der Systemanalyse einer Simulationsstudie angewendet wird (vgl. 2.3). Hier wird unter anderem mit Hilfe des Top-Down-Ansatzes die Komplexität des Systems durch Zergliederung in seine Elemente aufgelöst. Der Top-Down-Ansatz ist dabei durch eine Differenzierung, ausgehend vom Ganzen, gekennzeichnet. Für die Analyse der Kennzahlen-Logik bedeutet dies eine Zergliederung der Kennzahlensysteme in die Basisdaten und Kennzahlen. Ausgangspunkt bildet die zugehörige Spitzenkennzahl.

Zur Durchführung der Analyse wird zunächst auf Dokumentationen über die Kennzahlen und die entsprechenden Kennzahlensysteme z.B. in Form eines Kennzahlensteckbriefes zurückgegriffen (vgl. 3.4). Liegen diese nicht unvollständig oder in keiner eindeutigen Form vor, so ist es nicht möglich den vorhandenen Informationsbedarf ausschließlich mittels Dokumentationen abzudecken. Für diesen Fall sowie auch als Ergänzung zur Nutzung von Kennzahlendokumentationen wird das Gespräch mit einem Fachexperten gesucht. Der vorhandene Informationsbedarf bezieht

sich neben der Kennzahlen-Logik auch auf die Attribute der Kennzahlen. Handelt es sich um allgemein bekannte Kennzahlen, kann die zugehörige Berechnungsvorschrift auch aus der entsprechenden Fachliteratur entnommen werden. Für das in dieser Arbeit betrachtete Logistiknetzwerk des Werkstoffhandels stellt beispielsweise der durchschnittliche Bestand eine solche Kennzahl dar. Voraussetzung für die Verwendung einer allgemeingültigen Berechnungsvorschrift ist es sicherzustellen, dass diese keine unternehmensspezifischen Elemente enthält.

Ein wichtiger Aspekt innerhalb der Analyse der Berechnungsstruktur der Kennzahlen besteht darin, die jeweils weiteren zu messenden Kennzahlen zu beachten. Dies bedeutet, dass die Analysen der zu messenden Kennzahlen nicht vollkommen unabhängig voneinander durchgeführt werden. Hierdurch können mögliche Überschneidungen der Daten der verschiedenen Kennzahlensysteme ermittelt werden. Dies ist für die weitere Vorgehensweise von Bedeutung. In Abbildung 14 ist ein Beispiel für die Überschneidung der Daten zweier Kennzahlensysteme mit den Spitzenkennzahlen KPI 1 und KPI 2, die auch als Key Performance Indicator (KPI) bezeichnet werden, graphisch dargestellt. Das durch das X gekennzeichnete Element in der Abbildung repräsentiert Daten, die zur Berechnung von KPI 1 und KPI 2 benötigt werden.



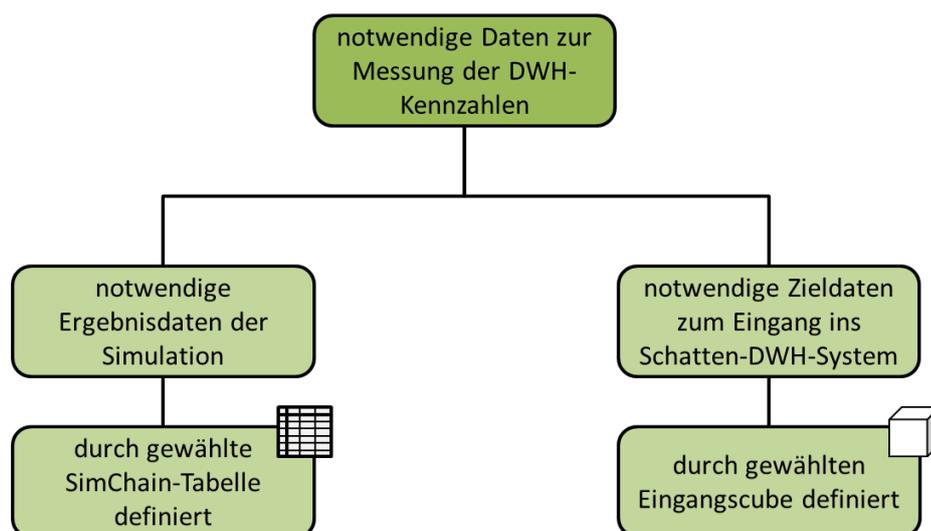
**Abbildung 14: überschneidende Daten zweier Kennzahlensysteme**

Durch die in diesem Schritt gewonnenen Informationen über die Kennzahlen, erhält der Anwender einen Überblick über die Logik sowie die beschreibenden Attribute der Kennzahlen. Zur Dokumentation der gewonnenen Informationen können die zugehörigen Kennzahlensysteme beispielsweise in Form einer Baumstruktur in einem Kennzahlenbaum graphisch dargestellt werden. Dieser wird zusammen mit einem Fachexperten für die Kennzahlen strukturiert überprüft. *Rabe et al.* bezeichnen diese

V&V-Technik als das strukturierte Durchgehen [RSW08a, S.104]. Da dieses Dokument die Basis für die weitere Vorgehensweise darstellt, ist die Überprüfung der ermittelten Informationen über die Kennzahlen von Bedeutung. Auf Grundlage dieser Kenntnisse wird die Implementierung der Kennzahlenberechnung im DWH-System untersucht werden. Weiterhin erhält der Anwender durch die Kenntnis über die Zusammensetzung der Kennzahlen eine erste Vorstellung davon, welche Elemente im Simulationsmodell dargestellt und welche Ergebnisdaten generiert werden müssen, sodass darauf aufbauend die Kennzahlenberechnung erfolgen kann.

#### 4.1.1.2 Identifikation der notwendigen Daten zur Messung der DWH-Kennzahl

Nachdem im vorherigen Schritt der Fokus zunächst nur auf der Logik der zu messenden Kennzahlen lag, wird in diesem Schritt, aufbauend auf den zuvor gewonnenen Informationen, die Kennzahlenberechnungen im DWH-System analysiert. Diese stellt sich anschaulich als Verknüpfung der in multidimensionalen Datenwürfeln gespeicherten Fakten und abgeleiteten Kennzahlen dar (vgl. 3.4). In diesem Zusammenhang wird im Folgenden der Begriff Cube als Synonym für die multidimensionalen Datenwürfel verwendet. Dabei kann ein Cube auch mehrere multidimensionale Datenwürfel enthalten. Wie zuvor beschrieben, basiert die Berechnung der Kennzahlen im Schatten-DWH-System auf den Ergebnisdaten der Simulation. Innerhalb dieses Prozessschrittes werden neben notwendigen Ergebnisdaten auch die notwendigen Zieldaten zum Eingang ins Schatten-DWH-System bestimmt. Die Vereinigungsmenge dieser Daten wird übergeordnet als die notwendigen Daten zur Messung der DWH-Kennzahlen bezeichnet (vgl. Abbildung 15). Innerhalb der Vorgehensweise zur Aufbereitung der Ergebnisdaten werden die Ergebnisdaten der Simulation zu den Zieldaten, die ins Schatten-DWH-System eingehen, aufbereitet.



**Abbildung 15:** Unterteilung der notwendigen Daten zur Messung der DWH Kennzahlen

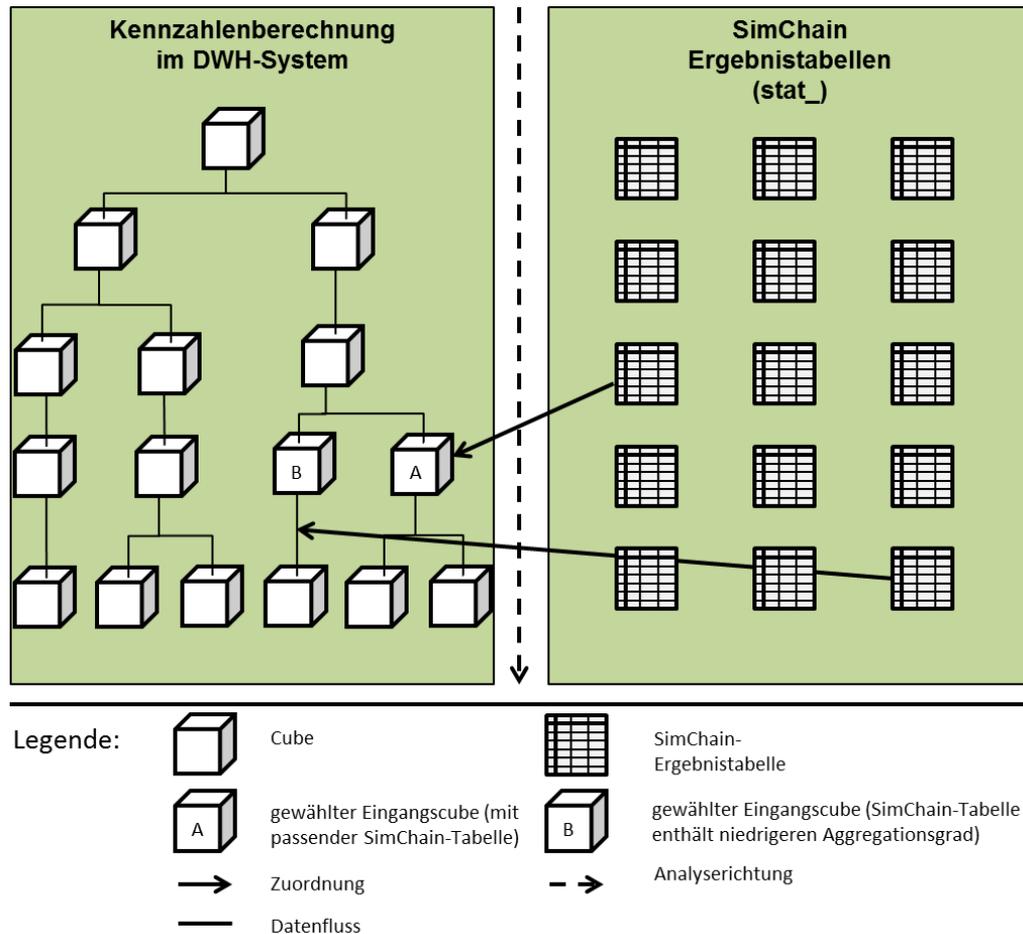
Wie nachfolgend beschrieben wird, werden auf der einen Seite die Ergebnistabellen von SimChain (vgl. 2.4) bestimmt, welche die notwendigen Ergebnisdaten beinhalten. Auf der anderen Seite werden die Cubes im Schatten-DWH-System ermittelt, in welche die Ergebnisdaten der Simulation eingepflegt werden. Diese sind als Eingangscubes bezeichnet. Demnach werden die notwendigen Ergebnisdaten durch die ausgewählten Ergebnistabellen von SimChain definiert. Die gewählten Eingangscubes bestimmen hingegen die notwendigen Zieldaten zum Eingang ins Schatten-DWH-System. (vgl. Abbildung 15)

### **Ermittlung der Eingangscubes und der zu verwendenden Ergebnistabellen**

Die Identifikation der Eingangscubes sowie der zu verwendenden Ergebnistabellen von SimChain bedingt sich gegenseitig. Da die Ergebnisdaten in die Eingangscubes geladen werden, ist es das Ziel, die zu verwendenden Ergebnistabellen und die Eingangscubes möglichst passend zueinander auszuwählen. Durch eine möglichst passende Wahl wird ein minimaler Aufbereitungsaufwand der erzeugten Ergebnisdaten zum Eingang in das Schatten-DWH-System erzielt. Im Idealfall bedeutet dies, dass die Eingangscubes und die zu verwendenden Ergebnistabellen sowohl inhaltlich als auch hinsichtlich des Aggregationsgrades übereinstimmen. Demnach ist an dieser Stelle mit Ausnahme eventuell notwendiger Formatierungen keine Aufbereitung der Ergebnisdaten notwendig. In Abbildung 16 wird dieser Idealfall durch den mit dem Buchstaben A gekennzeichneten Cube, dem eine übereinstimmende Ergebnistabelle zugeordnet ist, dargestellt. Hier enthalten beispielsweise Eingangscube und die zu verwendende Ergebnistabelle die Bestände pro Kunde, Monat und Lager. Ergibt sich keine Übereinstimmung zwischen Cubes und Ergebnistabellen, wird der Cube als Eingangscube gewählt, zu dem sich die Ergebnisdaten der Simulation möglichst einfach aufrechnen lassen. In Abbildung 16 ist dieser Fall durch den mit dem Buchstaben B gekennzeichneten Datenwürfel dargestellt. Hier beinhaltet der Eingangscube beispielsweise den Umsatz pro Jahr, Kunde und Lager. Die zugeordnete zu verwendende Ergebnistabelle enthält hingegen den Umsatz pro Monat, Kunde und Lager. Hier ist eine entsprechende Aggregation über die Zeit von Monat auf Jahr notwendig, um die erzeugten Ergebnisdaten in das Schatten-DWH-System laden zu können.

Grundsätzlich werden die Eingangscubes und die zu verwendenden Ergebnistabellen durch den Abgleich der Cubes innerhalb der Analyse der Kennzahlenberechnung im DWH-System mit den Ergebnistabellen ermittelt (vgl. Abbildung 16). Da sich die Berechnungsvorschrift von komplexen DWH-Kennzahlen, aus vielen Funktionen zusammensetzt (vgl. 3.4), bedarf es eines systematischen Vorgehens zur Analyse der Kennzahlenberechnung. Analog zur Systemanalyse innerhalb einer Simulationsstudie

(vgl. 2.3) werden auch im Kontext des Data-Warehouse-Systems die zwei Analyserichtungen Top-Down und Bottom-Up unterschieden (vgl. 3.2).



**Abbildung 16: Ermittlung der Eingangscubes im Schatten-DWH-System und der zugehörigen Ergebnistabellen in SimChain**

Für die Analyse der Kennzahlenberechnung und der damit einhergehenden Ermittlung der Eingangscubes und der zu verwendenden Ergebnistabellen wird das Top-Down-Vorgehen gewählt. Die Wahl der Top-Down-Richtung beruht auf der Grundidee, die Eingangscubes und damit einhergehend auch die zu verwendenden Ergebnistabellen auf einer möglichst aggregierten Ebene innerhalb der Kennzahlenberechnung zu definieren. Daraus resultiert, dass auch die notwendigen Daten zur Messung der DWH-Kennzahlen (vgl. Abbildung 15) einen höchstmöglichen Aggregationsgrad aufzeigen. Hierdurch wird die Berechnung der Kennzahl im Schatten-DWH-System vereinfacht. Aufgrund der entfallenden Rechenschritte wird eine effiziente Arbeitsweise und Leistungssteigerung des Schatten-DWH-Systems erzielt. Je höher die Ebene ist, in der die gewählten Eingangscubes liegen, desto weniger Berechnungsschritte sind anschließend im Schatten-DWH-System durchzuführen.

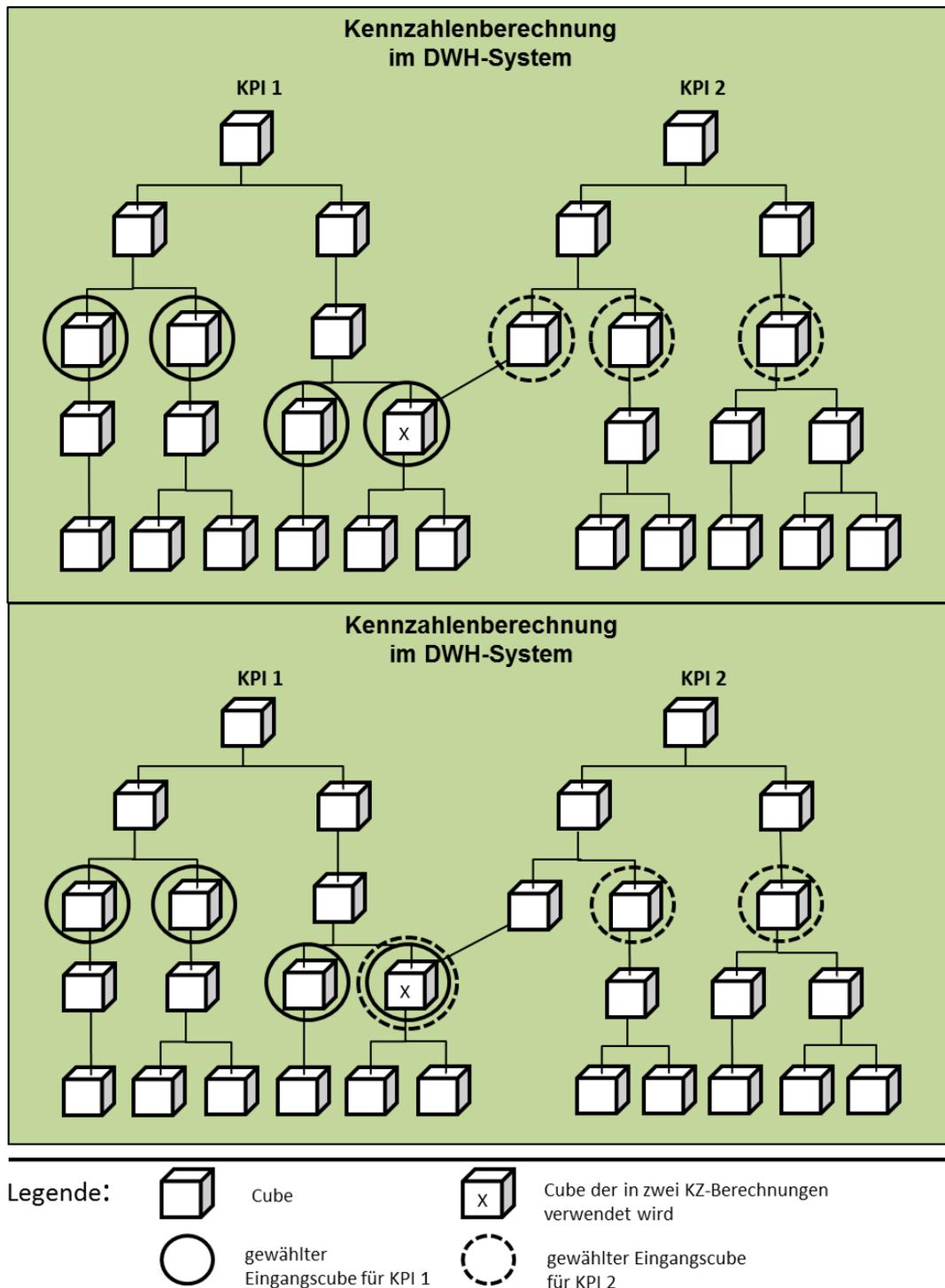
Dies bedeutet, dass innerhalb der Top-Down-Analyse der erste mögliche Cube als Eingangscube gewählt wird, zu dem eine inhaltlich passende Ergebnistabelle in

SimChain existiert. Der gewählte Eingangscube kann daher verschiedene Stufen innerhalb der logischen Kennzahlenberechnungen repräsentieren. In Abhängigkeit der zugeordneten zu verwendenden Ergebnistabellen werden sowohl Trace-Daten sowie Zählvariablen und Kennzahlen (vgl. 4), welche von SimChain zur Verfügung gestellt werden, genutzt (vgl. 2.4). Beispielsweise wird in SimChain der Lieferservicegrad berechnet. Durch die Verwendung der von SimChain bereitgestellten Kennzahlen, als höchstmögliche Aggregationsebene der notwendigen Daten, entfällt im Schatten-DWH-System die größtmögliche Menge an Berechnungsschritten. Ein weiterer Grund für die Verwendung der von SimChain bereits berechneten Kennzahlen ergibt sich für den Fall, dass die zur Berechnung dieser Kennzahl zugehörigen Trace-Daten von SimChain nicht gespeichert werden. An dieser Stelle wäre es demnach notwendig, die von SimChain berechnete Kennzahl in ihre Bestandteile zu zerlegen. Dies erfordert in Abhängigkeit von der jeweiligen Kennzahl einen hohen Aufwand oder ist gar unmöglich. Voraussetzung zur Nutzung der von SimChain berechneten Kennzahl ist der vorherige Abgleich mit der unternehmensspezifischen Berechnung dieser Kennzahl. Ergeben sich hierbei Abweichungen zwischen den Berechnungsvorschriften werden, wie oben beschrieben, die zugehörigen untergeordneten Ergebnisdaten verwendet.

### **Besonderheiten hinsichtlich weiterer zu messenden Kennzahlen**

Für den Fall, dass zwei oder mehrere DWH-Kennzahlen am Simulationsmodell gemessen werden sollen, ist eine mögliche Überschneidung der identifizierten notwendigen Daten zur Berechnung der verschiedenen Kennzahlen zu beachten. Ebenso kann sich eine Überschneidung der Kennzahlenberechnungen im DWH-System ergeben. Ein erster Anhaltspunkt für eine mögliche Überschneidung wird bereits innerhalb der allgemeinen Analyse (vgl. Abbildung 14) in Bezug auf die zugehörigen Kennzahlensysteme aufgezeigt. Die Überschneidung zweier Kennzahlenberechnungen im DWH-System ist in Abbildung 17 graphisch dargestellt. Die umkreisten Cubes stellen die ausgewählten Eingangscubes dar. Cube X ist Bestandteil der Kennzahlenberechnungen von KPI 1 und KPI 2 ein. Wird aufgrund der zuvor beschriebenen Gedanken dieser Cube für eins der beiden Kennzahlensysteme ausgewählt, so ist dieser gewählte Cube zwingend auch für das andere Kennzahlensystem als Eingangscube zu wählen (vgl. Abbildung 17 unten). Dies gilt ebenfalls für den Fall, dass ein Cube ausgewählt wird, der Cube X untergeordnet ist. In der oberen Hälfte von Abbildung 17 ist beispielhaft der Fall dargestellt, in dem die Ergebnisdaten der Simulation für KPI 1 in den gemeinsamen Cube X und für KPI 2 in einen höheren darauf aufbauenden Cube geladen werden. Innerhalb der Kennzahlenberechnung im DWH-System beinhalten die Cubes transformierte Daten der untergeordneten Cubes (vgl. 3.4). Im beschriebenen Fall werden für KPI 2 Ergebnisdaten in Cube X, als auch in den übergeordneten Cube geladen. Da die beiden Cubes innerhalb der Kennzahlenberechnung durch eine entsprechende Transformation

miteinander verbunden sind, erhält der übergeordnete Cube automatisch die Daten aus Cube X. Somit werden die Ergebnisdaten, die in den übergeordneten Cube geladen werden, überschrieben. Hierdurch können Fehler im Schatten-DWH-System entstehen.



**Abbildung 17: Überschneidung zweier Kennzahlenberechnungen im DWH-System**

Aufgrund der Tatsache, dass DWH-Systeme oftmals über einen langen Zeitraum historisch gewachsen sind, ist eine entsprechende Überschneidung der Kennzahlenberechnungen im DWH-System selten. Daher kann es ebenso der Fall sein,

dass zwei verschiedenen Eingangscubes aus zwei verschiedenen Kennzahlenberechnungen die gleiche Ergebnistabelle in SimChain zugeordnet ist. Hier ist darauf zu achten, dass die Ergebnisdaten für die verschiedenen Cubes gegebenenfalls unterschiedlich aufbereitet werden müssen.

In vielen Unternehmen liegt oftmals keine oder nur eine unzureichende Dokumentation über solche technischen Sachverhalten vor (vgl. 3.4). Daher wird zur Durchführung der Analyse der Kennzahlenberechnung im DWH-System der Austausch mit den IT-Fachexperten gesucht. Weiterhin dienen die im vorherigen Schritt ermittelten Informationen über die Logik der Kennzahl zum Grundverständnis der Berechnung.

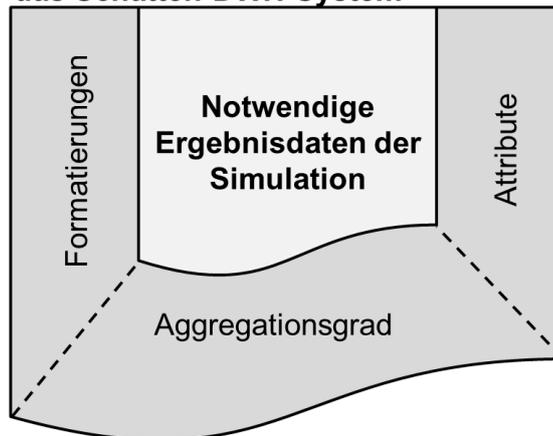
Die Ergebnisse dieses Schrittes sind einerseits die ermittelten Eingangscubes und der daraus resultierende Informationsbedarf bezüglich der notwendigen Zieldaten zum Eingang in Schatten-DWH-System. Andererseits wurden die zu verwendenden Ergebnistabellen von SimChain und damit einhergehend der Informationsbedarf bezüglich der notwendigen Ergebnisdaten der Simulation bestimmt. Die Eingangscubes und zu verwendenden Ergebnistabelle werden dabei so gewählt, dass auf der einen Seite der Aufbereitungsaufwand der notwendigen Ergebnisdaten zum Eingang in das Schatten-DWH-System minimal ausfällt. Auf der anderen Seite werden die Eingangscubes auf einer möglichst hohen Ebene gewählt, sodass die Berechnungsschritte im Schatten-DWH-System minimiert werden.

Abschließend wird der ermittelte Informationsbedarf validiert. Hierzu erfolgt eine sorgfältige Überprüfung des ermittelten notwendigen Informationsbedarfs durch den Anwender selbst. Dies wird als die V&V-Technik Schreibtischtest [RSW08a, S.102] bezeichnet. Zusätzlich erfolgt gemäß der V&V-Technik Strukturiertes Durchgehen [RSW08a, S.104] eine Überprüfung durch eine außenstehende Person, die nicht an der Definition des Informationsbedarf beteiligt war. Innerhalb der Validierung wird überprüft, ob durch den ermittelten notwendigen Informationsbedarf bezüglich der notwendigen Daten zur Messung der DWH-Kennzahl die Berechnung der DWH-Kennzahl möglich ist.

#### **4.1.2 Bestimmung der Abweichungen zwischen den notwendigen Ergebnisdaten und den notwendigen Zieldaten**

Die notwendigen Ergebnisdaten der Simulation sowie die notwendigen Zieldaten zum Eingang ins Schatten-DWH-System können sich in Bezug auf verschiedene Aspekte voneinander unterscheiden. Die Abweichungen der beiden Datensätze werden innerhalb dieses Prozessschrittes herausgestellt und zu einem großen Teil innerhalb der Vorgehensweise zur Aufbereitung der Ergebnisdaten eliminiert (vgl. 4.2). In Abbildung 18 sind die drei Abweichungsmöglichkeiten dargestellt.

### Notwendige Zieldaten zum Eingang in das Schatten-DWH-System



**Abbildung 18:** Abweichungsmöglichkeiten zwischen den notwendigen Ergebnisdaten der Simulation und den notwendigen Zieldaten zum Eingang in das Schatten-DWH-System

Die erste Abweichungsmöglichkeit besteht in Bezug auf die Formatierungen der beiden Datensätze. Diese ergeben sich durch Unterschiede zwischen den Formatierungsregeln für das DWH-System und denen für die Ergebnistabellen von SimChain. Die Unterschiede der Formatierungen beziehen sich beispielsweise auf die verwendeten Datentypen. Für Datentypen, die eine Zeichenkette enthalten, können sich zusätzlich Unterschiede hinsichtlich der syntaktischen Formatierung ergeben. Zur Ermittlung der Abweichungen werden die entsprechenden Formatierungsvorschriften von SimChain mit denen des DWH-Systems abgeglichen.

Eine weitere Abweichungsmöglichkeit resultiert aus den Aggregationsgraden der beiden Datensätze. In Abhängigkeit der Struktur der Kennzahlenberechnung im DWH-System, ist es nicht immer möglich dem ausgewählten Eingangscube eine, hinsichtlich des Aggregationsgrades, passende Ergebnistabelle zuzuordnen (vgl. Abbildung 16). Damit einhergehend weichen auch die notwendigen Ergebnisdaten und die notwendigen Zieldaten in Bezug auf den Aggregationsgrad voneinander ab.

Ein dritter möglicher Unterschied ergibt sich zwischen den Attributen der Datensätze. Auf der einen Seite ist es möglich, dass die Ergebnistabellen Attribute enthalten, die im zugehörigen Eingangscube nicht vorgesehen sind und somit aussortiert werden müssen. Auf der anderen Seite führen Attribute, die für den Eingang ins Schatten-DWH-System erforderlich sind, aber in den Ergebnistabellen der Simulation nicht abgebildet werden, zu einem höheren Bearbeitungsaufwand. Die erforderlichen Attribute umfassen nicht alle Attribute, die im Eingangscube vorhanden sind, sondern ausschließlich die Attribute, die für die Berechnung der Kennzahl erforderlich sind.

Die fehlenden Attribute in den Ergebnistabellen von SimChain lassen sich auf die folgenden Gründe zurückführen. Simulationsmodelle stellen nur eine Abstraktion des

realen Systems dar (vgl. 2.1) und spiegeln somit nicht alle Abläufe des realen Systems im Detail wider. Daher werden möglicherweise nicht alle, für die Kennzahlenberechnung relevanten Attribute, in der Simulation erzeugt. Die fehlenden Attribute werden nicht nur auf Vorgänge zurückgeführt, die im Simulationsmodell nicht abgebildet werden, sondern auch auf nicht vorhandene Stammdaten. Beispielsweise wird in SimChain nur der Name eines Kunden hinterlegt, während zum Eingang der Daten in das Schatten-DWH-System auch die Kundennummer benötigt wird. Außerdem kann es sich bei den für die Kennzahlenberechnung benötigten Attributen der Daten um unternehmensspezifische Definitionen handeln, die somit in SimChain ebenfalls nicht abgebildet werden.

Die ermittelten fehlenden Attribute sind für die Berechnung der Kennzahlen im Schatten-DWH-System zwingend notwendig. Daher bedarf es einer entsprechenden Erzeugung dieser Attribute. Die glaubwürdigste, aber auch aufwändigste Methode ist es, die entsprechend benötigten Vorgänge zur Erzeugung der fehlenden Attribute in das Simulationsmodell zu integrieren. Hierzu ist eine entsprechende Veränderung des SimChain-Codes notwendig. Dies ist nur von entsprechenden Fachexperten durchzuführen. Ist dies nicht möglich, müssen die fehlenden Daten auf eine andere Art und Weise in die Ergebnisdaten der Simulation integriert werden (vgl. 4.2.3).

Es folgt die Überprüfung der identifizierten Abweichungen. Hierzu werden erneut die V&V-Techniken Schreibtischtest und Strukturiertes Durchgehen [RSW08a, S.102 ff.] eingesetzt. Dies bedeutet, dass die Abweichungen sowohl durch den Anwender selbst, als auch durch außenstehende Person validiert werden. Hierzu wird überprüft, ob durch die Eliminierung der ermittelten Abweichungen der Informationsbedarf bezüglich der notwendigen Zieldaten für den Eingang ins Schatten-DWH gedeckt werden kann. Die Überprüfung des möglicherweise erweiterten SimChain-Codes erfolgt innerhalb der V&V der Ergebnisdaten (vgl. 4.2).

Die ermittelten Abweichungen repräsentieren den Aufbereitungsaufwand, der notwendig ist, um die Ergebnisdaten der Simulation in die notwendigen Zieldaten zum Eingang ins Schatten-DWH-System zu überführen. Die entsprechende Durchführung der Aufbereitung wird in der Vorgehensweise zur Aufbereitung der Ergebnisdaten beschrieben (vgl. 4.2).

### **4.1.3 Vorbereitung der Erhebung der realen Zieldaten**

Im Folgenden werden die zu den notwendigen Zieldaten zugehörigen realen Daten aus dem DWH-System erhoben und erfasst. Diese werden im Folgenden als reale Zieldaten bezeichnet. Die realen Zieldaten bilden im weiteren Verlauf der Vorgehensweise die Grundlage für die V&V der simulierten Ergebnisdaten. Ebenso werden auf Grundlage der erhobenen realen Zieldaten die notwendigen Eingangsdaten für die Simulation

ermittelt. Zu den notwendigen Ergebnisdaten existieren im DWH-System keine zugehörigen realen Daten. Die erhobenen notwendigen Zieldaten können, trotz der möglichen Abweichungen zu den notwendigen Ergebnisdaten, in den zuvor beschriebenen Anwendungssituationen verwendet werden.

In diesem Prozessschritt erfolgt die Erhebungsplanung. Im Gegensatz zum Erhebungsvorgang der Eingangsdaten im IGP beschränkt sich die Erhebungsplanung für die realen Zieldaten auf die Vorbereitung der Erhebung. Die realen Zieldaten werden aus dem ausgewählten Eingangscube im DWH-System gezogen. Demnach ist das DWH-System bereits als Informationsquelle festgelegt. Daher entfällt ein Großteil der vorbereitenden Schritte für die Erhebung, die im IGP aufgezeigt werden (vgl. 2.2.2.2). Hierzu zählen alle für die Auswahl der Informationsquelle notwendigen Schritte wie die Informationsangebotsanalyse, die Nutzbarkeitsüberprüfung der Informationsquellen sowie die Auswahl adäquater Informationsquellen. Weiterhin erübrigt sich die Auswahl der Erhebungsmethoden, da diese durch das DWH-System als Informationsquelle bestimmt ist. Die Datenbanken des DWH-Systems (vgl. 3.2) stellen strukturierte, zeichenorientierte und elektronische Dokumente dar, die mittels einer entsprechenden Datenbankabfrage ausgewertet werden können [JoMa04, S.17].

Zur Vorbereitung der Erhebung der realen Zieldaten ist demnach die Datenbankabfrage zu erstellen. Für den Zugriff auf die entsprechenden Eingangscubes muss zuvor gegebenenfalls eine entsprechende IT-Berechtigung eingeholt werden. Bei der Erstellung der Abfrage ist darauf zu achten, dass diese alle erforderlichen Attribute, die für die Berechnung der DWH-Kennzahlen benötigt werden, enthält. Anschließend wird die Datenbankabfrage durch entsprechende Voruntersuchungen auf gültige und zuverlässige Ergebnisdaten überprüft. Die Bewertung sowie die Überprüfung der erstellten Datenbankabfrage erfolgt durch einen IT-Fachexperten.

#### **4.1.4 Erhebung der realen Zieldaten**

Innerhalb der Datenerhebung werden die realen Zieldaten aus dem gewählten Eingangscube mittels der zuvor erstellten Datenbankabfrage extrahiert und für die nachfolgenden Schritte zur Verfügung gestellt. Da die Ergebnisse der Datenbankabfrage elektronisch gespeichert werden und somit in digitaler Form vorliegen, handelt es sich um eine integrierte Datenerfassung. Es ist daher kein zusätzlicher Prozessschritt zur Datenerfassung notwendig. (vgl. 2.2.2.2)

Es folgt die Überprüfung der erhobenen realen Zieldaten. In einem ersten Schritt werden die erhobenen realen Zieldaten in einem Gespräch mit einem Fachexperten validiert. Dieser kann vor seinem Erfahrungshintergrund einschätzen, ob die erhobenen realen Zieldaten die gewünschten notwendigen Zieldaten abbilden. Diese V&V-Technik wird als Validierung im Dialog bezeichnet [RSW08a, S.109]. Zur Verifikation der

erhobenen realen Zieldaten werden die Kennzahlen auf Basis der erhobenen realen Zieldaten nachberechnet und mit den realen historischen gemessenen Kennzahlen verglichen. Dies beschreibt die V&V-Technik Vergleich mit aufgezeichneten Daten [RSW08a, S.111]. Die Berechnung erfolgt dabei anhand der in vorherigen Prozessschritten ermittelten Berechnungslogik der Kennzahlen. Die Berechnung kann beispielsweise in einem Tabellenkalkulationsprogramm durchgeführt werden. Zur Aufklärung von Abweichungen der berechneten Kennzahlen von den real gemessenen Kennzahlen wird zusätzlich das Gespräch mit dem zugehörigen Fachexperten gesucht.

Ziel des gesamten Prozessschrittes ist die Bereitstellung der realen Zieldaten. Aus diesen werden im nächsten Prozessschritt die entsprechend notwendigen Eingangsdaten für das Simulationsmodell abgeleitet. Weiterhin bildet die überprüfte Datenbasis die Grundlage für die später folgende Verifikation der tatsächlich simulierten Ergebnisdaten.

#### **4.1.5 Zieldefinition bezüglich der notwendigen Eingangsdaten**

Aufbauend auf dem identifizierten Informationsbedarf bezüglich der notwendigen Daten zur Messung der DWH-Kennzahlen sowie der erhobenen Zieldaten selbst, werden im Folgenden die notwendigen und nutzbaren Eingangsdaten für die Simulation ermittelt. Hierzu wird an dieser Stelle der entwickelten Vorgehensweise der IGP von *Bernhard et al.* integriert. Dabei werden die Prozessschritte grundsätzlich übernommen. Innerhalb der Beschreibung des integrierten IGP liegt der Fokus allerdings auf der Spezifizierung des IGP in Bezug auf den vorliegenden Sachverhalt sowie auf den V&V-Elementen innerhalb der Prozessschritte.

Der IGP beginnt mit der Zieldefinition bezüglich der notwendigen Eingangsdaten (vgl. 2.2.2.2). Diese umfasst die Ermittlung des objektiven Informationsbedarfs bezüglich der notwendigen Eingangsdaten. Die darin enthaltene Analyse der Aufgabenstellung bezieht sich in dieser Vorgehensweise explizit auf den zuvor definierten Informationsbedarf bezüglich der notwendigen Daten zur Messung der DWH-Kennzahlen. Anschließend werden die Ziele für die Informationsgewinnung, der objektive Informationsbedarf bezüglich der notwendigen Eingangsdaten, hergeleitet. Im Folgenden werden der objektive und subjektive Informationsbedarf bezüglich der notwendigen Eingangsdaten ausschließlich als objektiver und subjektiver Informationsbedarf bezeichnet. Die Eindeutigkeit der Bezeichnung bleibt dabei erhalten.

##### **4.1.5.1 Analyse der notwendigen Daten zur Messung der DWH-Kennzahl**

Analog zur Analyse der Aufgabestellung im IGP erfolgt im vorliegenden Fall die Analyse der notwendigen Daten zur Messung der DWH-Kennzahl. Die Analyse beinhaltet eine systematische Untersuchung der Entstehung der notwendigen

Ergebnisdaten. Vor diesem Hintergrund wird ermittelt welche Elemente und zugehörigen Attribute des Logistiknetzwerks im Simulationsmodell abgebildet werden müssen, sodass die notwendigen Ergebnisdaten generiert werden können. Für den Fall eines Logistiknetzwerk des Werkstoffhandels, stellen beispielsweise Lager, Kunden oder Lieferanten Elemente und die Zuordnung von Kunden zu Lägern Beziehungen des Modells dar (vgl. 2.1). Es erfolgt somit eine Vorauswahl der zur Beschreibung der notwendigen Ergebnisdaten zu verwendenden Elemente und Standardprozessen.

#### **4.1.5.2 Ableitung der Ziele für die Informationsgewinnung**

Im Anschluss an die Analyse der notwendigen Daten zur Messung der DWH-Kennzahl erfolgt die Festlegung des objektiven Informationsbedarfs. Die Festlegung erfolgt auf Grundlage der detaillierten Untersuchung der notwendigen Daten zur Messung der DWH-Kennzahl in den vorherigen Schritten sowie den zuvor aufgezeigten abzubildenden Elementen und Standardprozessen. (vgl. 2.2.2.2)

Im Rahmen einer Simulationsstudie werden hier beispielsweise Checklisten zur Datenerhebung von *Baron et al.* verwendet [BDKR01, S.151 ff.]. Darüber hinaus kann zur Ermittlung der notwendigen Eingangsdaten die *Methodik zur systematischen Informationsgewinnung für Simulationsstudien* von *Csanady et al.* genutzt werden [CBW08]. (vgl. 2.2.2)

Es folgt die Überprüfung des objektiven Informationsbedarfs auf Vollständigkeit. Hierzu wird die V&V-Technik strukturiertes Durchgehen verwendet [RSW08a, S.104 f.]. Dies bedeutet, dass der ermittelte objektive Informationsbedarf gemeinsam mit einem Fachexperten für die Kennzahl, der nicht an der Erstellung des objektiven Informationsbedarfs beteiligt war, strukturiert kontrolliert wird. Dabei wird überprüft, inwieweit die im objektiven Informationsbedarf aufgezeigten Daten zur Abbildung der notwendigen Ergebnisdaten genügen. Der Vorteil dieser V&V-Technik besteht in der unvoreingenommenen Sicht des Fachexperten und dessen Kenntnissen.

#### **4.1.6 Informationsidentifikation**

Der Prozessschritt der Informationsidentifikation umfasst im IGP drei Komponenten. Zu diesen zählen die Informationsbedarfsanalyse, die Informationsangebotsanalyse und –bewertung sowie die Nutzbarkeitsprüfung der Informationsquellen. (vgl. 2.2.2.2)

##### **4.1.6.1 Informationsbedarfsanalyse**

Innerhalb der entwickelten Vorgehensweise liegt der Fokus der Informationsbedarfsanalyse auf der Spezifizierung des objektiven Informationsbedarfs bezüglich der notwendigen Eingangsdaten. Die Spezifizierung erfolgt in Bezug auf die Anforderungen, welche sich durch den Einsatz der Simulationssoftware SimChain

ergeben. Diese können sich beispielsweise auf die Granularität und die Darstellungsform der Eingangsdaten beziehen.

#### **4.1.6.2 Informationsangebotsanalyse und -bewertung**

Die Komponente der Informationsangebotsanalyse und -bewertung umfasst die Bewertung der zur Deckung des subjektiven Informationsbedarfs möglichen Informationsquellen anhand verschiedener Kriterien. (vgl. 2.2.2.2)

Die Kennzahlenberechnungen in einem DWH-System basieren teilweise auf den Output-Daten des Logistiknetzwerks, wie beispielsweise dem Bestand oder dem Warenausgang. Daher sind die benötigten Eingangsdaten für die Simulation oftmals nicht in den Kennzahlenberechnungen im DWH-System enthalten. Lediglich konstante Größen, die keine Output-Größen des Logistiknetzwerks darstellen, können teilweise der Kennzahlenberechnung im DWH-System entnommen werden. Dies wird am Beispiel der Kennzahl Transportauslastung aufgezeigt. Die Transportauslastung ergibt sich allgemein aus der Gegenüberstellung von Warenausgang pro Standort und der Kapazitäten der Transportfahrzeuge. Die Berechnung der Transportauslastung basiert im DWH-System direkt auf dem Warenausgang, als Output-Größe des Logistiknetzwerks und den Kapazitäten der Transportfahrzeuge, als Stammdaten des Logistiknetzwerkes. Dieser wird aus einem entsprechenden operativen System direkt in das DWH-System geladen. Die zur Generierung des Warenausgangs erforderlichen Eingangsdaten sind daher kein Bestandteil der Kennzahlenberechnung im DWH-System. Für diese Daten sind somit andere Informationsquellen zu ermitteln. Im Gegensatz dazu kann die Angabe über die Kapazität der Fahrzeuge in der Kennzahlenberechnung enthalten sein, da diese Stammdaten darstellen.

Grundsätzlich existieren verschiedene mögliche Informationsquellen für die Datenbeschaffung innerhalb einer Simulationsstudie. Hierzu zählen unter anderem manuelle Messungen und Aufzeichnungen, Datenbanken, Textdateien und Tabellenkalkulationen sowie automatische Erfassung der Daten durch Lesestationen [RSW08a, S.88]. Vor dem Hintergrund die entwickelte Vorgehensweise in einem automatisierten Prozess umzusetzen, werden primär Informationsquellen genutzt, welche eine automatisierte Datenerhebung ermöglichen (vgl. 4). Hierzu zählen elektronische Dokumente wie beispielsweise Datenbanken und Tabellen deren Analyse im Rahmen einer Sekundärerhebung erfolgt [JHL04, S.23 ff.]. Für Dokumente in Papierform, wie beispielsweise Bilder, Grafiken oder Texte, lässt sich der Erhebungsvorgang nicht automatisieren.

#### **4.1.6.3 Nutzbarkeitsprüfung der Informationsquellen**

Der Prozessschritt Informationsidentifikation schließt mit der Nutzbarkeitsprüfung der Informationsquellen ab. Diese beinhaltet den Abgleich zwischen dem zuvor

aufgezeigten Informationsangebot und dem ermittelten notwendigen subjektiven Informationsbedarf bezüglich der Eingangsdaten. Hierbei muss beachtet werden, dass alle im subjektiven Informationsbedarf aufgezeigten notwendigen Eingangsdaten in der geforderten Granularität in den Datenquellen enthalten sind. (vgl. 2.2.2.2)

#### **4.1.7 Erhebungsplanung**

Im IGP erfolgt als nächster Prozessschritt die Erhebungsplanung. Diese setzt sich aus den Teilschritten Auswahl adäquater Informationsquellen, Auswahl der Erhebungsmethoden sowie Vorbereitung der Erhebung zusammen. (vgl. 2.2.2.2)

##### **4.1.7.1 Auswahl adäquater Informationsquellen**

Der Auswahlprozess der adäquaten Informationsquellen erfolgt vor dem Hintergrund der bestmöglichen Effizienz der Erhebung. Anhand der ermittelten Effizienz der Informationsquellen erfolgt eine entsprechende Priorisierung. (vgl. 2.2.2.2)

In Bezug auf die entwickelte Vorgehensweise liegt hier der Schwerpunkt auf der Auswahl von Informationsquellen, die eine automatisierte Erhebung ermöglichen. Dies ist vor dem Hintergrund, die Vorgehensweise in einem automatisierten Prozess umzusetzen, von Bedeutung. Demnach erhalten beispielsweise elektronische Dokumente innerhalb der entwickelten Vorgehensweise eine hohe Effizienz.

##### **4.1.7.2 Auswahl der Erhebungsmethoden**

Im Anschluss an die Auswahl der Informationsquellen werden für diese geeignete Erhebungsmethoden bestimmt. (vgl. 2.2.2.2)

Auch an dieser Stelle werden aufgrund der gewünschten Umsetzung der entwickelten Vorgehensweise in einem automatisierten Prozess vorrangig Erhebungsmethoden gewählt, die automatisierbar sind. Die Datenbankabfrage stellt eine solche Erhebungsmethode dar. Grundsätzlich ist es möglich sowohl Primär- als auch Sekundärerhebungen zu automatisieren.

Sowohl die ausgewählten Informationsquellen, als auch die festgelegten Erhebungsmethode werden durch das Management auf Auftraggeber- und Auftragnehmer-Seite genehmigt. Dies entspricht der V&V-Technik Begutachtung [RSW08a, S.97].

##### **4.1.7.3 Vorbereitung der Erhebung**

Der Teilschritt Vorbereitung der Erhebung umfasst alle zur unmittelbaren Durchführung der Erhebung notwendigen Schritte. (vgl. 2.2.2.2)

Für die Erstellung der erforderlichen Erhebungsunterlagen liegt der Fokus in der entwickelten Vorgehensweise auf der Erstellung einer Datenbankabfrage. Entsprechend der Forderung im IGP werden die erstellten Erhebungsinstrumente durch

Voruntersuchungen auf gültige und zuverlässige Erhebungsergebnisse überprüft sowie eventuell notwendige Genehmigungen eingeholt und involvierte Personen informiert.

#### **4.1.8 Erhebung und Datenerfassung**

Die Datenerhebung beinhaltet sowohl die Extraktion der erforderlichen Daten aus der gewählten Informationsquelle sowie die entsprechende Bereitstellung für die nachfolgenden Prozessschritte. (vgl. 2.2.2.2)

Da die Daten in der entwickelten Vorgehensweise oftmals mittels einer Datenbankabfrage aus elektronischen Dokumenten erhoben werden, ist keine gesonderte Datenerfassung notwendig. Erfolgt die Erhebung nicht aus elektronischen Dokumenten, so ist zusätzlich eine entsprechende Datenerfassung durchzuführen.

##### **4.1.8.1 Erhebungsvalidierung**

Nach dem IGP erfolgt innerhalb der Erhebungsvalidierung der Abgleich der erhobenen Eingangsdaten mit dem in der Informationsbedarfsanalyse aufgezeigten subjektiven Informationsbedarf. Neben den erhobenen Eingangsdaten soll auch der Erhebungsvorgang auf Qualität und Vollständigkeit hin überprüft werden. (vgl. 2.2.2.2)

Die erhobenen Eingangsdatendaten können beispielsweise zunächst durch eine eigene Überprüfung auf Vollständigkeit in Bezug auf den subjektiven Informationsbedarf hin untersucht werden. Da diese als Schreibtischtest [RSW08a, S.102] bezeichnete V&V-Technik die Gefahr birgt, eigene Fehler auch innerhalb der Überprüfung nicht zu erkennen, wird zusätzlich der Austausch mit einem Fachexperten gesucht. Hier erfolgt die sogenannte Validierung im Dialog [RSW08a, S.109]. Durch seinen Erfahrungshintergrund ist es dem Fachexperten möglich die erhobenen Eingangsdaten hinsichtlich Vollständigkeit und Qualität in Bezug auf den subjektiven Informationsbedarf auf Gültigkeit zu überprüfen.

#### **4.1.9 Datenstrukturierung**

Der Prozessschritt Datenstrukturierung umfasst im IGP von *Bernhard et al.* fünf Teilschritte. Zu diesen zählen die syntaktische Formatierung, die Definition von Relationen, die Fehlerbereinigung sowie die Anpassung der Granularität. Abschließend erfolgt eine Plausibilitätsprüfung. (vgl. 2.2.2.2)

Für die entwickelte Vorgehensweise ist hier Folgendes zu beachten. Da die Ergebnisdaten der Simulation zur Kennzahlenberechnung im Schatten-DWH-System genutzt werden, ist es erforderlich, dass die Ergebnisdaten in die notwendige Form der Zieldaten transformiert werden. Dies erfolgt durch verschiedene Aufbereitungsschritte (vgl. 4.2). Für den Fall, dass die Eingangsdaten in einem Datentypen vorliegen, der eine Zeichenkette enthält und diese Daten in SimChain in einem numerischen Datentypen gespeichert werden, ist eine Datenmapping erforderlich. Dies bedeutet, dass die

Eingangsdaten für den Eingang in SimChain in diesem Prozessschritt auf Zahlen umgestellt werden müssen. Innerhalb der Aufbereitung der Ergebnisdaten werden diese wieder in ihre ursprüngliche Form umgewandelt.

Zur Durchführung der Plausibilitätsprüfung der vorgenommen Änderungen können beispielsweise die folgenden V&V-Techniken eingesetzt werden. Die Datenverifikation umfasst verschiedene Punkte. So werden durch Anwendung statistischer Techniken [RSW08a, S.103] mögliche Eigenschaften der Daten, beispielsweise in Form von Verteilungen nachgewiesen. Weitere Eigenschaften der Daten können sich unter anderem in Bezug auf die Anzahl der Ziffern in numerischen Daten, wie zum Beispiel bei der Kundennummer oder hinsichtlich des Wertebereiches ergeben. Vor dem Hintergrund, dass die Eigenschaften der Daten dem Simulationsexperten bekannt sind, überprüft dieser die Daten zunächst selbstständig im Sinne eines Schreibtischtests [RSW08a, S.102]. Zusätzlich wird zum Nachweis der Eigenschaften der Daten sowie zur Überprüfung der festgelegten Relationen ein Fachexperte mit entsprechendem Fachwissen mit einbezogen. Dies beschreibt die V&V-Technik Validierung im Dialog [RSW08a, S.109] Weiterhin werden die identifizierten fehlenden Werte und Ausreißer bewertet.

Aufbauend auf positiven Ergebnissen der Verifikation wird innerhalb der anschließenden Datenvalidierung einerseits die Vollständigkeit der Daten in einem Gespräch mit einem Fachexperten überprüft. Andererseits untersucht das Management von Auftraggeber- und Auftragnehmerseite, ob die Daten trotz Qualitätsverschlechterung durch Fehler und Ausreißer ausreichend repräsentativ sind. Diese V&V Technik wird als Begutachtung [RSW08a, S.97] bezeichnet.

#### **4.1.1 Statistische Datenanalyse**

Die statistische Datenanalyse beinhaltet vier Teilschritte. Diese sind die deskriptive Analyse, die Auswahl der notwendigen statistischen Verfahren sowie die anschließende Durchführung der ausgewählten Verfahren unter Verwendung statistischer Software. Außerdem erfolgt abschließend eine Ergebnisüberprüfung. (vgl. 2.2.2.2)

Innerhalb der Ergebnisüberprüfung kann beispielsweise durch den Einsatz von Anpassungstest im Rahmen der statistischen Techniken zur V&V [RSW08a, S.103 f.] die hinreichend genaue Abbildung der Realität durch die ermittelten Verteilungen untersucht werden.

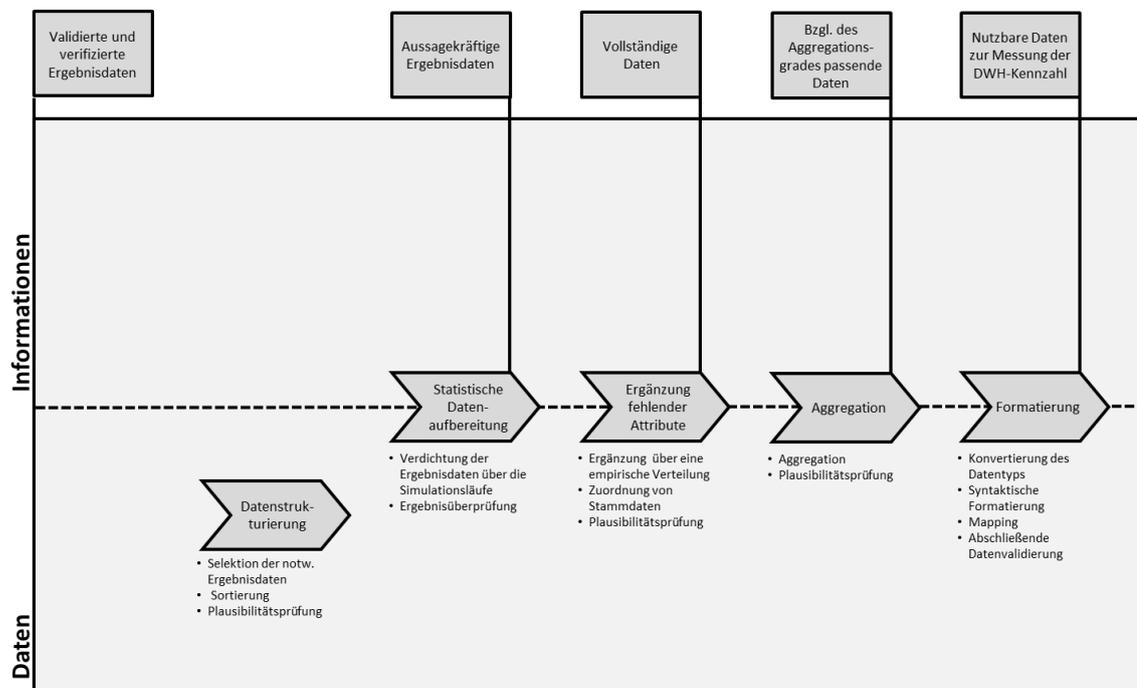
#### **4.1.2 Datennutzbarkeitsprüfung**

Im letzten Prozessschritt des IGPs, der Datennutzbarkeitsprüfung, werden die für das übergeordnete ausführbare Simulationsmodell nutzbaren Eingangsdaten erzeugt. Hierzu wird zunächst innerhalb der abschließenden Datenvalidierung überprüft, ob die

vorliegenden nutzbaren Eingangsdaten den subjektiven Informationsbedarf abdecken. Anschließend werden durch die Nutzbarkeitsbewertung die validierten potenziellen nutzbaren Eingangsdaten in Bezug auf die Eignung für das Simulationsmodell untersucht. Zur Durchführung der Überprüfung können verschiedene V&V-Techniken [RSW08a, S.113] genutzt werden. (vgl. 2.2.2.2)

## 4.2 Vorgehensweise zur Aufbereitung der Ergebnisdaten

Im Gegensatz zu den Eingangsdaten existiert für die Ergebnisdaten in der Literatur keine Vorgehensweise zur Aufbereitung der Daten (vgl. 2.2.4). Daher orientiert sich die entwickelte Vorgehensweise für die Ergebnisdaten an dem Aufbereitungsschritt, der zu den drei Hauptschritten zur Auswertung von Simulationsergebnissen in der VDI-Richtlinie 3633 zählt. Dieser beinhaltet das Handhabarmachen sowie die anschließende statistische Aufbereitung der Ergebnisdaten (vgl. 2.2.4). Die Darstellung der entwickelten Vorgehensweise zur Aufbereitung der Ergebnisdaten erfolgt analog zur entwickelten Vorgehensweise für die Eingangsdaten (vgl. Abbildung 13) in einem prozessorientierten Vorgehensmodell (vgl. Abbildung 19). In diesem bleiben ebenfalls die Darstellungsformen, wie beispielsweise die Trennung von Daten und Informationen, die im IGP verwendet werden (vgl. 2.2.2.2) erhalten, sodass sich eine einheitliche Darstellung für die gesamte Vorgehensweise ergibt.



**Abbildung 19: Prozessorientiertes Vorgehensmodell zur Aufbereitung von Ergebnisdaten zur glaubwürdigen Messbarkeit von komplexen DWH-Kennzahlen.**

In den folgenden Abschnitten werden die einzelnen Schritte des in Abbildung 19 dargestellten prozessorientierten Vorgehensmodells erläutert. Ausgangspunkt für die Vorgehensweise sind die validierten und verifizierten Ergebnisdaten der Simulation. Voraussetzung hierfür ist die kontinuierliche V&V innerhalb der gesamten Simulationsstudie. Zur V&V der Ergebnisdaten existieren nach *Rabe et al.* verschiedene Techniken (vgl. [RSW08a, S.113]). Ein Teil der Techniken kann erst im Anschluss an die statistische Aufbereitung der Ergebnisdaten angewendet werden (vgl. 4.2.2). Im

Zusammenhang mit der Vorgehensweise zur glaubwürdigen Messbarkeit von komplexen DWH-Kennzahlen eignet sich zur Überprüfung der Ergebnisdaten beispielsweise der Dimensionstest [RSW08a, S.98]. Mit Hilfe des Dimensionstests können die in SimChain berechneten, insbesondere die zur Messung der Kennzahlen eigens integrierten Zählvariablen oder Kennzahlen (vgl. 2.4), durch die Nachrechnung der zugehörigen Dimensionen überprüft werden. Durch den Test der internen Validität [RSW08a, S.105] ist es möglich, die Schwankungsbreite der Replikationen mit der Schwankungsbreite des realen Systems abzugleichen und somit die Ergebnisdaten zu verifizieren. Eine weitere Überprüfungsmöglichkeit besteht beispielsweise in der Validierung der Ergebnisdaten im Dialog mit einem Fachexperten. Dieser kann vor seinem Erfahrungshintergrund einschätzen, inwiefern die Ergebnisdaten die Realität hinreichend genau widerspiegeln [RSW08a, S.109]. Die Überprüfung der simulierten Ergebnisdaten erfolgt wird in den Prozessschritt Nutzbarkeitsprüfung der Vorgehensweise zur Aufbereitung der Eingangsdaten eingeordnet (vgl. 4.1.2).

Die Menge der Ergebnisdaten ergibt sich aus der Anzahl der Replikationen der Simulation. Allgemein weisen Ergebnisdaten die in Abschnitt 2.2.3 beschriebene Form auf, in der innerhalb einer Zeile die Ergebnisdaten eines Simulationsdurchlaufes aufgezeigt werden. Die Darstellung der Ergebnisdaten in SimChain ist in Kapitel 2.4 beschrieben worden.

Innerhalb des entwickelten prozessorientierten Vorgehensmodells werden die erzeugten Ergebnisdaten in einem ersten Schritt durch eine entsprechende Datenstrukturierung handhabbar gemacht. Es folgt die statistische Aufbereitung der strukturierten Ergebnisdaten in Bezug auf die einzelnen Replikationen. Darauf aufbauend werden die statistisch aufbereiteten Ergebnisdaten in Bezug auf die zuvor ermittelten Abweichungen von den Zieldaten zum Eingang ins Schatten-DWH-System (vgl. 4.1.2) entsprechend weiter aufbereitet. Abschließend werden die im Zieldatenformat vorliegenden Ergebnisdaten in die entsprechenden Eingangscubes des Schatten-DWH-Systems geladen. Zur Aufbereitung der Ergebnisdaten werden diese über die ODBC-Schnittstelle aus der MySQL-Datenbank in ein Tabellenkalkulationsprogramm exportiert (vgl. 2.4).

### **4.2.1 Datenstrukturierung**

Der erste Prozessschritt der entwickelten Vorgehensweise für die Ergebnisdaten ist die Datenstrukturierung. Diese umfasst in Anlehnung an die VDI Richtlinie 3633 verschiedene Aufbereitungspunkte, durch deren Ausführung die Daten für den Anwender handhabbar gemacht werden (vgl. 2.2.4). Voraussetzung für die Durchführung der Datenstrukturierung ist der vorherige Export der zu Berechnung der Kennzahlen benötigten Ergebnistabellen in ein Tabellenkalkulationsprogramm. Die erforderlichen Ergebnistabellen werden im Prozessschritt Zieldefinition bezüglich der

notwendigen Daten zur Messung der DWH-Kennzahl bestimmt (vgl. 4.1.1.2). Das primäre Ziel dieses Prozessschrittes ist die Aufbereitung der simulierten Ergebnisdaten in eine für die weiteren Aufbereitungsschritte notwendige Form.

Angelehnt an die Selektion bestimmter Daten aus dem Datenbestand nach VDI Richtlinie 3633 (vgl. 2.2.4), werden zunächst die für die Kennzahlenberechnung notwendigen Ergebnisdaten aus den exportierten Ergebnistabellen selektiert. Die notwendigen Ergebnisdaten werden im Informationsbedarf bezüglich der notwendigen Daten zur Messung der DWH-Kennzahlen definiert. Die entsprechenden erforderlichen Attribute der notwendigen Ergebnisdaten werden durch die Klassifikationsstufen der Dimensionen des Eingangscubes bestimmt. Hierbei sind allerdings nur solche Dimensionen relevant, welche für die Berechnung der Kennzahlen benötigt werden.

Anschaulich werden innerhalb der Selektion die Spalten der exportierten Ergebnistabellen ausgewählt, die den erforderlichen Attributen entsprechen. Der Aggregationsgrad der Ergebnistabellen von SimChain stimmt nicht grundsätzlich mit dem Aggregationsgrad der Eingangscubes überein (vgl. 4.1.1.2). Demnach entsprechen die zu selektierenden Spalten nicht immer den erforderlichen Attributen. Für den Fall, dass die Ergebnistabellen einen geringeren Aggregationsgrad als die Eingangscubes aufweisen, entspricht die selektierte Spalte einer untergeordneten Klassifikationsstufe der Dimensionen. So werden die Daten im Eingangscube beispielsweise pro Jahr gespeichert. Da in der zugeordneten SimChain-Tabelle die Ergebnisdaten allerdings nicht auf Jahresbasis zur Verfügung gestellt werden, wird mit den Monatswerten eine untergeordnete Klassifikationsstufe ausgewählt. Neben den für die Kennzahlenberechnung notwendigen Ergebnisdaten wird zusätzlich die Spalte IdRun zur Identifikation der verschiedenen Läufe mit ausgewählt. Diese Informationen sind für die anschließende statistische Datenaufbereitung erforderlich.

Nachdem die für die Kennzahlenberechnung notwendigen Ergebnisdaten aus den erzeugten Ergebnisdaten ausgewählt wurden, werden diese entsprechend den verschiedenen Cubes sortiert. Dies wird innerhalb der VDI-Richtlinie 3633 als die Sortierung der Daten nach einzelnen Kriterien bezeichnet. Die Sortierung erfolgt allerdings nicht in eine auftrags- oder elementorientierte Sichtweise (vgl. 2.2.4), sondern in Bezug auf die ausgewählten Eingangscubes, in welche die Ergebnisdaten im Schatten-DWH-System geladen werden. Für den Fall, dass die notwendigen Ergebnisdaten für einen Cube in verschiedenen Ergebnistabellen vorliegen, müssen an dieser Stelle adäquate Relationen zwischen den erzeugten Ergebnisdaten definiert werden. Hierzu wird beispielsweise in Excel für jeden Cube ein eignes Tabellenblatt angelegt, in dem ausschließlich die Daten für den jeweiligen Cube enthalten sind.

Der Prozessschritt der Datenstrukturierung schließt mit einer Plausibilitätsprüfung der selektierten und sortierten Ergebnisdaten ab. Diese beinhaltet die Überprüfung der

aufbereiteten Daten in Bezug auf die Richtigkeit der vorgenommenen Selektion und Sortierung sowie hinsichtlich der Weiterverwendbarkeit in den folgenden Prozessschritten. Für den Fall, dass die Ergebnisdaten eines Cubes in verschiedenen Ergebnistabellen vorliegen, sind die entsprechend definierten Relationen zwischen diesen Ergebnisdaten auf Widerspruchsfreiheit hin zu untersuchen. Hierzu wird der Schreibtischtest durchgeführt [RSW08a, S.102]. Das bedeutet, dass die definierten Relationen durch den Anwender selbst überprüft werden. Da innerhalb der eigenen Überprüfung Fehler häufig unerkannt bleiben, werden die definierten Relationen zusätzlich gemeinsam mit einem Fachexperten untersucht. In einem gemeinsamen Gespräch kann dieser vor seinem Erfahrungshintergrund bestätigen, dass die definierten Relationen korrekt sind. Dieses Vorgehen beschreibt erneut die V&V-Technik Validierung im Dialog [RSW08a, S.109].

#### **4.2.2 Statistische Datenaufbereitung**

Im Anschluss an die Datenstrukturierung erfolgt in Bezug auf den Aufbereitungsschritt der VDI-Richtlinie 3633 die statistische Aufbereitung der Ergebnisdaten (vgl. 2.2.4). Wie in Abschnitt 2.2.4 beschrieben, werden mittels statistischer Verfahren auf der einen Seite die Ergebnisdaten eines einzelnen Laufes verdichtet und auf der anderen Seite die Daten der Replikationen zu aussagekräftigen Daten zusammengefasst. In diesem Prozessschritt erfolgt lediglich die Verdichtung der Daten über die Replikationen. Hierbei werden aus den, aufgrund stochastischer Eingangsgrößen schwankenden Ergebnisdaten, als korrekt anzunehmende Daten bestimmt (vgl. 2.2.4). Eine weitere, zum Eingang ins Schatten-DWH-System erforderliche Verdichtung der Daten innerhalb der zusammengefassten Replikationen, wird im nachfolgenden Prozessschritt Aggregation durchgeführt. Dies bedeutet, dass innerhalb dieses Prozessschrittes zum Beispiel die Umsatzdaten pro Monat der einzelnen Replikationen durch die Ermittlung der jeweiligen Mittelwerte pro Monat verdichtet werden. Die Aggregation der jeweiligen monatlichen Mittelwerte zu einem Jahreswert erfolgt dabei erst innerhalb des Aggregationsschrittes. Innerhalb der Kennzahlenberechnung im Schatten-DWH-System werden die Daten weiter aggregiert.

Abschließend erfolgt eine Ergebnisüberprüfung. Diese gliedert sich in zwei Abschnitte. Zunächst wird die in diesem Prozessschritt stattfindende statistische Aufbereitung durch eine entsprechende Verifikation überprüft. Anschließend erfolgt die Validierung der statistisch aufbereiteten Ergebnisdaten. Zur Überprüfung der Richtigkeit der Daten wird sowohl die statistische als auch die inhaltlich logische Richtigkeit der Daten untersucht (vgl. 4.1.1). Für zuerst genanntes wird die Gültigkeit der jeweiligen Modellannahmen, die für die Anwendung des statistischen Verfahrens getroffen wurden, überprüft. So ist für die statistische Aufbereitung beispielsweise in Form der Mittelwertbildung an dieser Stelle zu zeigen, dass die Ergebnisdaten der Replikationen der dafür notwendigen

Normalverteilung entsprechen (vgl. 2.2.4). Dies kann zum Beispiel durch die Erstellung eines Histogramms erfolgen. Die Überprüfungsmethode zählt zu den statistischen Techniken der V&V [RSW08a, S.103]. Außerdem sind die statistisch aufbereiteten Ergebnisdaten auf ihre inhaltliche Widerspruchsfreiheit hin zu untersuchen. Hierzu werden diese, beispielsweise in Bezug auf ihre Größenordnung, in einem Gespräch mit einem Fachexperten überprüft.

Anschließend werden die statistisch aufbereiteten Ergebnisdaten validiert. Dieser Teil beschreibt die in Kapitel 4.2 benannte Validierung der Ergebnisdaten im Anschluss an die statistische Aufbereitung. Zur statistischen Absicherung der Aussagekraft der statistisch aufbereiteten Ergebnisdaten wird beispielsweise für die errechneten Mittelwerte das Konfidenzintervall ermittelt (vgl. 2.2.4). Hierdurch wird aufgezeigt, wie genau der gewogene Mittelwert einer Größe geschätzt wurde. Auch dieses statistische Verfahren ist Teil der Methoden zur V&V [RSW08a, S.103]. Außerdem werden die statistisch aufbereiteten Ergebnisdaten durch den Vergleich mit real aufgezeichneten Daten validiert. Dies entspricht der V&V-Technik Vergleich mit aufgezeichneten Daten [RSW08a, S.111]. Basis für die Validierung sind die im Prozessschritt Erhebung der notwendigen Zieldaten (vgl. 4.1.4) erhobenen Zieldaten.

Mit Abschluss dieses Prozessschrittes ist die Aufbereitung der Ergebnisdaten im Sinne der VDI-Richtlinie 3633 beendet. An dieser Stelle liegen die selektierten und sortierten Ergebnisdaten in einer aussagekräftigen Form vor. Zum Eingang der Daten in das Schatten-DWH-System bedarf es einiger weiterer Aufbereitungsschritte. Innerhalb dieser werden die drei in Abbildung 18 aufgezeigten Abweichungsmöglichkeiten hinsichtlich Formatierung, Aggregationsgrad und Attribute zwischen den simulierten Ergebnisdaten der Simulation und den notwendigen Zieldaten zum Eingang in das DWH-System eliminiert.

### **4.2.3 Ergänzung fehlender Attribute**

Innerhalb dieses Prozessschrittes werden die Ergebnisdaten um die zuvor ermittelten fehlenden Attribute ergänzt (vgl. 4.1.2). Die fehlenden Attribute unterteilen sich in zwei verschiedene Kategorien. Für einen Teil der fehlenden Attribute ergeben sich die Werte durch die jeweilige aktuelle Situation im Logistiknetzwerk. Einen anderen Teil stellen Stammdaten dar, welche in den SimChain-Tabellen nicht vorgesehen sind. Grundsätzlich wird eine Integration der fehlenden Attribute in SimChain angestrebt. Hierzu muss der SimChain-Code durch einen Fachexperten entsprechend verändert werden. Anschaulich werden hierbei die Ergebnistabellen von SimChain um die fehlenden Attribute erweitert. (vgl. 4.1.2)

Ist dies nicht möglich, müssen die Werte zu den fehlenden Attributen auf eine andere Art und Weise erzeugt und anschließend in die simulierten Ergebnisdaten integriert

werden. Ein Attribut, dessen Werte sich durch die Abläufe im System ergeben, ist zum Beispiel die Umsatzart. Die Ausprägung der Umsatzart hängt davon ab, woher der verkaufte Artikel stammt. Dabei kann der verkaufte Artikel zum Beispiel sowohl aus dem eigenen Lager, als auch aus einem externen Zukauf kommen. Die Ausprägung der Umsatzart ergibt sich durch die jeweilige aktuelle Situation im Logistiknetzwerk. Solche Attribute können mit Hilfe einer Wahrscheinlichkeitsverteilung, welche auf historischen Daten des jeweiligen Attributes basieren, generiert werden [Fahr10, S.31 ff.]. Die Nutzung einer solchen empirischen Verteilung birgt allerdings die Gefahr der Verfälschung der Daten, wodurch die Aussagekraft der Daten reduziert werden kann. So passen die Ergebnisse der Verteilung möglicherweise nicht zu den Ergebnisdaten, da diese in Abhängigkeit von weiteren, in der Verteilung nicht berücksichtigten, Einflussgrößen stehen. Für das Beispiel der Umsatzart würde dies bedeuten, dass durch die Ausführung der empirischen Verteilung einem Artikel, der nie zugekauft wird, die Umsatzart externer Zukauf zugeordnet wird.

Weiterhin stellt eben die Abbildung solcher Abläufe den eigentlichen Nutzen der Simulation dar. Besonders in Bezug auf die Durchführung verschiedener Szenarien und der damit einhergehenden Veränderung des Simulationsmodells verlieren die empirischen Verteilungen ihre Gültigkeit. Dies resultiert aus der Tatsache, dass sich aufgrund der Veränderungen im System die Abläufe und somit auch möglicherweise die Ausprägungen des mittels der Verteilung dargestellten Attributes verändern würden. Diese Veränderungen sind in den auf Vergangenheitsdaten aufgestellten statistischen Verteilungen nicht berücksichtigt.

Für Attribute die Stammdaten darstellen können die Werte oftmals über ein, in den Ergebnistabellen vorhandenes Attribut, in die Ergebnisdaten integriert werden. Dieses übernimmt die Funktion eines Schlüssels, über den die Werte des fehlenden Attributes dem jeweiligen Datensatz der Ergebnisdaten zugeordnet werden können. Im Folgenden wird hierfür ein Beispiel beschrieben. Zur Berechnung der Kennzahlen im Schatten-DWH-System wird die Kundennummer benötigt. In SimChain werden die Kunden innerhalb der Tabelle customer erzeugt. In dieser kann allerdings nur der Kundenname und nicht die Kundennummer hinterlegt werden. Durch eine eindeutige Beziehung zwischen dem Kundennamen und der Kundennummer wird im Anschluss an die Simulation die Kundennummer dem Kundennamen zugeordnet und somit in die Ergebnisdaten integriert.

Ergebnis dieses Prozessschrittes sind die vollständigen Daten, welche zum Eingang in die gewählten Eingangscubes im Schatten-DWH-System benötigt werden. Diese werden hinsichtlich der definierten Relationen zwischen den vorhandenen und den zugeordneten Daten auf Unstimmigkeiten untersucht. In einem ersten Schritt werden die vorliegenden Daten im Sinne der V&V Technik Schreibtischtest [RSW08a, S.102]

durch eine eigene Überprüfung verifiziert. Zusätzlich werden die vollständigen Daten einem Fachexperten vorgelegt. Dieser kann auf Grundlage seines Erfahrungshintergrundes die Widerspruchsfreiheit vollständigen Daten bestätigen.

#### 4.2.4 Aggregation

Innerhalb dieses Prozessschrittes werden die simulierten Ergebnisdaten insofern aggregiert, als dass diese anschließend dem Aggregationsgrad der notwendigen Zieldaten entsprechen. Stimmt der Aggregationsgrad der simulierten Ergebnisdaten mit dem der notwendigen Zieldaten überein, so entfällt dieser Prozessschritt zur Aggregation.

Die Aggregation erfolgt entlang einer oder mehrerer Dimensionen der Daten. Ist die Klassifikationshierarchie der entsprechenden Dimensionen allgemeingültig, wie beispielsweise die der Zeit (vgl. 3.3), sind keine weiteren Informationen für den Aggregationsvorgang erforderlich. Hier ist die Beziehung zwischen den einzelnen Klassifikationsstufen immer gleich. So ist festgelegt, welche Tage zu welchem Monat, welche Monate zu welchem Quartal und welche Quartale zu welchem Jahr zugeordnet sind (vgl. Abbildung 11). Für den Fall, dass es sich um eine unternehmensspezifische Klassifikationshierarchie handelt, ist die Beziehung zwischen den Klassifikationsstufen nicht allgemein bekannt und von Unternehmen zu Unternehmen verschieden. Dies gilt beispielsweise für die Dimension Produkt. So weist jedes Unternehmen ein anderes Artikelsortiment auf, das einer jeweiligen Produktgruppe, Produktfamilie und Sparte zugeordnet wird (vgl. Abbildung 11). Zur Aggregation müssen die entsprechenden Klassifikationshierarchien bekannt sein.

Voraussetzung zur Aggregation der Daten sind Vollständigkeit, Disjunktheit und Typverträglichkeit der Daten (vgl. 3.4). Vollständigkeit und Disjunktheit werden an dieser Stelle vorausgesetzt. Die Typverträglichkeit ist gewährleistet, wenn die angewendete Aggregationsoperation mit dem Summationstypen der Kennzahlen kompatibel ist (vgl. 3.4). Demnach werden die zu verwendenden Aggregationsoperationen in Abhängigkeit vom jeweiligen Summationstypen der zu aggregierenden Kennzahl bestimmt. Der Summationstyp wird durch die jeweilige Kennzahlenart definiert. Allgemein wird zwischen Bestandskennzahlen, Bewegungskennzahlen sowie relativen Kennzahlen unterschieden (vgl. 3.1.2). In Bezug auf das multidimensionale Datenmodell werden dementsprechend die Summationstypen stock, flow und value per unit unterschieden (vgl. 3.4). Im folgenden Beispiel soll der Umsatz pro Kunde von Monatswerten auf Quartalswerte aggregiert werden. Der Umsatz stellt eine Bewegungskennzahl dar, die im Kontext des multidimensionalen Datenmodells als flow bezeichnet wird. Für Kennzahlen des Typs flow wird die Aggregation der Daten in Form einer Summation durchgeführt (vgl. 3.4). Durch die

Ausführung der Aggregationsoperationen werden die Ergebnisdaten auf die Aggregationsstufe der notwendigen Zieldaten transformiert.

Abschließend erfolgt eine Plausibilitätsprüfung. Für den Fall, dass die Aggregation in einer Summation der Daten besteht, werden die Daten durch die Berechnung der Summen der beiden relevanten Verdichtungsstufen verifiziert. Die Summen beinhalten alle Knoten der jeweiligen Verdichtungsstufe. Vor dem Hintergrund der Vollständigkeit und Disjunktheit der Daten ist die Summe aller Knoten einer Verdichtungsstufe gleich der Gesamtsumme (vgl. 3.4). Demnach ist an dieser Stelle aufzuzeigen, dass die Summen der Knoten innerhalb der beiden Verdichtungsstufen übereinstimmen. Ist zur Aggregation der Daten das Minimum, Maximum oder der Durchschnitt bestimmt worden, werden die aggregierten Daten auf Unstimmigkeiten hinsichtlich der jeweiligen Wertebereiche untersucht. Hierzu werden die aggregierten Daten gemeinsam mit einem Fachexperten, der über die entsprechenden Kenntnisse in Bezug auf die Wertebereiche der Daten verfügt, strukturiert überprüft. Dies beschreibt die V&V-Technik strukturiertes Durchgehen [RSW08a, S.104].

#### **4.2.5 Formatierung**

An dieser Stelle liegen die erzeugten Ergebnisdaten in Bezug auf die notwendigen Zieldaten in einer vollständigen und hinsichtlich des Aggregationsgrades passenden Form vor. Ein letzter Aufbereitungsbedarf besteht in der Formatierung der Daten in die zum Schatten-DWH-System kompatible Form (vgl. Abbildung 18).

Die simulierten Ergebnisdaten können sowohl hinsichtlich der verwendeten Datentypen, als auch in Bezug auf ihre Syntax von den notwendigen Zieldaten abweichen (vgl. 4.1.2). Innerhalb der Datenstrukturierung der Eingangsdaten wurden die Eingangsdaten in die für SimChain notwendige Form formatiert (vgl. 4.1.9). Innerhalb dieses Prozessschrittes werden diese Formatierungen teilweise rückgängig gemacht. Dies gilt vor allem für die Eingangsdaten, welche aus dem DWH-System extrahiert wurden.

Somit werden innerhalb dieses Prozessschrittes die Datentypen der vorliegenden Ergebnisdaten in das notwendige Zieldatenformat konvertiert. Bei Datentypen, die keine Zeichenkette enthalten, ist das Datenformat durch den jeweiligen Datentypen bereits definiert. Daher erfolgt für diese, einhergehend mit der Konvertierung des Datentyps, auch die Formatierung des Datenformats. Im Datentyp float wird beispielsweise ein Punkt als Trennzeichen zu den Dezimalstellen verwendet. Erfolgt eine Konvertierung vom Datentyp float zum Datentyp double, wird der Punkt automatisch durch ein Komma ersetzt, da dies durch den Datentyp double vorgegeben ist. Für Datentypen, die eine Zeichenkette enthalten, wie beispielsweise ein string, ist zusätzlich eine syntaktische Formatierung der Daten erforderlich. Diese beinhaltet zum Beispiel die Anpassung von Groß- und Kleinschreibung innerhalb der Zeichenkette.

Außerdem wird das durchgeführte Mapping rückgängig gemacht (vgl. 4.1.9). Das bedeutet, dass die Daten wieder in ihre ursprüngliche Form umgewandelt werden.

Die Vorgehensweise zur Aufbereitung der Ergebnisdaten schließt mit einer Datenvalidierung ab. In dieser wird überprüft, inwieweit die Daten den am Anfang der gesamten Vorgehensweise aufgestellten Informationsbedarf bezüglich der notwendigen Zieldaten zur Messung der DWH-Kennzahl abdecken. Dies erfolgt im Sinne der V&V-Technik Validierung im Dialog [RSW08a, S.109] in einem Gespräch mit einem IT-Fachexperten. Ergebnis dieses Prozessschrittes und somit der gemachten Vorgehensweise sind die aufbereiteten Ergebnisdaten in Form der notwendigen Zieldaten zur Berechnung der Kennzahl im Schatten-DWH-System.

### **4.3 Einbettung der entwickelten Vorgehensweise in das Simulationsvorgehensmodell zur kontinuierlichen V&V**

Zur Anwendung der entwickelten Vorgehensweise innerhalb einer Simulationsstudie wird diese in das Simulationsvorgehensmodell zur kontinuierlichen V&V von *Rabe et al.* (vgl. Abbildung 5) eingebettet. Dies wird in Abbildung 20 dargestellt. Das Simulationsvorgehensmodell stellt aufgrund der darin vorgenommen Trennung von Daten und Modell eine gute Basis für die Integration der entwickelten Vorgehensweise dar. Weiterhin wurde der IGP von *Bernhard et al.* in dieses Vorgehensmodell integriert (vgl. Abbildung 3).

Die Grundlage für die Durchführung einer Simulationsstudie ist die vom Auftraggeber entwickelte Zielbeschreibung. Vor dem Hintergrund, dass am Simulationsmodell komplexe DWH-Kennzahlen gemessen werden sollen, beinhaltet die Zielbeschreibung die zu messenden DWH-Kennzahlen. In der nachfolgenden Aufgabendefinition wird die Zielbeschreibung von Simulationsfachexperten sowie der Auftragnehmer-Seite präzisiert und vervollständigt. Dies bedeutet im Zusammenhang mit der Messung von DWH-Kennzahlen diese zu spezifizieren. Hierzu wird ermittelt, welche Ergebnis- und Zieldaten notwendig sind, um darauf aufbauend die DWH-Kennzahlen messen zu können. Demnach wird der Prozessschritt Zieldefinition bezüglich der notwendigen Daten zur Messung der DWH-Kennzahl in die Phase Aufgabendefinition eingeordnet (vgl. Abbildung 20). Die Aufgabenspezifikation enthält sowohl Informationen zum Modell als auch zu den Daten. Daher wird der Informationsbedarf bezüglich der notwendigen Daten der Aufgabendefinition zugeordnet. (vgl. 2.3 )

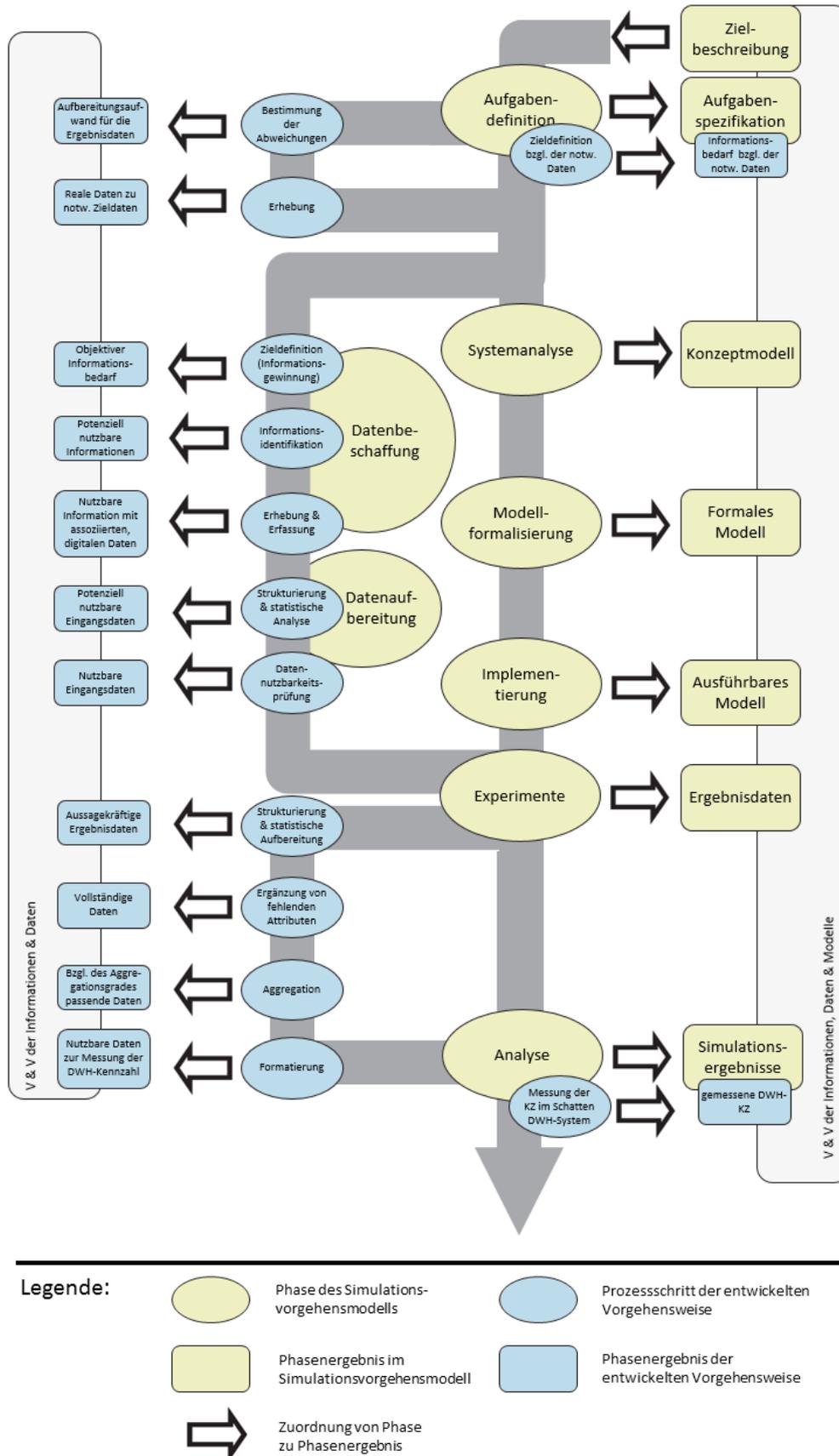


Abbildung 20: Integration der entwickelten Vorgehensweise in das Simulationsvorgehensmodell von Rabe et al. (i.A.a [JKW09, S.5])

Aufbauend auf dem in der Aufgabendefinition ermittelten Informationsbedarf werden die Abweichungen zwischen den notwendigen Ergebnisdaten und den notwendigen Zieldaten zum Eingang in das Schatten-DWH-System bestimmt. Es folgt die Erhebung der notwendigen Zieldaten. In Analogie zur Integration des IGP in das Simulationsvorgehensmodell werden diese beiden Prozessschritte graphisch in die linke Seite des Vorgehensmodells eingeordnet. Diese beschreibt den Daten- und Informationsaspekt innerhalb des Simulationsvorgehensmodells. Hiermit sind die ermittelten Abweichungen zeitlich vor der Modellbildung eingeordnet. Diese sind insofern Voraussetzung für die Modellbildung, als dass hier für den möglichen Fall fehlender Attribute, diese durch die Veränderung des SimChain-Codes in die Modellbildung integriert werden. Sowohl der Informationsbedarf bezüglich der notwendigen Zieldaten, als auch die zu den notwendigen Zieldaten zugehörigen realen Daten, stellen die Basis für die Bestimmung der Eingangsdaten innerhalb des IGP dar (vgl. 4.1.5).

Es folgt die Integration des IGP. Dieser wird entsprechend der bereits bestehenden Integration in das Simulationsvorgehensmodell (vgl. Abbildung 3) im Anschluss an die Aufgabendefinition zeitlich und organisatorisch parallel zur Modellbildung in das Simulationsvorgehensmodell eingeordnet. Hierbei werden die beiden Phasen Datenbeschaffung sowie Datenaufbereitung des Simulationsvorgehensmodells detailliert beschrieben und um die Informationsbeschaffung und -bewertung erweitert (vgl. 2.2.2.2).

Das Simulationsvorgehensmodell schließt mit der Phase Experimente und Analyse ab (vgl. Abbildung 5). Ergebnis dieser Phase sind die analysierten Simulationsergebnisse. Voraussetzung zur Durchführung dieser Phase ist die Verfügbarkeit der aufbereiteten Eingangsdaten sowie des ausführbaren Modells. Zur Einbettung der Vorgehensweise zur Aufbereitung der Ergebnisdaten wird die letzte Phase in zwei Teilphasen unterteilt. Dies ist erforderlich, da der Ausgangspunkt für die Vorgehensweise zur Aufbereitung der Ergebnisdaten, die Ergebnisdaten und nicht die bereits analysierten Simulationsergebnisse sind. In der Phase Experimente werden somit ausschließlich die Experimentpläne und die zu überprüfende Hypothesen festgelegt sowie die Experimente (vgl. 2.1) selbst durchgeführt. Ergebnis dieser Phase und demnach Zwischenergebnis der Phase Experimente und Analyse sind die Ergebnisdaten der Simulation. (vgl. 2.3) Die Verfügbarkeit der Ergebnisdaten ist die Voraussetzung für die nachfolgenden Prozessschritte zur Aufbereitung der Ergebnisdaten. Diese werden somit aufbauend auf der Phase Experimente in die Daten- und Informationsseite des Simulationsvorgehensmodells integriert. Die Verfügbarkeit der erzeugten Ergebnisdaten in Form der notwendigen Zieldaten ist wiederum Voraussetzung für die anschließende Messung der Kennzahl im Schatten-DWH-System. Da die Berechnung der Kennzahl eine Analyse der Daten darstellt, wird die Messung der Kennzahl im Schatten-DWH-

System in die Analyse Phase des Simulationsvorgehensmodells eingeordnet. Die daraus resultierende gemessene Kennzahl zählt demnach zu den Simulationsergebnissen. In den Vorüberlegungen zum Simulationsvorgehensmodell zur kontinuierlichen V&V war die Trennung der letzten Phase in zwei einzelne Prozessschritte enthalten (vgl. [SLR04, S.269]).

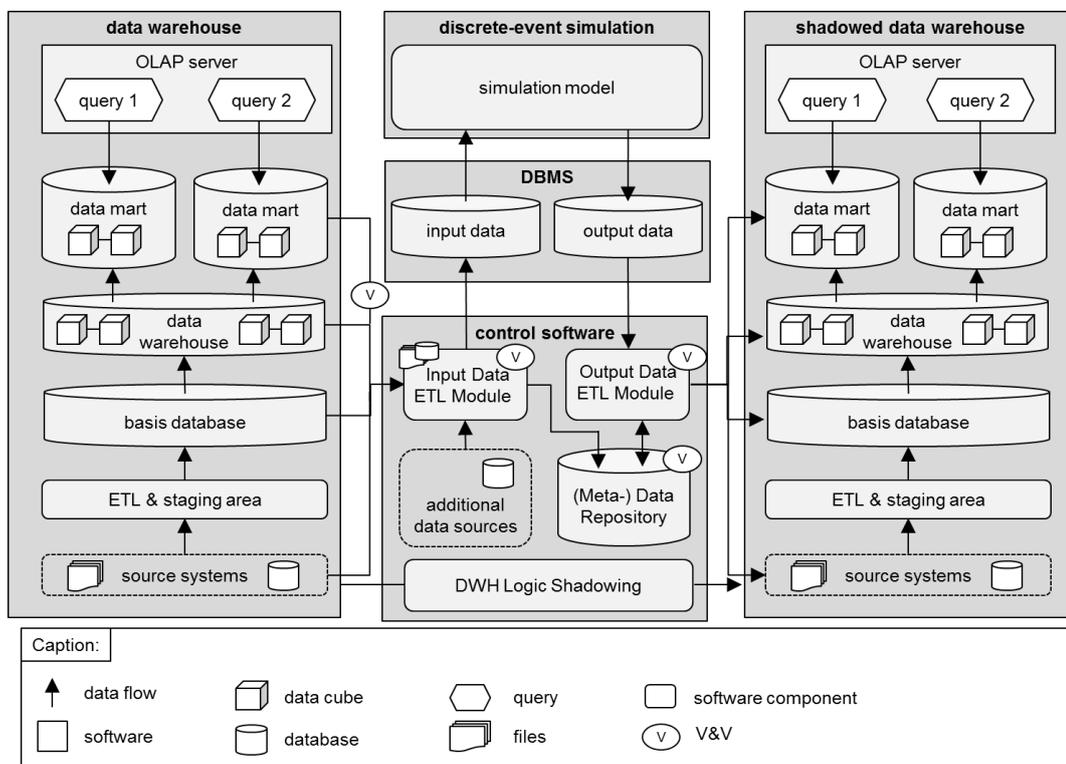
Grundsätzlich ist die Einbettung vor dem Hintergrund erfolgt, die Daten und Informationen entsprechend der Einbettung des IGP in die linke Seite des Simulationsvorgehensmodells zu integrieren. Die in der entwickelten Vorgehensweise erweiterten Schritte des IGP werden allerdings vor und nach der Modellbildung in das Simulationsvorgehensmodell eingebettet. Sowohl die Aufgabendefinition, als auch die Phase Experimente und Analyse enthalten Informationen über Daten und Modell. Daher wird die in der Einbettung des IGP eingeführte ausschließliche V&V der Modelle auf der rechten Seite (vgl. Abbildung 3) um die V&V der Informationen und Daten erweitert. Innerhalb der kontinuierlichen V&V werden die Prozessschritte nicht nur in Bezug auf sich selbst, sondern auch in Bezug auf die Ergebnisse der anderen Phasen untersucht. Somit sind die in Abbildung 20 aufgezeigten Prozessschritte durch V&V-Schritte miteinander vernetzt.

#### **4.4 Automatisierter Prozess**

Im Kontext der wachsenden Bedeutung von logistischen Assistenzsystemen ist es das Ziel den in der entwickelten Vorgehensweise beschriebenen Prozess zur glaubwürdigen Messbarkeit von komplexen DWH-Kennzahlen zu automatisieren. Allgemein stellen Assistenzsysteme rechnerbasierte Systeme dar, die den Menschen bei der Entscheidungsfindung und –durchführung unterstützen. Der automatisierte Prozess zur glaubwürdigen Messbarkeit von komplexen DWH-Kennzahlen repräsentiert allerdings nur die Grundlage für ein logistisches Assistenzsystem. In diesem werden ausschließlich Aussagen über die Kennzahlen und noch keine Handlungsempfehlungen zur Verfügung gestellt. Das in der Einleitung beschriebene zugrunde liegende Forschungsvorhaben [DrRa14] ist hingegen als logistisches Assistenzsystem einzuordnen. In diesem werden Handlungsempfehlungen zur Durchführung von Maßnahmen gegeben.[BCK+09, S.242 f.]

Durch den Ansatz logistischer Assistenzsysteme können Simulationen aufwandsarm in die operativen und taktischen Planungs- und Steuerungsinstrumente eingebunden werden. Für den hier betrachteten Prozess besteht die Einbindung in der Integration des Simulationsmodells sowie des kopierten DWH-System in die Planungs- und Steuerungssysteme des Unternehmens. Dies stellt eine laufende Planung auf Basis aktueller Daten sicher. [KGK+13]

Die Softwarearchitektur des automatisierten Prozesses zur glaubwürdigen Messbarkeit von komplexen DWH-Kennzahlen ist in Abbildung 21 dargestellt. Das zuvor als bestehendes beschriebene DWH-System ist auf der linken Seite der Abbildung in Anlehnung an die in Kapitel 3.2 beschriebene Architektur eines DWH-Systems erstellt worden. Im Zentrum der Abbildung befindet sich die Simulationsumgebung. Die ereignisdiskrete Simulation des Logistiknetzwerks erfolgt mittels eines datengetriebenen Simulationsmodells. Dieses setzt sich aus einem generischen Simulationsmodell und einer damit verbundenen Datenbank zusammen, über die das Modell initialisiert wird. Die Datenbank enthält alle notwendigen Eingangsdaten zum Aufbau und Lauf der Simulation. Weiterhin werden die Ergebnisdaten der Simulationsläufe in die Datenbank geladen. Die rechte Seite der Architektur-Abbildung zeigt das Schatten-DWH-System, in dem die Kennzahlenberechnung aufbauend auf den Ergebnisdaten der Simulation entsprechend der Logik des bestehenden DWH-Systems erfolgt. Das Zentrum der Softwarearchitektur bildet neben der Simulationsumgebung die control software. Diese beinhaltet verschiedene Softwarekomponenten in denen die in der entwickelten Vorgehensweise aufgezeigten Prozessschritte (vgl. Abbildung 13 und Abbildung 19) durchgeführt werden. Hierbei handelt es sich allerdings ausschließlich um die Prozessschritte, die für den Lauf und nicht für den Aufbau des automatisierten Prozesses relevant sind. Der erste automatisierte Schritt ist die Erhebung der Eingangsdaten. Alle zuvor stattfindenden Prozesse sind nur einmalig für den Aufbau des Systems notwendig.



**Abbildung 21: Automatisiertes System zur glaubwürdigen Messung von DWH-Kennzahlen**

Die beiden ETL-Module der control software beinhalten jeweils die Extraktion, Transformation sowie das Laden der Daten. Durch das Input Data ETL-Modul werden die Eingangsdaten aus den im Aufbau des automatisierten Prozesses ausgewählten Informationsquellen extrahiert und zwischengespeichert (vgl. 4.1.8). Diese sind in Abbildung 21 durch das bestehende DWH-System selbst sowie die Software Komponente, additional data sources, dargestellt. Im anschließenden Transformationsprozess werden die beiden Prozessschritte Datenstrukturierung und statistische Datenanalyse (vgl. 4.1.9 und 4.1.1) durchgeführt. Abschließend wird das Simulationsmodell parametrisiert. Hierzu werden die transformierten Ergebnisdaten in die Eingangsdatentabellen der Datenbank von SimChain geladen. Neben den Eingangsdaten werden mit Hilfe des Input Data ETL-Moduls ebenfalls die Informationen und Daten erhoben, die für die Aufbereitung der erzeugten Ergebnisdaten zum Eingang in das Schatten-DWH-System erforderlich sind. Hierzu zählen beispielsweise Formatierungsinformationen sowie Daten, die im Simulationsmodell nicht berücksichtigt werden. Diese werden in das (Meta-) Data repository geladen.

Mit Hilfe des Output Data ETL Modul werden die Ergebnisdaten der Simulation aus den Ergebnistabellen von SimChain extrahiert und mittels der im (Meta-)Data repository gespeicherten Daten entsprechend des notwendigen Zielformats für das Schatten-DWH-System transformiert. Dies umfasst alle in Abbildung 19 aufgezeigten Prozessschritte zur Aufbereitung der Ergebnisdaten. Anschließend werden die Daten an den zuvor gewählten Eingangscube im Schatten-DWH-System geladen. In Abhängigkeit vom verwendeten DWH-System kann ein Teil des Output Data ETL module in das Schatten-DWH-System integriert werden. Auf Grundlage der gepflegten Ergebnisdaten erfolgt analog zum bestehenden DWH-System die Berechnung der DWH-Kennzahl. Über das logic shadowing wird die Struktur des Schatten-DWH-Systems kontinuierlich aktualisiert. Ziel des automatisierten Prozesses ist es unter anderem, dass das Simulationsmodell durch die periodische Aktualisierung der Eingangsdaten die gegenwärtige Situation des Logistiknetzwerks widerspiegelt. Die periodische Aktualisierung gilt ebenfalls für die Daten des (Meta-) Data Repository.

Damit die glaubwürdige Messbarkeit der komplexen DWH-Kennzahlen auch im automatisierten System gewährleistet ist, muss die Glaubwürdigkeit des Systems nachgewiesen werden. Hierzu erfolgt die Integration entsprechender V&V-Schritten in den automatisierten Prozess. Grundsätzlich stellen die V&V-Schritte zur Automatisierung eine Untermenge der V&V-Schritte zum Aufbau des Systems dar. Weitere mögliche Fehlerquellen im automatisierten Prozess ergeben sich einerseits aus den Rückschlüssen, welche aus dem manuellen Aufbau des Systems gezogen werden können. Andererseits können beispielsweise mit Hilfe der Durchführung einer Fehlermöglichkeits- und Einflussanalyse (FMEA) potenzielle Schwachstellen des Systems identifiziert und somit Fehler vermieden werden. Die Notwendigkeit des

Handlungsbedarfs zur Durchführung von V&V-Schritten ergibt sich hierbei aus der resultierenden Risikoprioritätszahl. Diese spiegelt die Auftritts- und Entdeckungswahrscheinlichkeit sowie die Bedeutung des Fehlers wider. Je größer die Risikoprioritätszahl ist, desto mehr ist eine entsprechend detaillierter V&V notwendig. [Bund15]

In Abhängigkeit von der jeweiligen Fehlerart werden verschiedene V&V-Techniken eingesetzt. Ein Teil der genutzten V&V-Techniken kann automatisiert erfolgen. Dies gilt für die Überprüfungen, die mittels entsprechender Prüfroutinen in das System implementiert werden können. Mit Hilfe der Prüfroutinen ist allerdings nur eine bestimmte Vorarbeit hinsichtlich der Überprüfung der großen Datenmengen zu erzielen, da der Prüfvorgang nur bis zu einem bestimmten Grad automatisiert erfolgen kann. Für die restliche Überprüfung ist die punktuelle Einbindung des Anwenders notwendig. Grundsätzlich wird auch im automatisierten Prozess analog zur entwickelten Vorgehensweise eine durchgängige V&V durchgeführt werden, sodass eine frühzeitige Fehlererkennung möglich ist. Im Folgenden werden einige mögliche Fehlerquellen und Beispiele zur Umsetzung der V&V im automatisierten Prozess beschrieben.

Eine mögliche Fehlerquelle bildet der Datenübernahmeprozess aus den Informationsquellen in das Input Data ETL Modul ab. Diese Fehler würden im Rahmen einer FMEA eine hohe Risikoprioritätszahl aufweisen, da diese die Grundlage für alle weiteren Schritte bildet und daher entsprechend große Auswirkungen auf den gesamten Prozess hat. Im Datenübernahmeprozess besteht die Gefahr, dass die Daten aufgrund unterschiedlicher Datentypen und Datenformate im Quell- und Zielsystem nicht korrekt in das Input Data ETL Modul übernommen werden. Werden im verwendeten Zeichensystem der Informationsquellen beispielsweise Sonderzeichen genutzt, die im verwendeten Zeichensystem des Input Data ETL Moduls nicht codiert sind, so kann sich hierdurch eine Veränderung der Daten ergeben. Dies kann beispielsweise dazu führen, dass Artikelnummern, welche sich teilweise ausschließlich durch Sonderzeichen voneinander unterscheiden, durch den Übernahmeprozess in die gleiche Artikelnummer transformiert werden. Um dies zu vermeiden, wäre ein möglicher Lösungsschritt, zusätzlich zur Artikelnummer, die zugehörige Artikelbezeichnung zu übernehmen. Es wird eine Prüfroutine erstellt, in der die Zuordnung von Artikelnummer zur Artikelbezeichnung hinterlegt ist. Mit Hilfe dieser Prüfroutine ist es möglich fehlerhafte Artikelnummern zu identifizieren. Für den Fall fehlerhafter Artikelnummer werden diese vom System ausgegeben und durch den Anwender geprüft.

Da es sich beim DWH-System um ein lebendes System handelt, muss sichergestellt werden, dass die Daten an dem für sie vorgesehenen Speicherplatz hinterlegt wurden. Ebenso ist zu überprüfen, dass keine Leerdaten übertragen wurden. Hierzu werden in einem Gespräch mit einem Fachexperten Eigenschaften über die jeweiligen Daten

erarbeitet. Durch die Untersuchung der Daten in Bezug auf ihre Eigenschaften wird überprüft, ob die richtigen Daten an der richtigen Stelle im Input Data ETL Modul gespeichert wurden. Hierzu werden die Eigenschaften der Daten in einer Prüfroutine hinterlegt, durch welche die Überprüfung automatisiert erfolgt. Zu untersuchende Eigenschaften der Daten sind beispielsweise der Wertebereich, die Anzahl der Zeichen oder Ziffern sowie eine möglich statistische Verteilung der Daten. Für den Fall, dass die Prüfroutine Unstimmigkeiten in Bezug auf die Eigenschaften der Daten oder leere Datenfelder identifiziert, wird ein Fehlalarm ausgegeben. An dieser Stelle ist die Einbindung des Anwenders unumgänglich. Nur dieser kann die Ursache der Unstimmigkeiten ermitteln und beheben.

Es folgt die Betrachtung einiger V&V-Elemente innerhalb der Datenstrukturierung, welche ebenfalls im Input Data ETL Modul ausgeführt wird. Die zur Durchführung der Ausreißeranalyse eingesetzten statistischen Techniken der V&V werden im Rahmen des Transformationsschrittes des automatisierten Prozesses automatisiert. Die anschließende Bewertung der Repräsentativität der Daten, trotz festgestellter Qualitätseinbußen, kann hingegen nicht automatisiert werden. Dies erfolgt weiterhin durch die Begutachtung [RSW08a, S.97] des Managements von Auftraggeber und Auftragnehmer Seite (vgl. 4.1.9). Die Überprüfung der definierten Relationen durch das gemeinsame Durchgehen der Daten mit einem Fachexperten, im Sinne der V&V-Technik strukturiertes Durchgehen [RSW08a, S.104 f.], kann teilweise automatisiert erfolgen. Hierzu werden in einer zugehörigen Prüfroutine gültige Relationen hinterlegt. Wird innerhalb der Relationsdefinition beispielsweise einer Adresse das zugehörige Land zugeordnet, so kann dies durch die Prüfroutine untersucht werden. Für den Fall, dass die Prüfroutine einen Fehler entdeckt, muss eine Ausgabe des Systems an den Anwender erfolgen. Dieser untersucht den Fehler und die Ursache detailliert.

Auch im Output Data ETL Modul können einige V&V-Schritte teilweise automatisiert werden. Voraussetzung zur gültigen Mittelwertbildung über die Replikationen der Simulation ist die Normalverteilung der Ergebnisdaten. Um dies zu überprüfen werden statistische Techniken [RSW08a, S.103] eingesetzt. Hierbei kann beispielsweise automatisiert eine Verteilung über die Ergebnisdaten bestimmt und mit der Normalverteilung verglichen werde. Da der Vergleich jedoch qualitativ nur unzureichend durch einen automatisierten Prozess erfolgen kann, erfolgt an dieser Stelle eine Ausgabe der ermittelten Verteilung der Ergebnisdaten an den Anwender. Dieser beurteilt abschließend, ob die Verteilung der Daten hinreichend genau einer Normalverteilung entspricht. Des Weiteren kann der Modus und Median der Daten bestimmt und mit dem errechneten Mittelwert abgeglichen werden. Sind die Daten normalverteilt, so entsprechen Modus und Median dem Mittelwert.

Da das datengetriebene Simulationsmodell bei jeder Aktualisierung der Eingangsdaten neu initialisiert wird, ist im Anschluss jeweils zu überprüfen, ob die Ergebnisdaten sowie die berechneten Kennzahlen hinreichend genau sind. Hierzu wird die V&V Technik Vergleich mit aufgezeichneten Daten [RSW08a, S.111] eingesetzt. Das automatisierte System gibt sowohl die statistisch aufbereiteten Ergebnisdaten als auch die im Schatten-DWH-System berechneten DWH-Kennzahlen aus. Diese werden durch den Abgleich mit den entsprechenden realen Ausgabedaten und gemessenen Kennzahlen validiert.

In der Literatur existieren bereits umfangreiche Überlegungen zur V&V in der Informatik. Beispiele hierfür sind die Richtlinie des IEEE (Institute for Electrical and Electronics Engineers) sowie der Rational Unified Process. [SLR04, S.267]

## **5 Anwendung der Vorgehensweise am Bsp. bestehender DWH-Kennzahlen eines Werkstoffhändlers**

In diesem Kapitel wird die in Kapitel 4 beschriebene Vorgehensweise zur Aufbereitung von Eingangs- und Ergebnisdaten zur glaubwürdigen Messbarkeit von komplexen DWH-Kennzahlen anhand eines realen Fallbeispiels eines Werkstoffhändlers erprobt. Das Ziel ist die Validierung der Vorgehensweise und der darin enthaltenen Prozessschritte durch das Aufzeigen der glaubwürdigen Messung von einer DWH-Kennzahl in einem realen Anwendungsfall.

Hierzu wird zunächst in Kapitel 5.1 das vorliegende Fallbeispiel charakterisiert. Dies beinhaltet die Beschreibung der zu messenden DWH-Kennzahl, des verwendeten DWH-Systems sowie des derzeit entstehenden verwendeten Simulationsmodells. Anschließend werden die einzelnen Schritte der Vorgehensweise zur Aufbereitung der Eingangsdaten (vgl. 5.2.1) sowie der Ergebnisdaten (vgl. 5.2.2) für das Fallbeispiel durchlaufen. Für die Überprüfung der Vorgehensweise für die Eingangsdatengewinnung liegt der Schwerpunkt auf den erweiterten Schritten des IGP von *Bernhard et al.* (vgl. 4.1), da der IGP selbst bereits ein etabliertes und validiertes Vorgehensmodell darstellt. Zur Durchführung der Vorgehensweise wurden teilweise Microsoft Excel und gemäß der Ausrichtung der Vorgehensweise die Simulationssoftware SimChain verwendet. Im Rahmen der Validierung werden sowohl das Fallbeispiel-Unternehmen, die betrachteten Artikel sowie die Lagerstandorte anonymisiert. Weiterhin wird die Beschreibung der Kennzahlenberechnung im DWH-System abstrahiert.

### **5.1 Charakterisierung des Fallbeispiels**

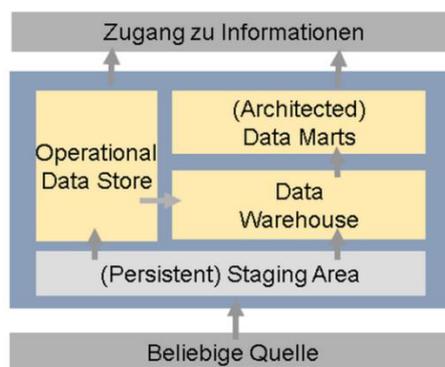
Das Fallbeispiel-Unternehmen ist durch ein großes und komplexes Logistiknetzwerk gekennzeichnet. Dieses umfasst eine Vielzahl von Lieferanten und Kunden sowie zentralen, regionalen und lokalen Lagern. Zur Lenkung und Kontrolle der Betriebs- und Logistikprozesse werden im Fallbeispiel-Unternehmen spezielle und komplexe Kennzahlensysteme verwendet. Durch die Nutzung eines DWH-Systems werden die Kennzahlen periodisch für die aktuellen Daten berechnet und den verantwortlichen Führungskräften in Form von monatlichen Berichten zur Verfügung gestellt. Zur Verbesserung der einzelnen Kennzahlen erfolgt die Durchführung verschiedener Maßnahmen. Aufgrund der komplexen Strukturen der Kennzahlensysteme sowie des Logistiknetzwerks selbst, bestehen unklare Interdependenzen zwischen den Kennzahlensystemen. Dies gilt auch für die Auswirkungen der Maßnahmen auf die Kennzahlen. So kann eine Maßnahme zur Verbesserung einer Kennzahl die Verschlechterung einer anderen Kennzahl zur Folge haben. Zur effizienten Steuerung

seines Logistiknetzwerks ist das Fallbeispiel-Unternehmen bestrebt diese unklaren Interdependenzen in Bezug auf die Auswirkungen der Maßnahmen auf die verschiedenen Kennzahlen aufzuzeigen. Dies wird durch das System, welches mittels der in dieser Arbeit entwickelten Vorgehensweise zur Messung von DWH-Kennzahlen aufgebaut wird, ermöglicht. Durch die Veränderung des erstellten Simulationsmodells, entsprechend der Durchführung bestimmter Maßnahmen, können Vorhersagen über die Entwicklung der einzelnen Kennzahlen getroffen werden.

Für das Beispiel-Unternehmen wird die Vorgehensweise für eine Kennzahl, welche die Rentabilität des Lagerbestandes (RDLB) angibt, erprobt. Diese sagt aus, wie viel Euro Rohgewinn pro eingesetzten Euro Lagerwert erwirtschaftet wurde. Außerdem wird die Kennzahl Lieferzuverlässigkeit in die Erprobung mit einbezogen. Diese umfasst sowohl die termingerechte als auch die vollständige Lieferung der Ware an den Kunden (On Time In Full, OTIF).

### 5.1.1 SAP NetWeaver Business Warehouse

Im Fallbeispiel-Unternehmen wird das DWH-System von SAP, das sogenannte SAP NetWeaver Business Warehouse (SAP BW), eingesetzt. Abbildung 22 bietet einen groben Überblick über die Architektur des SAP BW.



**Abbildung 22:** Architektur des SAP Business Warehouse [SAP15]

Im Gegensatz zu der in Abschnitt 3.2 aufgezeigten Referenzarchitektur eines DWH-Systems, wird im SAP BW die Datenbank, welche die integrierten, flexiblen und nicht auswertungsorientierten Daten beinhaltet, nicht als Basisdatenbank, sondern als Data Warehouse bezeichnet (vgl. 3.1.1). Die auf Analysezwecke fokussierte Datenbank, welche in der Referenzarchitektur unter den Begriffen Data Warehouse und Data Mart zusammengefasst ist, wird im SAP BW wiederum als Architected Data Marts benannt (vgl. 3.1.1). Die übrigen Schichten der in Abbildung 22 aufgezeigten Architektur des SAP BW Systems entsprechen denen der in Kapitel 3.2 aufgezeigten Referenzarchitektur eines DWH-Systems (vgl. Abbildung 10). So werden die Daten aus einer beliebigen Quelle extrahiert und in die hier als Persistent Staging Area, kurz PSA, bezeichnete Eingangsschicht des SAP BW geladen. Analog zum ETL-Prozess in der

Referenzarchitektur (vgl. Abbildung 10) werden die Daten aus der PSA unter Durchführung einer Transformation und Bereinigung der Daten in eine einheitliche und integrierte Sicht in die nächste Schicht des SAP BW geladen. Die in Abbildung 22 dargestellte Schicht Zugang zu den Informationen repräsentiert innerhalb der Referenzarchitektur die Datenbereitstellungsebene sowie die Präsentationsschicht. Die Komponente Operational Data Store ist eine Komponente zur operativen Datenanalyse und kann daher für diese Arbeit vernachlässigt werden. Die Gleichsetzung erfolgt vor dem Hintergrund, dass die Schichtenarchitektur des SAP BWs Systems lediglich einen konzeptionellen Ansatz darstellt und somit die Grenzen dieser Schichten organisationspezifisch sind. [SAP15]

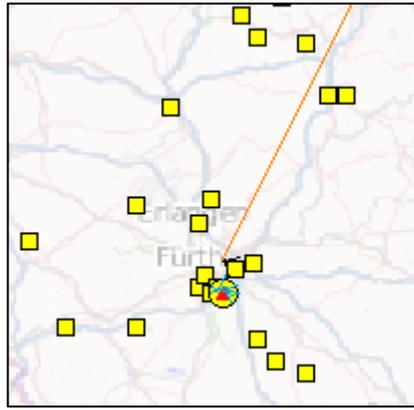
Die in der Vorgehensweise als Cubes bezeichneten multidimensionalen Datenwürfel werden im SAP-BW-Umfeld als Infoprovider bezeichnet. Diese dienen der Speicherung sowie der Zusammenführung von Daten. Es werden verschiedene Infoprovider unterschieden. Für diese Arbeit sind nur die DataStore-Objekte (DSOs) relevant. Die Attribute zur eindeutigen Identifizierung eines Datensatzes innerhalb eines DSO werden als Schlüsselfelder bezeichnet. Der Datenfluss zwischen den Infoprovicidern erfolgt im SAP BW über sogenannte Datentransferprozesse. Diese steuern die Selektion und Verbuchung der Daten. [WoYa10, S.67 ff.]

Die Implementierung der Kennzahlenberechnung des RDLBs stellt einen Data Mart da und ist somit in die Architected Data Mart-Schicht des SAP-BW einzuordnen (vgl. Abbildung 22). Die Kennzahlberechnung umfasst mehrere DSOs. Die Auswertung der DSO erfolgt mittels BEx Queries. Diese können mit Hilfe des BEx Query Designer erstellt werden [WoYa10, S.69].

### **5.1.2 Simulationsmodell**

Das aufgebaute Simulationsmodell wurde mittels des Simulationswerkzeugs SimChain erstellt. Da es sich hierbei um eine datenbankgetriebene Simulation handelt, erfolgt der Aufbau des Simulationsmodells erst zur Laufzeit (vgl. 2.4). Es werden daher sowohl die für den Lauf des Modells notwendigen Eingabedaten, als auch die zum Aufbau des Simulationsmodells erforderlichen internen Modelldaten in die Eingangsdatentabellen der verknüpften MySQL-Datenbank eingepflegt (vgl. Abbildung 1 und Abbildung 6). Die SimChain-GUI wird demnach nicht genutzt.

Im erstellten Simulationsmodell wird ein Teilausschnitt des Logistiknetzwerks des Beispiel-Unternehmens dargestellt. Dabei werden der Standorte A sowie die Artikel der Gruppe A und B über einen Zeitraum von einem Jahr betrachtet. In Abbildung 23 ist ein Teilausschnitt des Logistiknetzwerks abgebildet.



**Abbildung 23:** Screenshot - Teilausschnitt des Simulationsmodells

## 5.2 Anwendung der entwickelten Vorgehensweise

In diesem Kapitel wird die Funktionsweise der entwickelten Vorgehensweise anhand der zwei zuvor vorgestellten DWH-Kennzahlen des charakterisierten Beispiel-Unternehmens demonstriert. Der Fokus liegt dabei auf der Kennzahl zur Messung der Rentabilität des Lagerbestandes. Mit Hilfe der zweiten Kennzahl OTIF kann die Vorgehensweise in ihrem gesamten Umfang erprobt werden. So existieren verschiedene Punkte innerhalb der Vorgehensweise an denen weitere zu messende Kennzahlen zu beachten sind (vgl. 4.1.1). Die Anwendung erfolgt entsprechend der in den Abbildung 13 und Abbildung 19 aufgezeigten Prozessschritte der entwickelten Vorgehensweise. Weiterhin werden Empfehlungen und Hinweise in Bezug auf die Umsetzung des aufgebauten Systems in ein automatisiertes Szenario gegeben.

Entsprechend der Zweiteilung der entwickelten Vorgehensweise, erfolgt in Kapitel 5.2.1 zunächst die Anwendung der Vorgehensweise für die Eingangsdaten und anschließend in Kapitel 5.2.2 für die Ergebnisdaten.

### 5.2.1 Aufbereitung der Eingangsdaten zur glaubwürdigen

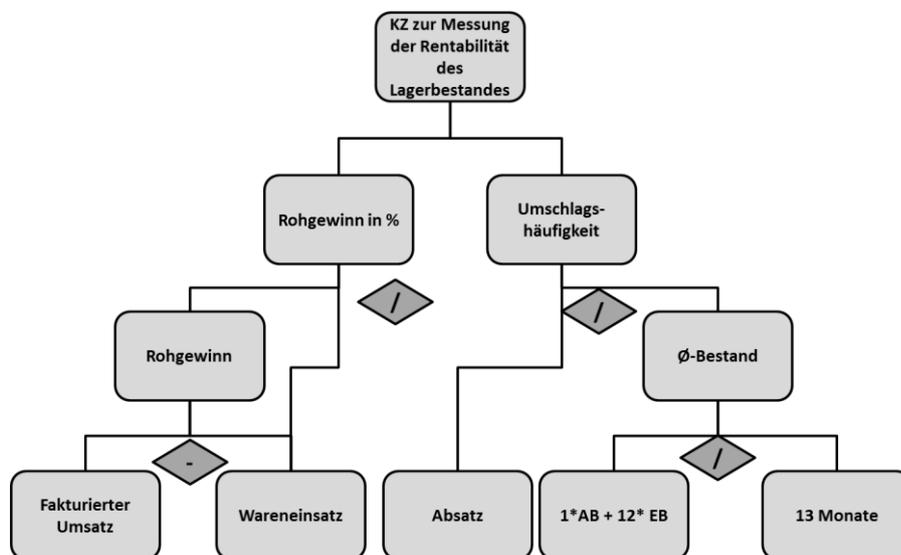
#### Messbarkeit der komplexen DWH-Kennzahlen RDLB und OTIF

##### 5.2.1.1 Zieldefinition der Ergebnisdaten

Im ersten Prozessschritt Zieldefinition bezüglich der notwendigen Daten zur Messung der DWH-Kennzahl ist der Informationsbedarf für die notwendigen Daten zur Messung der DWH-Kennzahlen zu bestimmen. Hierzu gehört sowohl die Ermittlung der notwendigen Ergebnisdaten der Simulation, die durch die Ergebnistabellen von SimChain bestimmt werden, als auch die notwendigen Zieldaten zum Eingang ins Schatten-DWH-System, welche durch den gewählten Cube definiert werden (vgl. Abbildung 15).

Zu Anfang ist für die betrachteten DWH-Kennzahlen eine allgemeine Analyse durchzuführen. Ausgehend von einem vorliegenden Kennzahlensteckbrief über die

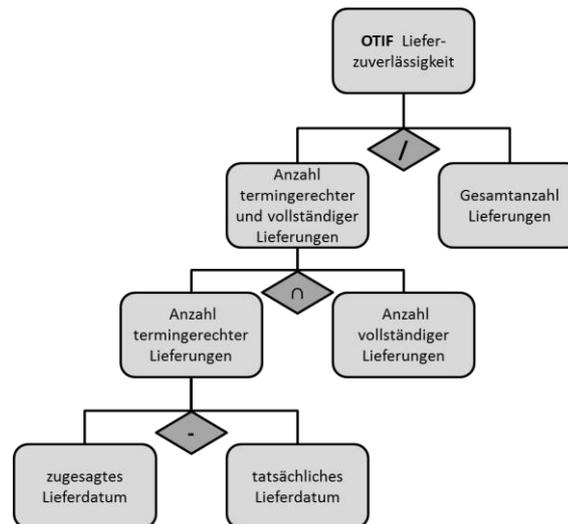
Kennzahl zu Messung der Rentabilität des Lagerbestandes (vgl. Anhang Abbildung 1) erfolgte die Zerlegung der zugehörigen Berechnungsvorschrift. Diese setzt sich aus der Umschlagshäufigkeit und dem Rohgewinn in % zusammen. Die weitere Gliederung dieser beiden Komponenten und der daraus resultierenden Komponenten wird ebenfalls auf Grundlage entsprechender Kennzahlensteckbriefe durchgeführt. Das Ergebnis der Zergliederung der Berechnungsvorschrift der Kennzahl RDLB wurde in Form eines Kennzahlenbaumes (vgl. Abbildung 24) graphisch veranschaulicht. Neben der Berechnungsvorschrift sind die Attribute der Kennzahl zu ermitteln. Diese konnten ebenfalls dem vorliegenden Kennzahlensteckbrief (vgl. Anhang Abbildung 1) entnommen werden. Die RDLBs wird pro Artikel und pro Standort für einen Betrachtungszeitraum von einem Jahr jährlich berechnet.



**Abbildung 24: Kennzahlenbaum für die Kennzahl RDLB**

Ebenso wurde für die Kennzahl OTIF ein entsprechender Kennzahlenbaum erstellt (vgl. Abbildung 25). Hierfür fand ein Austausch mit einem Fachexperten für die OTIF-Kennzahl statt (vgl. Anhang Abbildung 3). Die Attribute der Kennzahl OTIF sind der Lagerstandort und das Jahr.

Die beiden Kennzahlensysteme der betrachteten Kennzahlen RDLB und OTIF weisen auf den ersten Blick keine Überschneidungen auf. Es muss aber beachtet werden, dass der fakturierte Umsatz und die Lieferungen in Zusammenhang stehen.



**Abbildung 25: Kennzahlenbaum für die Kennzahl OTIF**

Wie in der entwickelten Vorgehensweise an dieser Stelle gefordert, wurden im Anschluss an die allgemeine Analyse die notwendigen Ergebnisdaten und die notwendigen Zieldaten für die Kennzahlenmessung bestimmt. Hierzu wurden zunächst, aufbauend auf den zuvor gewonnenen Kenntnissen über die Kennzahlen, deren Berechnungen im SAP BW-System in mehreren Gesprächen mit einem Fachexperten der IT untersucht (vgl. 4.1.1). Die Grundlage für die Untersuchung der Berechnung des RDLBs im SAP BW-System bestand in einem entsprechenden Datenmodell über den RDLB (vgl. Anhang Abbildung 4).

Im Folgenden wurden durch die Top-Down-Analyse der Kennzahlenberechnung im SAP BW-System und einem zeitgleichen Abgleich mit den Ergebnistabellen von SimChain sowohl die erforderlichen Ergebnistabellen als auch die Eingangscubes im DWH-System in Abhängigkeit voneinander bestimmt. Dabei wurden zunächst die beiden DSOs 1 und 2 innerhalb der Kennzahlenberechnung des RDLBs ausgewählt (vgl. Anhang Abbildung 4). In DSO 1 sind unter anderem die Daten über den Umsatz und den Wareneinsatz und in DSO 2 über die Umschlagshäufigkeit enthalten. Die zugehörige ausgewählte SimChain-Tabelle ist in Abbildung 26 abgebildet. Die für die Kennzahlenmessung relevanten Daten aus dieser Tabelle sind einerseits die Umschlagshäufigkeiten der einzelnen Artikel am jeweiligen Lagerstandort. In SimChain werden die Artikel als Stock Keeping Unit (SKU) und der Lagerstandort als Location bezeichnet. Andererseits ist die Anzahl der Artikel relevant, die beim Kunden eingegangen sind und in SimChain als NumItems-Inbound bezeichnet werden. Letzteres dient zur Berechnung von Umsatz und Wareneinsatz, welche in SimChain nicht abgebildet werden. Ziel ist es an dieser Stelle die beiden Basiskennzahlen Umsatz und Wareneinsatz in den SimChain-Code zu integrieren, sodass deren Werte den Ergebnisdamentabellen zu entnehmen sind. Da dies innerhalb dieser Arbeit nicht möglich

war, erfolgte die Generierung von Umsatz und Wareneinsatz durch die Integration von Verkaufs- und Einkaufspreis in die erzeugten Ergebnisdaten.

<b>ScenarioMan_Id_Scenario</b>	<b>INT</b>	<b>11</b>
<b>IdRun</b>	<b>INT</b>	<b>11</b>
<b>Location</b>	<b>VARCHAR</b>	<b>50</b>
<b>SKU</b>	<b>VARCHAR</b>	<b>45</b>
TotalCosts_Euro	DOUBLE	
ProductionCosts_Euro	DOUBLE	
WorkingCapitalCosts_Stock_Euro	DOUBLE	
WorkingCapitalCosts_Transport_Inbound_Euro	DOUBLE	
CustomsCosts_Euro	DOUBLE	
Num_Carrier_Inbound	INT	11
Num_Carrier_Outbound	INT	11
NumItems_Inbound	INT	11
NumItems_Outbound	INT	11
Avg_Inventory_Items	FLOAT	
Avg_Inventory_Carrier	FLOAT	
Avg_DaysOfInventory_Days	FLOAT	
Avg_Backlog_Production_Items	INT	11
A_ServiceLevel_Inbound	FLOAT	
B_ServiceLevel_Inbound	FLOAT	
TurnoverFrequency	FLOAT	
Label_1	VARCHAR	45
Label_2	VARCHAR	45
TotalCO2Emission	FLOAT	

**Abbildung 26: Screenshot – Spalten der ausgewählten Ergebnistabelle stat\_summary\_skus zur Messung des RDLBs**

Des Weiteren musste zur Verwendung der von SimChain berechneten Umschlagshäufigkeit zunächst geklärt werden, ob diese mit der Berechnungsvorschrift des Beispiels-Unternehmens für die Umschlagshäufigkeit übereinstimmt. In Abbildung 27 ist die Berechnung der Umschlagshäufigkeit in SimChain erläutert. Diese wurde mit der Berechnungsvorschrift für die Umschlagshäufigkeit im Beispiel-Unternehmen abgeglichen. Hierzu diente wiederum ein entsprechender Kennzahlensteckbrief. Hierbei hat sich ergeben, dass die beiden Berechnungsvorschriften nicht übereinstimmen. Demnach konnte die von SimChain berechnete Umschlagshäufigkeit nicht genutzt werden. Entsprechend der Berechnung im Beispiel-Unternehmen werden daher der durchschnittliche Bestand und die abgesetzte Menge zur Berechnung der Umschlagshäufigkeit genutzt. Die entsprechenden Daten sind ebenfalls in der bereits ausgewählten Ergebnistabelle enthalten. Die abgesetzte Menge entspricht dabei in SimChain den NumItems\_Outbound.

```

m_EvalTurnoverRateInventory:
For all SKUs
Calculate consumption of SKU (according
t_SKUStatList) = Inventory –
Inventory prev. day –
production amount prev. day –
quantity received prev. day

result = sum of usage for all simulation days
(normalised to a yearly value)

TurnoverFrequency = consumption / average daily stock

The average value for all SKUs (in warehouse source)

```

**Abbildung 27: Berechnung der Umschlagshäufigkeit in SimChain [FeGu14, S.197]**

Durch die Verwendung der ausgewählten SimChain-Tabelle (Abbildung 26) und den DSOs 1 und 2 wird ein minimaler Aufbereitungsaufwand zum Eingang der Ergebnisdaten in das Schatten-DWH-System sowie auch innerhalb des Schatten-DWH-Systems erzielt.

Aufgrund der besonderen Struktur der Berechnung des RDLBs im SAP BW-System wurde von einem IT-Fachexperten ein Vorschlag zur Performanceoptimierung der vorgestellten Lösung beschrieben. Aufbauend auf dem DSO 1, in dem die Umsatz- und Wareneinsatzdaten enthalten sind, erfolgen innerhalb des Transformationsschrittes zur Berechnung des Rohgewinns und dem Rohgewinn in % viele weitere Rechenschritte, welche für die Berechnung des RDLBs nicht relevant sind. Diese würden ebenfalls für die in das DSO 1 geladenen Ergebnisdaten durchgeführt. Hierdurch würde ein unnötiger Performanceverlust entstehen. Daher bestand der Vorschlag des IT-Experten in der Erzeugung eines zusätzlichen Eingang-DSO-Objekts (vgl. Anhang Abbildung 4)) in das die aufbereiteten Ergebnisdaten der Simulation geladen werden. Von hier aus werden die Daten direkt in den DSO 3 geladen (vgl. Anhang Abbildung 4). Dieser ist den DSO 1 und 2 in der Kennzahlenberechnung übergeordnet. In diesem sind unter anderem die Umschlagshäufigkeit und der Rohgewinn in % enthalten. Die Berechnung des Rohgewinns und des Rohgewinns in % sowie die Bestimmung der Umschlagshäufigkeit wird durch einen dazwischen liegenden Transformationsschritt erreicht.

Der vom IT-Fachexperten beschriebene Eingangscube definiert die notwendigen Zieldaten und die zugehörigen erforderlichen Attribute. Diese werden in der Anwendung des nächsten Prozessschrittes beschrieben. Durch die ausgewählte

Ergebnistabelle von SimChain (vgl. Abbildung 26) werden die notwendigen zu verwendenden Ergebnisdaten bestimmt.

Weiterhin ist nach der Vorgehensweise die mögliche Überschneidung der Berechnungen verschiedener Kennzahlen im DWH-System zu beachten. Für die Berechnung der beiden betrachteten Kennzahlen RDLB und OTIF ergibt sich keine Überschneidung der Implementierung im SAP BW.

### **5.2.1.2 Ermittlung der Abweichungen zwischen notwendigen Ergebnisdaten und den notwendigen Zieldaten**

Im nächsten Prozessschritt der entwickelten Vorgehensweise sind die Abweichungen zwischen den notwendigen Ergebnisdaten und den notwendigen Zieldaten zu bestimmen. Für diesen Anwendungsfall ist der Eingangscube das neu erzeugte Eingangs-DSO. Entsprechend der Vorgehensweise wurden die Abweichungen in Bezug auf die fehlenden Attribute, die Aggregationsstufe sowie die Formatierung der Daten ermittelt.

Die notwendigen Attribute zur Berechnung des RDLBs, innerhalb des neu erzeugten Eingangs-DSO, entsprechen den notwendigen Attributen im DSO 3. Die notwendigen Schlüsselfelder (vgl. 5.1.1) zur Berechnung des RDLBs umfassen unter anderem das Geschäftsjahr, den Lagerstandort, den Artikel, die Sparte und den Buchungskreis (vgl. Anhang Abbildung 5). Wie in Abbildung 26 abgebildet, sind in der ausgewählten Ergebnistabelle sowohl der Lagerstandort, als auch der Artikel angegeben. Demnach wurden sowohl die Sparte, als auch der Buchungskreis als fehlende Attribute identifiziert. Da sich die Sparte aus dem Material ergibt und die Untersuchung auf einen bekannten Buchungskreis beschränkt wurde, konnten diese im Folgenden ohne Einschränkungen in die Ergebnisdaten integriert werden (vgl. 5.2.2). Weiterhin konnte die Angabe zum Geschäftsjahr frei gewählt werden.

Hinsichtlich der Aggregationsstufe der notwendigen Daten wurde kein Unterschied festgestellt. So werden der Umsatz, der Wareneinsatz und die Umschlagshäufigkeit laut den Schlüsselfeldern des DSO 3 pro Lagerstandort und Geschäftsjahr benötigt. Innerhalb dieser Erprobung der Vorgehensweise wurde nur ein Geschäftsjahr simuliert. Da die ausgewählte Ergebnistabelle die Werte für den gesamten Betrachtungszeitraum zusammenfasst, ist demnach in Bezug auf die Zeit keine Aggregation notwendig. Im Folgenden wird aufgezeigt, dass der Warenausgang mit den gesamten Lieferungen, die bei den Kunden eingehen, übereinstimmt. Daher kann zur Berechnung von Umsatz und Wareneinsatz auch der Warenausgang des Lagerstandortes genutzt werden. Dieser wird innerhalb der ausgewählten Ergebnistabelle als NumItems-Outbound bezeichnet (vgl. Abbildung 26).

In einem nächsten Schritt wurden die notwendigen Daten hinsichtlich Abweichungen bezüglich der Formatierungen untersucht. In Tabelle 1 sind die ermittelten Abweichungen zwischen den Datentypen aufgezeigt. Die Datentypen sind der ausgewählten SimChain-Tabelle (vgl. Abbildung 26) sowie der Übersicht über die Schlüsselfelder des Eingangs-DSOs (vgl. Anhang Abbildung 5) entnommen worden. Für alle zu ergänzenden Attribute erfolgte die direkte Speicherung im notwendigen Datentyp zum Eingang in das Eingangsdaten-DSO. Abweichungen in Bezug auf die syntaktische Formatierung wurden an dieser Stelle nicht festgestellt.

**Tabelle 1: Abweichungen der Formatierungen**

Attribut	Datentyp in SimChain	Datentyp im SAP BW
durchschnittlicher Bestand	float	dec
Lagerstandort	varchar	char
Artikel	varchar	char

### 5.2.1.3 Erhebungsplanung für die notwendigen Zieldaten

Im nächsten Schritt wurde die Vorbereitung der Erhebung der realen Zieldaten durchgeführt. Hierzu soll eine Datenbankabfrage aufbauend auf dem entsprechend ausgewählten Eingangscube im DWH-System erstellt werden. Da im Anwendungsbeispiel der ausgewählte Eingangscube im Schatten-DWH zusätzlich erzeugt wird, können aus diesem keine Daten extrahiert werden. Das DSO 3 enthält ebenfalls die notwendigen realen Zieldaten. Daher sollte die BEx Query (vgl. 5.1.1) aufbauend auf dem DSO 3 definiert werden. Da diese BEx Query A bereits im Beispiel-Unternehmen existiert, war eine entsprechende Erstellung nicht notwendig. Die BEx Query A wird monatlich durch die Mitarbeiter des Beispiel-Unternehmens überprüft.

### 5.2.1.4 Erhebung der realen Zieldaten

Zur Erhebung der realen Zieldaten wurde lediglich die BEx Query A ausgeführt. Zur Überprüfung der erhobenen Daten wurde für diese, auf Grundlage der in der allgemeinen Analyse ermittelten Berechnungsvorschrift, die RDLBs statisch berechnet. Anschließend sind die nachberechneten Werte der RDLBs durch einen Abgleich mit den entsprechenden gemessenen Werten der RDLBs im DWH-Systems verifiziert worden.

### 5.2.1.5 Zieldefinition bezüglich der notwendigen Eingangsdaten

Im Anschluss an die Ermittlung der notwendigen Daten zur Messung der Kennzahl RDLB wurden im Folgenden die entsprechenden notwendigen Eingangsdaten für die Simulation bestimmt.

Zur Ermittlung der notwendigen und nutzbaren Eingangsdaten ist zunächst der notwendige objektive Informationsbedarf für die Eingangsdaten zu definieren (vgl. 4.1.5). Hierzu ist eine systematische Analyse der Entstehung der notwendigen Ergebnisdaten erfolgt. Die notwendigen Ergebnisdaten sind durch die ausgewählte Ergebnistabelle (Abbildung 26) bestimmt. Dies waren ausschließlich der durchschnittliche Bestand und die ausgehende Menge. Aus letzterem wurde im Folgenden der Umsatz und Wareneinsatz berechnet. Innerhalb der Analyse der Ergebnisdaten wurden die Elemente und Prozesse identifiziert, die zur Abbildung der notwendigen Ergebnisdaten benötigt werden.

Die Ermittlung der Elemente und Prozesse ist vor dem Hintergrund erfolgt, dass das Simulationsmodell nur eine Abstraktion des realen Systems darstellt (vgl. 2.1). So galt hier beispielsweise die vereinfachte Annahme, dass der wertmäßige Lagerabgang den Wareneinsatz darstellt und dieser wiederum linear mit dem Umsatz einhergeht. Der Lagerabgang korreliert dabei mit der Kundennachfrage. Dies bedeutet, dass innerhalb des Simulationsmodells jede Nachfrage eine entsprechende Lieferung impliziert. Jegliche Sonderfälle wurden nicht betrachtet. Ebenso wurde die Bestellpolitik vereinfacht dargestellt.

Durch Gespräche mit Fachexperten des Beispiel-Unternehmens und unter Verwendung der Checklisten von *Baron et al.* [BDKR01, S.155 ff.] (vgl. 4.1.5.2) wurden die notwendigen Eingangsdaten, die zur Erzeugung der notwendigen Ergebnisdaten erforderlich sind, identifiziert. Nachfolgend werden die wichtigsten Elemente und Prozesse benannt. Hierzu zählen unter anderem Daten über die Lieferanten und Kunden. Für diese werden beispielsweise die Adressen und für die Kunden zusätzlich die zugehörige Nachfrage benötigt. Des Weiteren müssen die Artikel spezifiziert werden. In Bezug auf den Lagerstandort sind der Initialbestand sowie die Bestellpolitik und die Wiederbeschaffungszeit zu ermitteln. Der Transport der Artikel erfolgt innerhalb von LKW-Touren. Somit sind beispielweise die maximale Ladekapazität sowie die Zuordnung der Kunden zu den Touren zu bestimmen.

#### **5.2.1.6 Informationsidentifikation**

Der nächste Schritt besteht in der Informationsidentifikation. Ausgehend vom zuvor ermittelten objektiven Informationsbedarf, ist hier der subjektive Informationsbedarf in Bezug auf die Anforderungen von SimChain bestimmt worden. Dies ist durch einen Abgleich des zuvor bestimmten objektiven Informationsbedarfs mit den Eingangsdatentabellen von SimChain erfolgt. Im Folgenden werden zwei Beispiele aufgezeigt. So wurde festgestellt, dass zur Erzeugung der Kunden, Lieferanten und Lagerstandorte in SimChain die zugehörigen geographischen Daten erforderlich sind. Ebenso wurde ermittelt, dass SimChain für die Wiederbeschaffungszeit eine Dreiecks-Verteilung

vorsieht. In dieser werden die minimale, maximale und durchschnittliche Wiederbeschaffungszeit berücksichtigt.

Anschließend wurden alle vorhandenen Informationsquellen, die möglicherweise die zuvor identifizierten notwendigen Eingangsdaten liefern können, hinsichtlich der Verfügbarkeit, der Zugriffsmöglichkeit, des Erhebungsaufwandes und der Qualität der Daten bewertet. Im nächsten Schritt wurde ermittelt, welche Daten aus welchen Informationsquellen den subjektiven Informationsbedarf abdecken können.

#### **5.2.1.7 Erhebungsplanung**

Im nächsten Prozessschritt der Vorgehensweise steht die Planung der Erhebung. Zu den an dieser Stelle ausgewählten Informationsquellen zählen sowohl operative SAP-Systeme, Datenblätter, Fachexperten, als auch das SAP BW. Insbesondere die operativen SAP-Systeme und das SAP BW sind interessant, da die Vorgehensweise im Beispiel-Unternehmen in einem automatisierten Prozess umgesetzt werden soll. Die operativen SAP-Systeme und das SAP BW ermöglichen eine automatisierte Datenerhebung. Für wenige Angaben, wie beispielsweise über die Größe der verwendeten LKWs, die Anzahl pro Ladungsträger im LKW und die Wiederbeschaffungszeit beim Lieferanten, mussten allerdings andere Informationsquellen, in Form von Fachexperten und Datenblättern, genutzt werden. Für die Erstellung eines automatisierten Prozesses zur glaubwürdigen Messbarkeit des RDLBs ist es erforderlich, diese digital zu erheben, so dass auch für diese Daten eine automatisierte Erhebung möglich ist. Zur Vorbereitung der Erhebung wurden einerseits entsprechende BEx Queries zum Abzug der Daten aus dem SAP-System erstellt und durch entsprechende Voruntersuchungen getestet. Andererseits wurden beispielsweise die Daten für die Wiederbeschaffungszeit durch ein Interview mit dem zuständigen Dispositionsleiter erhoben. Zur Vorbereitung wurde ein Fragenkatalog erstellt.

#### **5.2.1.8 Erhebung und Datenerfassung der notwendigen Eingangsdaten**

Im Anschluss an die Erstellung der BEx Queries und die Vorbereitung der Interviews erfolgte die eigentliche Durchführung der Erhebung. Während die Daten, welche mittels der BEx Queries erhoben wurden bereits digital vorlagen, mussten die resultierenden Daten aus den Interviews sowie aus der Analyse der Datenblätter zusätzlich digital erfasst werden. Anschließend wurden die erhobenen Daten gemeinsam mit einem Fachexperten des Beispiel-Unternehmens und einem weiteren SimChain-Fachexperten validiert. Hierbei wurde überprüft, inwiefern die erhobenen Daten den zuvor ermittelten objektiven Informationsbedarf abdecken.

#### **5.2.1.9 Datenstrukturierung**

Als nächster Prozessschritt folgt innerhalb der entwickelten Vorgehensweise die Datenstrukturierung. Die gesamte Datenstrukturierung sowie die anschließende

statistische Aufbereitung sind innerhalb eines Excel-Tools erfolgt. Dieses bildet einen Software Demonstrator des in Kapitel 4.4 beschriebenen Input Data ETL Moduls ab. In diesem werden die Eingangsdaten in die sogenannten Input-Tabellenblätter geladen. In den als Hilfstabellen bezeichneten Tabellenblättern werden die Eingangsdatendaten mittels Excel-Formeln strukturiert und statistisch aufbereitet. Die aufbereiteten Daten werden wiederum über Excel-Formeln in die sogenannten SimChain-Tabellenblätter geladen. Diese zeichnen sich durch die gleiche Struktur wie auch die Simchain-Tabellen in der MySQL Datenbank aus. Mittels Visual Basic Application werden die Daten dann aus den SimChain-Tabellenblättern des Excel-Tools in die MySQL Datenbank eingepflegt.

Im Folgenden werden einige Beispiele aus der durchgeführten Datenstrukturierung beschrieben. Angaben bezüglich des Datums lagen in den erhobenen Daten in der Form TT.MM.JJJJ vor. In SimChain wird das Datum in der Form JJJJ-MM-TT benötigt. Daher wurden die entsprechenden Daten diesbezüglich syntaktisch formatiert. Die Angaben über die Kundennachfragen wurden mittels zweier verschiedener BEx Queries erhoben. Durch die eine BEx Query sind der nachgefragte Artikel und der jeweilige Kunde und mit Hilfe der anderen BEx Query das jeweilige Bestelldatum und der geplante Liefertermin ermittelt worden. Über eine Kombination aus Verkaufsbelegnummer und der Positionsnummer, die in den Datensätzen beider BExQueries enthalten sind, konnte entsprechende Relationen definiert werden. Des Weiteren beinhalten die Nachfragedaten ausschließlich die bestellte Menge pro kg. Da die Nachfragemenge in SimChain in der Einheit Stück gefordert wird, fand hier eine entsprechende Umrechnung statt. Ebenso wurden fehlende Werte innerhalb der Lieferzeiten festgestellt. Durch die Analyse des jeweiligen Kunden und die Absprache mit einem Fachexperten stellte sich heraus, dass es sich hierbei um Barverkäufe handelt. Demnach existieren für diese auch keine Lieferzeiten. Die zugehörigen Bestellungen wurden im Folgenden nicht weiter betrachtet.

Durch eine gemeinsame Überprüfung der aufbereiteten Eingangsdaten mit einem Fachexperten konnte anschließend die Korrektheit der vorgenommenen Datenstrukturierung bestätigt werden.

#### **5.2.1.10 Statistische Datenanalyse**

Analog zur Datenstrukturierung wurde ebenfalls die statistische Datenanalyse innerhalb der Hilfstabellen des entwickelten Excel-Tools mittels Excel-Formeln durchgeführt. In SimChain ist es beispielsweise notwendig für die Wiederbeschaffungszeit eine Dreiecksverteilung zu bestimmen. Hierzu wurden im Interview mit dem zuständigen Dispositionsleiter nicht nur die mittlere Wiederbeschaffungszeit, sondern auch die minimale und maximale Wiederbeschaffungszeit erfragt. Dies sind die notwendigen Daten zur Abbildung der Dreiecksverteilung. Es wird angestrebt den Einkaufs- und

Verkaufspreis in SimChain zu integrieren. Da beide Preise aufgrund wirtschaftlicher Einflüsse schwanken, müssten hierfür entsprechende Verteilungen in SimChain implementiert werden. Hierzu wäre es notwendig innerhalb der statistischen Datenanalyse entsprechende Verteilungen über die Vergangenheitsdaten der Einkaufs- und Verkaufspreise ermittelt werden. Da die Integration in SimChain innerhalb dieser Arbeit nicht möglich war, werden die Preise wie zuvor beschrieben in die Ergebnisdaten integriert.

#### **5.2.1.11 Datennutzbarkeitsprüfung**

Nachfolgend wurden die erhobenen Daten gemeinsam mit einem Fachexperten des Beispiel-Unternehmens und einem weiteren SimChain-Fachexperten validiert. Hierbei wurde überprüft, inwiefern die erhobenen Daten den zuvor ermittelten subjektiven Informationsbedarf abdecken.

Die Simulation wurde anhand eines deterministischen Systems durchgeführt. Dies bedeutet, dass für die Eingangsdaten ausschließlich konstante Werte verwendet wurden (vgl. 2.2.3). Für die Kundenachfrage wurden die gesammelten, realen Daten direkt während der Simulation eingelesen. Dies wird als trace-driven-Simulation bezeichnet (vgl. 2.2.4). Für die zuvor bereits beschriebene Dreiecksverteilung der Wiederbeschaffungszeit wurde der gleiche Wert für die minimale, maximale und durchschnittliche Wiederbeschaffungszeit definiert. Die Überprüfung der aufbereiteten Eingangsdaten im Kontext mit dem Simulationsmodell ist durch den Vergleich der Ergebnisdaten der Simulation mit den realen Output-Daten des Logistiknetzwerks erfolgt. Diese wurden zuvor erhoben und werden innerhalb der entwickelten Vorgehensweise als reale Zieldaten bezeichnet.

### **5.2.2 Aufbereitung der Ergebnisdaten zur glaubwürdigen Messbarkeit der komplexen DWH- Kennzahl RDLB**

Im Anschluss an die durchgeführte Simulation wurden die verifizierten und validierten simulierten Ergebnisdaten entsprechend der in Abbildung 19 aufgezeigten Prozessschritte der entwickelten Vorgehensweise aufbereitet. Hierzu wurde ein weiteres Excel-Tool entwickelt, das einen Software Demonstrator für das Output Data ETL Modul im automatisierten Prozess (vgl. 4.4) darstellt. In diesem werden erneut drei verschiedene Tabellenblätter-Typen unterschieden. In die Eingangstabellenblätter werden die Ergebnisdaten der Simulation aus der MySQL-Datenbank geladen. Innerhalb der Hilfstabellenblätter erfolgt die Aufbereitung der simulierten Ergebnisdaten mittels Excel-Formeln. Die Ausgangsdatentabellenblätter beinhalten die aufbereiteten Ergebnisdaten in der Form der notwendigen Zieldaten zur Messung der DWH-Kennzahlen im Schatten-DWH-System.

### **5.2.2.1 Datenstrukturierung**

Als erster Prozessschritt der entwickelten Vorgehensweise zu Aufbereitung der Ergebnisdaten steht die Datenstrukturierung. Innerhalb der Erprobung sind nur eine Ergebnistabelle (vgl. Abbildung 26) und ein Eingangscube (vgl. 5.2.1) bestimmt worden. Entsprechend der Vorgabe in der Vorgehensweise werden alle notwendigen Ergebnisdaten für jeweils einen Eingangscube in einem Tabellenblatt gespeichert. Da hier nur ein Eingangscube ausgewählt wurde, wurden alle notwendigen Ergebnisdaten in ein Tabellenblatt des Excel-Tools geladen. Dies sind einerseits der durchschnittliche Bestand und andererseits die abgehende Menge (vgl. 5.2.1). Ebenso wurden die beiden zuvor ermittelten notwendigen Attribute Artikel und Lagerstandort mit selektiert.

### **5.2.2.2 Statistische Datenanalyse**

Wie oben beschrieben, bildet das simulierte Logistiknetzwerk ein deterministisches System ab. Da hierbei die Ergebnisdaten für alle Simulationsdurchläufe gleich sind (vgl. Kapitel 2.2.3), ist in der Erprobung lediglich ein Simulationslauf durchgeführt worden. Dies hatte zur Folge, dass hier keine statistische Aufbereitung der Ergebnisdaten notwendig war.

In der entwickelten Vorgehensweise wird beschrieben, dass die simulierten Ergebnisdaten im Anschluss an die statistische Datenanalyse durch den Vergleich mit den erhobenen den Zieldaten zugehörigen realen Daten validiert werden sollen. Hierzu war es zunächst erforderlich den fakturierten Umsatz und den Wareneinsatz zu erzeugen. Dies ist durch die Integration des durchschnittlichen Verkaufspreises und des durchschnittlichen Wareneinsatz pro Stück in die Ergebnisdaten erfolgt. Hierbei wurde die ausgehende Menge der einzelnen Artikel aus dem Lagerstandort A mit dem entsprechenden durchschnittlichen Verkaufspreis und dem durchschnittlichen Wareneinsatz pro Stück multipliziert. Diese Erzeugung von Umsatz und Wareneinsatz dient lediglich als Ersatz für die Integration der Umsatz- und Wareneinsatzberechnung im SimChain-Code. Daher ist die beschriebene Generierung von Umsatz und Wareneinsatz in der entwickelten Vorgehensweise nicht enthalten. Die Veränderung des SimChain-Codes war innerhalb dieser Erprobung nicht möglich.

Aufgrund der Tatsache, dass das Simulationsmodell nur eine vereinfachte Nachbildung (vgl. 2.1) des realen Logistiknetzwerks des Werkstoffhändlers darstellt, stimmen die simulierten Ergebnisdaten nicht eindeutig mit den realen Daten überein. So wurden beispielsweise die identifizierten Barverkäufe innerhalb des Simulationsmodells nicht berücksichtigt. Weiterhin wurde die Bestellpolitik vereinfacht dargestellt. Dementsprechend wurde in Absprache mit dem Beispiel-Unternehmen festgelegt, dass Ergebnisdaten, welche eine Abweichung von bis zu 5 % aufweisen, als hinreichend genau eingestuft werden. Da sich für die simulierten Ergebnisdaten keine

Abweichungen von mehr als 5 % gegenüber den erhobenen realen Zieldaten ergeben haben, wurden die Ergebnisdaten als valide eingeordnet.

### 5.2.2.3 Ergänzung fehlender Attribute

Entsprechend der Vorgehensweise wurden in einem nächsten Schritt die simulierten Ergebnisdaten um die fehlenden Attribute ergänzt. Wie in Kapitel 5.2.1 beschrieben, wurde zuvor ermittelt, dass die Attribute Sparte und Lagerstandort in den Ergebnisdaten der Simulation zur Berechnung der Kennzahl im Schatten-DWH-System fehlen. Da sich die Erprobung ausschließlich auf den Buchungskreis K35 beschränkt hat, konnten die Ergebnisdaten mit geringem Aufwand um das Attribut Buchungskreis ergänzt werden. Die Sparte ergibt sich durch den jeweiligen Artikel, da jeder Artikel einer bestimmten Sparte zugeordnet ist. Mit Hilfe einer Auflistung dieser Zuordnung wurden die simulierten Ergebnisdaten um die zugehörige Sparte erweitert. Die um die fehlenden Attribute erweiterten Ergebnisdaten sind in Abbildung 28 dargestellt.

Attribute (aus SimChain)		ergänzte Attribute		Kennzahlen (aus SimChain)		berechnete Kennzahlen	
SKU (Artikel)	location (Lagerstandort)	Sparte	Buchungskreis	NumItems_Outbound	Avg_Inventory_Items	fakturierter Umsatz	Wareneinsatz
Artikel der Gruppe A	Standort A	B2	K35	226	40	2.846,34	1.617,87
	Standort A	B2	K35	208	44	3.473,80	1.857,18
	Standort A	B2	K35	543	71	10.768,26	5.945,47
	Standort A	B2	K35	291	53	6.742,55	3.606,39
	Standort A	B2	K35	433	89	11.821,09	5.860,87
	Standort A	B2	K35	369	76	11.932,16	6.360,61
	Standort A	B2	K35	380	94	14.176,77	7.980,44
	Standort A	B2	K35	402	94	19.214,85	9.569,76
Artikel der Gruppe B	Standort A	B2	K35	168	33	2.611,39	1.407,54
	Standort A	B2	K35	261	36	5.104,52	2.768,95
	Standort A	B2	K35	419	49	9.703,77	5.427,19
	Standort A	B2	K35	239	38	6.582,04	3.440,58
	Standort A	B2	K35	148	53	4.578,42	2.392,09
	Standort A	B2	K35	129	31	4.852,56	2.634,35
	Standort A	B2	K35	89	39	3.163,49	2.154,57
	Standort A	B2	K35	35	18	1.988,48	1.042,33

**Abbildung 28: Screenshot - Aufbereitung der Ergebnisdaten in den Hilfstabellen des Excel-Tools**

### 5.2.2.4 Aggregation

Nachdem zuvor ermittelt wurde (vgl. 5.2.1), dass die notwendigen Ergebnisdaten die gleiche Aggregationsstufe wie die notwendigen Zieldaten aufweisen, war keine Aggregation der simulierten Ergebnisdaten erforderlich.

### 5.2.2.5 Formatierung

Laut der entwickelten Vorgehensweise werden die simulierten Ergebnisdaten in einem letzten Schritt in die für den Eingang in das Schatten-DWH-System notwendige Form gebracht. Für diesen Anwendungsfall hätten die Ergebnisdaten in die in Tabelle 1 (vgl. 5.2.1.2) aufgezeigten Datentypen konvertiert werden müssen. Da im Beispiel-Unternehmen keine Lizenz für ein zweites DWH-System vorhanden war, wurde die Berechnungslogik der RDLBs mit Hilfe von entsprechenden Formeln in Excel nachgebaut. Demnach war eine Konvertierung der Datentypen nicht erforderlich.

Die aufbereiteten Ergebnisdaten wurden dem IT-Fachexperten vorgelegt. Dieser bestätigte, dass die aufbereiteten Ergebnisdaten, vor dem Hintergrund der notwendigen Konvertierung der Datentypen, die notwendigen Zieldaten zur Berechnung der Kennzahl im Schatten-SAP BW darstellen. Anschließend wurde die Berechnungslogik der RDLB durch Formeln in Excel nachgebaut. Die Formeln zur Berechnung des RDLBs wurden aus den Ergebnissen der allgemeinen Analyse des RDLBs (vgl. Abbildung 24) entnommen. Die Korrektheit der nachgebauten Berechnungslogik in Excel wurde durch einen Fachexperten für die Kennzahl RDLB bestätigt.

Abschließend wurden die berechneten simulierten Werte für die RDLB mit den entsprechenden real gemessenen Werten der Kennzahl im SAP BW verglichen. Hierbei ergaben sich Abweichungen von bis zu 15 % gegenüber den zugehörigen realen Werten der RDLBs. Diese konnten allerdings auf die bereits abweichenden simulierten Ergebnisdaten von den entsprechenden realen Daten zurückgeführt werden.

In den durchgeführten Gesprächen mit dem IT-Fachexperten wurde ebenfalls die Umsetzung der glaubwürdigen Messung des RDLBs in einem automatisierten System thematisiert. Hierbei stellte sich heraus, dass beispielsweise die Software Komponente Output Data ETL Modul des automatisierten Prozesses (vgl. Abbildung 21) zu einem Teil im SAP BW erfolgen kann. Die darin enthaltene Datenstrukturierung und statistische Datenanalyse müssen weiterhin außerhalb des SAP BWs durchgeführt werden. Fehlende Attributen, die im SAP BW bereits gespeichert sind, wie beispielsweise zugehörige Stammdaten, können innerhalb des SAP BW ergänzt werden. Ebenso kann die Formatierung und Aggregation innerhalb des SAB BW durchgeführt werden.

### **5.3 Fazit**

In diesem Kapitel wurde die im Rahmen dieser Arbeit entwickelte Vorgehensweise zur Aufbereitung von Eingangs- und Ergebnisdaten zur glaubwürdigen Messbarkeit von komplexen DWH-Kennzahlen anhand eines realen Fallbeispiels eines Werkstoffhändlers erprobt.

Innerhalb der Erprobung konnten alle in der Vorgehensweise beschriebenen Prozessschritte ohne auftretende Unstimmigkeiten durchgeführt werden. Ebenso wurden keine weiteren Schritte zur Durchführung der Vorgehensweise benötigt. Dies bedeutet, dass die Vorgehensweise als in sich konsistent und vollständig betrachtet werden kann. Im Anwendungsfall wurde für die betrachtete Kennzahl RDLB kein bestehendes DSO der Kennzahlenberechnung im SAP BW, sondern ein dafür eigens neu erzeugtes DSO als Eingangscube gewählt. Obwohl dies in der Vorgehensweise nicht vorgesehen ist, war es trotzdem möglich die Vorgehensweise durchzuführen. Weiterhin wurde aufgezeigt, dass sich durch die Anwendung der Vorgehensweise

bereits erste Anhaltspunkte für die automatisierte Messung der DWH-Kennzahl ergeben. Beispielsweise wurde identifiziert, welche Daten hierfür im Beispiel-Unternehmen elektronisch erfasst werden müssten.

Ergebnis der Erprobung ist die Validierung der entwickelten Vorgehensweise durch die glaubwürdige Messung der DWH-Kennzahl RDLB. Ein Simulationsmodell und damit einhergehend die daraus resultierenden Ergebnisse gelten als glaubwürdig, wenn sie vom Auftraggeber als korrekt angesehen werden (vgl. 2.1). Dies gilt ebenso für die Glaubwürdigkeit der entwickelten Vorgehensweise sowie der gemessenen simulierten DWH-Kennzahlen. So wurden die gemessenen simulierten Werte der Kennzahl RDLB vom Beispiel-Unternehmen als korrekt angesehen. Auf der einen Seite konnten die Abweichungen der gemessenen Werte von bis zu 15% gegenüber den realen Werten auf die bereits abweichenden Ergebnisdaten der Simulation zurückgeführt werden. Diese wurden ebenfalls zuvor vom Beispiel-Unternehmen als glaubwürdig eingeordnet. Auf der anderen Seite wurde die Glaubwürdigkeit der Vorgehensweise sowie der gemessenen Werte der Kennzahl RDLB durch verschiedene Maßnahmen zur Überprüfung der einzelnen Prozessschritte erhöht. So wurden beispielsweise an mehreren Stellen innerhalb der Anwendung der Vorgehensweise Fachexperten des Beispiel-Unternehmens in die Überprüfung der einzelnen Zwischenergebnisse mit einbezogen.

## 6 Zusammenfassung und Ausblick

Im Rahmen dieser Arbeit wurde eine Vorgehensweise zur Aufbereitung von Eingangs- und Ergebnisdaten zur glaubwürdigen Messbarkeit von komplexen DWH-Kennzahlen entwickelt. In dieser werden auf Grundlage der Ergebnisdaten der Simulation die DWH-Kennzahlen in einem sogenannten Schatten-DWH-System gemessen.

In der bestehenden Literatur wird die Kombination einer ereignisdiskreten Simulation mit der DWH-Technologie bereits beschrieben (vgl. [EGM+11; VEC04]), es besteht zur Zeit aber noch kein Bezug zur glaubwürdigen Messbarkeit von komplexen DWH-Kennzahlen.

Die entwickelte Vorgehensweise unterteilt sich in zwei einzelne Vorgehensweisen für die Aufbereitung der Eingangs- und die der Ergebnisdaten. Die Vorgehensweise für die Eingangsdaten stellt eine Erweiterung des IGP von *Bernhard et al.* dar (vgl. Kapitel 4.1). Die Erweiterung besteht aus den vier Schritten, welche sich auf die notwendigen Ziel- und Ergebnisdaten zur Berechnung der DWH-Kennzahlen beziehen. Für die Aufbereitung von Ergebnisdaten ist in der Literatur keine Vorgehensweise vorhanden. Daher stützt sich dieser Teil der entwickelten Vorgehensweise lediglich auf den Aufbereitungsschritt von Ergebnisdaten aus der VDI Richtlinie 3633. Zur Entwicklung der Vorgehensweise wurden die in Kapitel 2 und 3 enthaltenen Informationen zur ereignisdiskreten Simulation und DWH-Kennzahlen verwendet.

Zur Nutzung der entwickelten Vorgehensweise innerhalb einer Simulationsstudie wurde diese in das Simulationsvorgehensmodell zur kontinuierlichen V&V von *Rabe et al.* integriert (vgl. 4.3). Abschließend wurde die entwickelte Vorgehensweise am Fallbeispiel eines Werkstoffhändlers durch die glaubwürdige Messung einer bestehenden DWH-Kennzahl validiert (vgl. 5).

Mit Hilfe der entwickelten Vorgehensweise ist es möglich komplexe DWH-Kennzahlen glaubwürdig an einer ereignisdiskreten Simulation eines Logistiknetzwerks des Werkstoffhandels zu messen. Hierdurch ergibt sich die Möglichkeit die Auswirkungen von Veränderungen in einem Logistiknetzwerk auf die Entwicklung der DWH-Kennzahlen vorherzusagen. Veränderungen können sich auf der einen Seite durch externe Einflüsse wie beispielsweise die wirtschaftliche Lage ergeben. Auf der anderen Seite resultieren interne Veränderungen in einem Logistiknetzwerk beispielsweise durch die Anwendung von Maßnahmen zur Steigerung der Performanz des Logistiknetzwerks. Durch die Veränderung der Eingangsdaten der Simulation können verschiedene Szenarien im simulierten Logistiknetzwerk abgebildet und die damit einhergehenden veränderten Werte der DWH-Kennzahlen bestimmt werden.

Ebenso kann die entwickelte Vorgehensweise in dem dieser Arbeit zugrunde liegenden Forschungsvorhaben eingesetzt werden (vgl. 1). Einen wichtigen Punkt stellt hier die

Ermittlung der Auswirkungen von Maßnahmen zur Verbesserung des Wertes einer DWH-Kennzahl auf weitere DWH-Kennzahlen dar. Diese Interdependenzen zwischen den einzelnen DWH-Kennzahlen lassen sich in der Theorie nur schwer detailliert abbilden. Die entwickelte Vorgehensweise zeigt einen Lösungsweg zur Darstellung dieser Interdependenzen auf.

Vor dem Hintergrund der wachsenden Bedeutung von logistischen Assistenzsystemen [BCK+09, S.242 ff.] ist es das Ziel, die beschriebene Vorgehensweise in einem automatisierten Prozess umzusetzen (vgl. 4.4). Ebenso wird der Bereich der Datensammlung vom automatisierten Datensammlungsprozess dominiert [HiOn14]. Durch die periodische Aktualisierung der Eingangsdaten mit Hilfe eines automatisierten Datensammlungsprozesses bildet das Simulationsmodell stets die aktuellen Gegebenheiten des Logistiknetzwerks ab. Das automatisierte System läuft somit parallel zur realen Welt und dient als Steuerungsinstrument zur Lenkung der Betriebs- und Logistikprozesse in Unternehmen.

## Literaturverzeichnis

Arno08

Arnold, D.: *Handbuch Logistik*, 3. Auflage. Berlin: Springer Verlag, 2008.

BCN+14

Banks, J.; Carson ||, J. S.; Nelson, B. L.; Nicol, D. M.: *Discrete-event system simulation*, 5. Auflage. Harlow, Essex, England: Pearson, 2014.

BDKR01

Baron, C. P; Dietel, U.; Kreppenhofer, D.; Rabe, M.: *Handlungsanleitung Simulation*. In: Rabe, M.; Hellingrath, B. (Hrsg.): *Handlungsanleitung Simulation in Produktion und Logistik. Ein Leitfaden mit Beispielen für kleinere und mittlere Unternehmen*. San Diego, CA, USA: SCS International, 2001, S. 117–190.

Baue03

Bauer, A.: *Datenallokation und Anfrageoptimierung in verteilten, förderierten Data-Warehouse-Systemen*. Erlangen-Nürnberg: Institut für Informatik, 2003 Dissertation.

BaGü13

Bauer, A.; Günzel, H.: *Data-Warehouse-Systeme. Architektur, Entwicklung, Anwendung*. Heidelberg: Dpunkt.verlag, 2013.

BSJ+09

Bengtsson, N.; Shao, G.; Johansson, B.; Swee Leong, Y. T. L.; Skoogh, A.; Mclean, C.: *Input data management methodology for discrete event simulation*. In: Rossetti, M.; Hill, R.; Johansson, B.; Dunkin, A.; Ingalls, R. (Hrsg.): *Proceedings of the 2009 Winter Simulation Conference*. Austin, TX, USA: IEEE Press, 2009, S. 1335–1344.

Berg13

Bergmann, S.: *Automatische Generierung adaptiver Modelle zur Simulation von Produktionssystemen*. Ilmenau: TU Ilmenau Universitätsbibliothek, 2013  
Dissertation.

BJH+07

Bernhard, J.; Jodin, D.; Hömberg, K.; Kuhnt, S.; Schürmann, C.; Wenzel, S.: *Vorgehensmodell zur Informationsgewinnung. -Prozessschritte und Methodennutzung-*. Dortmund: Sonderforschungsbereich 559 - Modellierung großer Netze in der Logistik, 2007 Technical Report 06008 ISSN 1612-1376.

BeWe05

Bernhard, J.; Wenzel, S.: *Information acquisition for model based analysis of large logistics networks*. In: Merkurjev, Y.; Zobel, R.; Kerckhoffs, E. (Hrsg.): *Proceedings 19th European Conference on Modelling and Simulation*. Riga, Litauen, 2005, S. 37–43.

BCK+09

Blutner, D.; Cramer, S.; Krause, S.; Mönks, T.; Nagel, L.; Reinholz, A.; Witthaut,

M.: *Assistenzsysteme für die Entscheidungsunterstützung*. In: Buchholz, P.; Clausen, U. (Hrsg.): *Grosse Netze der Logistik. Die Ergebnisse des Sonderforschungsbereichs 559*. Berlin: Springer Verlag, 2009.

#### Bund15

Bundesministerium des Inneren: *Handbuch für Organisationsuntersuchungen und Personalbedarfsermittlung*, 2015,  
[http://www.orghandbuch.de/OHB/DE/Organisationshandbuch/6\\_MethodenTechniken/63\\_Analysetechniken/633\\_FehlermoeglichkeitUndEinflussanalyse/fehlermoeglichkeitundeinflussanalyse\\_inhalt.html](http://www.orghandbuch.de/OHB/DE/Organisationshandbuch/6_MethodenTechniken/63_Analysetechniken/633_FehlermoeglichkeitUndEinflussanalyse/fehlermoeglichkeitundeinflussanalyse_inhalt.html).  
Abgerufen am 25.03.2015.

#### CBW08

Csanady, K.; Bockel, B.; Wenzel, S.: *Methodik zur systematischen Informationsgewinnung für Simulationsstudien*. In: Rabe, M. (Hrsg.): *Advances in Simulation for Production and Logistics Applications*. Stuttgart: Fraunhofer-IRB-Verlag, 2008.

#### CuCh13

Currie, C. S.M.; Cheng, R. C.H.: *A practical introduction to analysis of simulation output data*. In: Pasupathy, R.; Kim, S.-H.; Tolk, A.; Hill, R.; Kuhl, M.; eds. (Hrsg.): *Proceedings of the 2013 Winter Simulation Conference*. Washington, D.C., USA, 2013, S. 328–341.

#### DaCh10

Daniluk, D.; Chisu, R.: *Simulation und Emulation im Internet der Dinge*. In: Günthner, W.; Hompel, M. (Hrsg.): *Internet der Dinge in der Intralogistik*. Berlin: Springer Verlag, 2010, S. 149–162.

#### DrRa14

Dross, F.; Rabe, M.: *A SimHeuristic Framework as a Decision Support System for large Logistics Networks with complex KPIs*. In: : *Proceedings of the 22nd Symposium Simulationstechnik (ASIM 2014)*. Deutschland: HTW Berlin, Deutschland, ARGESIM / ASIM, Wien, Österreich, 2014

#### EGM+11

Ehmke, J. Fabian; Großhans, D.; Mattfeld, D. Christian; Smith, L. Douglas: *Interactive analysis of discrete-event logistics systems with support of a data warehouse*. In: *Computers in Industry*, 62, 2011, S. 578–586.

#### Eley 2012

Eley, M.: *Simulation in der Logistik. Einführung in die Erstellung ereignisdiskreter Modelle unter Verwendung des Werkzeuges "Plant Simulation"*. Berlin: Springer Verlag, 2012.

## Fahr10

Fahrmeir, L.: *Statistik. Der Weg zur Datenanalyse*, 7. Auflage. Berlin: Springer Verlag, 2010.

## Fark11

Farkisch, K.: *Data-Warehouse-Systeme kompakt. Aufbau, Architektur, Grundfunktionen*. Berlin: Springer Verlag, 2011.

## Fech15

Fechteler, Till: *The bridge to reality. SimChain Supply Chain Simulation as a Service*. SimPlan AG, [http://www.plant-simulation.de/images/stories/Dokumente/SimChain\\_PlantUserMeeting\\_20131025\\_Siemens.pdf](http://www.plant-simulation.de/images/stories/Dokumente/SimChain_PlantUserMeeting_20131025_Siemens.pdf).

Abgerufen am 12.02.2015.

## FeGu14

Fechteler, Till; Gutenschwager, Kai: *SimChain. Technical Documentation*, 2014.

## Gada12

Gadatsch, A.: *Grundkurs Geschäftsprozess-Management. Methoden und Werkzeuge für die IT-Praxis: Eine Einführung für Studenten und Praktiker*, 7. Auflage.

Wiesbaden: Vieweg+Teubner Verlag, 2012.

## Goek06

Goeken, M.: *Entwicklung von Data-Warehouse-Systemen.*

*Anforderungsmanagement, Modellierung, Implementierung*. Wiesbaden: Dt. Univ.-Verlag, 2006.

## GuAl04

Gutenschwager, K.; Aliche, K.: *Supply Chain Simulation mit ICON-SimChain*. In: Spengler, T.; Voß, S.; Kopfer, H. (Hrsg.): *Logistik Management*. Heidelberg: Physica-Verlag HD, 2004, S. 161–178.

## Gute13

Gutenschwager, Kai: *ICON-SimChain. Handbuch und Tutorial*, 2013.

## HiOn12

Hill, J.; Onggo, S.: *Data Identification and collection methodology in a simulation project: an action research*. In: Tjahjono, B.; Heavey, C.; Onggo, S.; van der Zee, D.-J. (Hrsg.): *Proceedings of the Operational Research Society Simulation Workshop*. Worcestershire, England, 2012.

## HiOn14

Hill, J.; Onggo, S.: *Data identification and data collection methods in simulation: a case study at ORH Ltd*. In: *Journal of Simulation*, 8, 2014, S. 195–205.

## HoNo09

Hofmann, E.; Nothardt, F.: *Logistics Due Diligence. Analyse, Bewertung, Anlässe, Checklisten*. Berlin: Springer Verlag, 2009.

## Holt99

Holthuis, J.: *Der Aufbau von Data-Warehouse-Systemen. Konzeption - Datenmodellierung - Vorgehen*, 2. Auflage. Wiesbaden: Dt. Univ.-Verlag, 1999.

## Inmo05

Inmon, W. H.: *Building the data warehouse*, 4. Auflage. Indianapolis, IN, USA: John Wiley & Sons, 2005.

## JHL04

Jodin, D.; Hömberg, K.; Leppin, M.: *Methoden der Informations- und Datenerhebung*. Dortmund: Sonderforschungsbereich 559 - Modellierung großer Netze in der Logistik, 2004 Technical Report 04002 ISSN 1612-1376.

## JKW09

Jodin, D.; Kuhnt, S.; Wenzel, S.: *Methodennutzungsmodell zur Informationsgewinnung in großen Netzen der Logistik*. In: Buchholz, P.; Clausen, U. (Hrsg.): *Grosse Netze der Logistik. Die Ergebnisse des Sonderforschungsbereichs 559*. Berlin: Springer Verlag, 2009, S. 1–15.

## JoMa04

Jodin, D.; Mayer, A.: *Automatisierte Methoden und Systeme der Datenerhebung*. Dortmund: Sonderforschungsbereich 559 - Modellierung großer Netze in der Logistik, 2004 Technical Report 05004 ISSN 1612-1376.

## JuWi00

Jung, R.; Winter, R.: *Data Warehousing: Nutzungsaspekte, Referenzarchitektur und Vorgehensmodell*. In: Jung, R.; Winter, R. (Hrsg.): *Data-Warehousing-Strategie. Erfahrungen, Methoden, Visionen*. Berlin: Springer Verlag, 2000, S. 4–20.

## KGK+13

Kamphues, J.; Groß, S.; Korth, B.; Zajac, M.; Hegmanns, T.: *Serviceorientierte Referenzarchitektur für Logistische Assistenzsysteme zur simulationsbasierten Entscheidungsunterstützung*. In: Dangelmaier, W.; Laroque, C.; Klaas, A. (Hrsg.): *Entscheidungsunterstützung von der Planung bis zur Steuerung. Simulation in Produktion und Logistik*, S. 145–155.

## Kimb96

Kimball, R.: *The Data Warehouse Toolkit. Practical Techniques for Building Dimensional Data Warehouses*. New York, USA: John Wiley & Sons, 1996.

## Koth11

Kothes, L.: *Grundlagen der Technischen Dokumentation. Anleitungen verständlich und normgerecht erstellen*. Berlin: Springer-Verlag, 2011.

## Küpp95

Küpper, H.-U.: *Controlling. Konzeption, Aufgaben und Instrumente*. Stuttgart: Schäffer-Poeschel, 1995.

## Kütz11

Kütz, M.: *Kennzahlen in der IT. Werkzeuge für Controlling und Management*, 4. Auflage. Heidelberg: dpunkt-Verlag, 2011.

## LaKe00

Law, A. M.; Kelton, W. David: *Simulation modeling and analysis*. Boston, MA, USA: McGraw-Hill, 2000.

## Lehn03

Lehner, W.: *Datenbanktechnologie für Data-Warehouse-Systeme. Konzepte und Methoden*. Heidelberg: dpunkt-Verlag, 2003.

## LeSe97

Lehtonen, J.-M.; Seppälä, U.: *A methodology for data gathering and analysis in a logistics simulation project*. In: *Integrated Manufacturing*, 8, 1997, S. 351–358.

## MBK+12

Mertens, P.; Bodendorf, F.; König, W.; Picot, A.; Schumann, M.; Hess, T.: *Grundzüge der Wirtschaftsinformatik*, 11. Auflage. Berlin: Springer Verlag, 2012.

## MuBe00

Mucksch, H.; Behme, W.: *Das Data Warehouse-Konzept als Basis einer unternehmensweiten Informationslogistik*. In: Mucksch, H.; Behme, W. (Hrsg.): *Das Data-Warehouse-Konzept*(4. Aufl. Wiesbaden: Gabler, 2000, S. 5–71.

## Pidd04

Pidd, M.: *Computer simulation in management science*, 5. Auflage. Hoboken, NJ, USA: Wiley, 2004.

## Prei08

Preißler, P. R.: *Betriebswirtschaftliche Kennzahlen. Formeln, Aussagekraft, Sollwerte, Ermittlungsintervalle*. München: Oldenbourg Verlag, 2008.

## RSW08b

Rabe, M.; Spieckermann, S.; Wenzel, S.: *A new procedure model for verification and validation in production and logistics simulation*. In: Mason, S. J.; Hill, R. R.; Mönch, L.; Rose, O.; Jefferson, T.; Fowler, J. W. (Hrsg.): *Proceedings of the 40th Winter Simulation Conference (WSC)*. Miami, FL, USA: IEEE Press, 2008, S. 1717–1726.

## RSW08a

Rabe, M.; Spieckermann, S.; Wenzel, S.: *Verifikation und Validierung für die Simulation in Produktion und Logistik. Vorgehensmodelle und Techniken*. Berlin: Springer Verlag, 2008.

## RSW09

Rabe, M.; Spieckermann, S.; Wenzel, S.: *Verification and validation activities within a new procedure model for V&V in production and logistics simulation*. In: Rossetti, M. D.; Hill, R. R.; Johansson, B.; Dunkin, A.; Ingalls, R. G. (Hrsg.): *Proceedings of*

*the 2009 Winter Simulation Conference (WSC)*. Piscataway, NJ: IEEE, 2009, S. 2509–2519.

## Reic11

Reichmann, T.: *Controlling mit Kennzahlen. Die systemgestützte Controlling-Konzeption mit Analyse- und Reportinginstrumenten*, 8. Auflage. München: Vahlen, 2011.

## RiWi98

Ritzschke, M.; Wiedemann, T.: *Gewinnung und Aufbereitung von Simulationsdaten zu Vergleichszwecken*. In: Engeli, M.; Hrdliczka, V. (Hrsg.): *Simulationstechnik. 12. Symposium in Zürich, September 1998 : Tagungsband*. Zürich: Hochschulverlag an der ETH, 1998, S. 283–290.

## RoPe02

Robertson, N.; Perera, T.: *Automated data collection for simulation? In: Simulation Practice and Theory*, 9, 2002, S. 349–364.

## Robinson 2003

Robinson, S.: *Simulation modelling. The practice of model development & use*. Hoboken, NJ, USA: Wiley, 2003.

## RoMäl1

Rose, O.; März, L.: *Simulation*. In: März, L. (Hrsg.): *Simulation und Optimierung in Produktion und Logistik. Praxisorientierter Leitfaden mit Fallbeispielen*. Berlin: Springer, 2011, S. 13–20.

## SSH10

Saake, G.; Sattler, K.-U.; Heuer, A.: *Datenbanken - Konzepte und Sprachen*, 4. Auflage. Heidelberg: Verlagsgruppe Hüthig Jehle Rehm, 2010.

## SAP15

SAP: *Die Architektur eines Data Warehouses*, 2015,  
[http://help.sap.com/saphelp\\_nw73/helpdata/de/4a/547f4a6a8a1cd4e10000000a421937/content.htm?frameset=/de/4a/124597ca771b41e10000000a42189c/frameset.htm&current\\_toc=/de/ad/6b023b6069d22ee10000000a11402f/plain.htm&node\\_id=5](http://help.sap.com/saphelp_nw73/helpdata/de/4a/547f4a6a8a1cd4e10000000a421937/content.htm?frameset=/de/4a/124597ca771b41e10000000a42189c/frameset.htm&current_toc=/de/ad/6b023b6069d22ee10000000a11402f/plain.htm&node_id=5).  
Abgerufen am 17.03.2015.

## Schu12

Schuster, T.: *Modellierung, Integration und Analyse von Ressourcen in Geschäftsprozessen*. Karlsruhe: KIT Scientific Publishing, 2012.

## SimP15b

SimPlan AG: *SimChain - Supply Chain Bausteinkasten. Modellierung und Analyse von Liefernetzwerken*, <http://www.plant-simulation.de/bausteine/simchain.html>.  
Abgerufen am 26.01.2015.

## SimP15a

SimPlan AG: *The innovation for more transparency and optimisation*,

<http://www.simchain.net/de/>.

Abgerufen am 26.01.2015.

SBU99

Sinz, E. J.; Böhnlein, M.; Ulbrich vom Ende, A.: *Konzeption eines Data Warehouse-Systems für Hochschulen*. In: Bamberger Beiträge zur Wirtschaftsinformatik, 1999.

SkJo08

Skoogh, A.; Johansson, B.: *A methodology for input data management in discrete event simulation projects*. In: S. J. Mason; R. R. Hill; L. Mönch; O. Rose, T. Jefferson, J.W. Fowler (Hrsg.): *Proceedings of the 2008 Winter Simulation Conference*.

SLR04

Spickermann, S.; Lehmann, A.; Rabe, M.: *V&V: Überlegungen zu einer integrierten Vorgehensweise*. In: Mertins, K.; Rabe, M. (Hrsg.): *Experiences from the future. New methods and applications in simulation for production and logistics*, Stuttgart: Fraunhofer IRB-Verlag, 2004, S. 263–274.

Spri15b

Springer Gabler Verlag: *Gabler Wirtschaftslexikon*. Stichwort: Kennzahlen, <http://wirtschaftslexikon.gabler.de/Archiv/54801/kennzahlen-v8.html>.  
Abgerufen am 28.01.2015.

Spri15a

Springer Gabler Verlag: *Gabler Wirtschaftslexikon*. Stichwort: Parameter, <http://wirtschaftslexikon.gabler.de/Archiv/54940/parameter-v10.html>.

StHa05

Stahlknecht, P.; Hasenkamp, U.: *Einführung in die Wirtschaftsinformatik*. Elfte, vollständig überarbeitete Auflage. Berlin, New York: Springer, 2005.

VEC04

Vasilakis, C.; El-Darzi, E.; Chountas, P.: *A Data Warehouse Environment for Storing and Analyzing Simulation Output Data*. In: Ingalls, R. G; Rossetti, M. D; Smith, J. S; Peters, B. A (Hrsg.): *Proceedings of the 36th Winter Simulation Conference*. USA: IEEE, 2004, S. 703–710.

Vere97

Verein Deutscher Ingenieure e.V.: ,Richtlinie 3633 Blatt 3 *Simulation von Logistik-, Materialfluß und Produktionssystemen - Experimentplanung und -auswertung*. Beuth, Berlin, 1997.

Vere14

Verein Deutscher Ingenieure e.V.: ,Richtlinie 3633 Blatt 1 *Simulation von Logistik-, Materialfluss- und Produktionssystemen - Grundlagen*. Beuth, Berlin, 2014.

Vinc98

Vincent, S.: *Input Data Analysis*. In: Banks, J. (Hrsg.): *Handbook of simulation*.

*Principles, methodology, advances, applications, and practice.* New York, [Norcross, Ga.]: Wiley; Co-published by Engineering & Management Press, 1998, S. 54–91.

VöSc10

Völker, S.; Schmidt, P. Michael: *Simulationsbasierte Optimierung von Produktions- und Logistiksystemen mit Tecnomatrix Plant Simulation.* In: Zülch, G.; Stock, P. (Hrsg.): *Integrationsaspekte der Simulation: Technik, Organisation und Personal.* Karlsruhe: KIT Scientific Publ, 2010, S. 93–100.

Wenz08

Wenzel, S.: *Qualitätskriterien für die Simulation in Produktion und Logistik. Planung und Durchführung von Simulationsstudien.* Berlin: Springer Verlag, 2008.

WeBe03

Wenzel, S.; Bernhard, J.: *Vorgehensmodell zur Informationsgewinnung für die Modellierung von Logistiksystemen.* In: Hohmann, R. (Hrsg.): *Simulationstechnik. Tagungsband zum 17. Symposium in Magdeburg, 2003.*

Wien02

Wiendahl, H.-P.: *Erfolgsfaktor Logistikqualität. Vorgehen, Methoden und Werkzeuge zur Verbesserung der Logistikleistung, 2. Auflage.* Berlin: Springer Verlag, 2002.

Witt99

Witte, T.: *Datenbankgetriebene objektorientierte Simulation. Grundkonzepte und betriebliche Anwendungen.* In: Biethahn, J.; Hummeltenberg, W.; Schmidt, B.; Stähly, P.; Witte, T. (Hrsg.): *Simulation als betriebliche Entscheidungshilfe. State of the Art und neuere Entwicklungen.* Heidelberg: Physica-Verlag HD, 1999, S. 22–49.

WoYa10

Wolf, F. K.; Yamada, S.: *Datenmodellierung in SAP NetWeaver BW, 1. Auflage.* Bonn: Galileo Press, 2010.

Zeh03

Zeh, T.: *Data Warehousing als Organisationskonzept des Datenmanagements.* In: *Informatik - Forschung und Entwicklung*, 18, 2003, S. 32–38.

---

## Eidesstattliche Versicherung

\_\_\_\_\_  
Name, Vorname

\_\_\_\_\_  
Matr.-Nr.

Ich versichere hiermit an Eides statt, dass ich die vorliegende Bachelorarbeit/Masterarbeit\* mit dem Titel

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

\_\_\_\_\_  
Ort, Datum

\_\_\_\_\_  
Unterschrift

\*Nichtzutreffendes bitte streichen

### **Belehrung:**

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden. (§ 63 Abs. 5 Hochschulgesetz - HG - )

Die Abgabe einer falschen Versicherung an Eides statt wird mit Freiheitsstrafe bis zu 3 Jahren oder mit Geldstrafe bestraft.

Die Technische Universität Dortmund wird gfls. elektronische Vergleichswerkzeuge (wie z.B. die Software „turnitin“) zur Überprüfung von Ordnungswidrigkeiten in Prüfungsverfahren nutzen.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

\_\_\_\_\_  
Ort, Datum

\_\_\_\_\_  
Unterschrift