

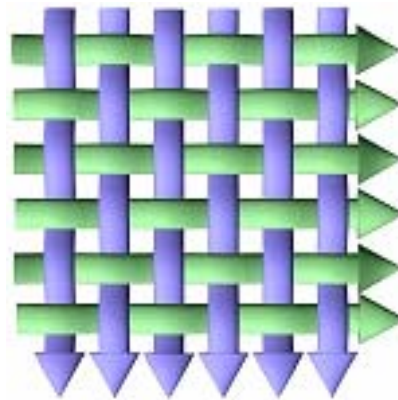


Technical Report 06008

ISSN 1612-1376

Vorgehensmodell zur Informationsgewinnung – Prozessschritte und Methodennutzung –

Jochen Bernhard, Dirk Jodin, Kay Hömberg, Sonja Kuhnt,
Christoph Schürmann, Sigrid Wenzel



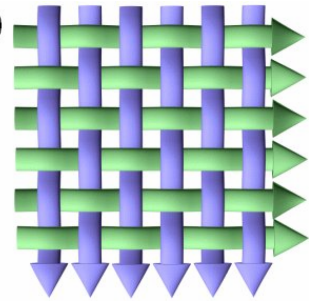
Sonderforschungsbereich 559
Modellierung großer Netze in der Logistik

Universität Dortmund
44221 Dortmund



Sonderforschungsbereich 559

**Modellierung großer
Netze in der Logistik**



Technical Report 06008

ISSN 1612-1376

**Vorgehensmodell zur Informationsgewinnung
- Prozessschritte und Methodennutzung -**

Teilprojekt M9:

Jochen Bernhard

Dirk Jodin

Kay Hömberg

Sonja Kuhnt

Christoph Schürmann

Sigrid Wenzel

Dortmund, den 25. Juni 2007

Inhalt

1	Einleitung	3
2	Integration der Informationsgewinnung in ein übergeordnetes Simulationsvorgehensmodell.....	3
3	Prozesse der Informationsgewinnung.....	7
3.1	Zieldefinition.....	7
3.1.1	Analyse der Aufgabenstellung	7
3.1.2	Ableitung der Ziele für die Informationsgewinnung	8
3.2	Informationsidentifikation	8
3.2.1	Informationsbedarfsanalyse	8
3.2.2	Informationsangebotsanalyse	9
3.2.3	Nutzbarkeitsprüfung der Informationsquellen	9
3.3	Erhebungsplanung.....	9
3.3.1	Auswahl adäquater Informationsquellen	9
3.3.2	Auswahl der Erhebungsmethode	9
3.3.3	Vorbereitung der Erhebung.....	10
3.4	Erhebung und Datenerfassung.....	10
3.4.1	Aufnahme.....	10
3.4.2	Transformation	10
3.4.3	Datenerfassung.....	10
3.4.4	Erhebungsvalidierung	11
3.5	Datenstrukturierung	11
3.5.1	Syntaktische Formatierung	11
3.5.2	Definition von Relationen	11
3.5.3	Fehlerbereinigung	12
3.5.4	Anpassung der Granularität	13
3.5.5	Plausibilitätsprüfung.....	13
3.6	Statistische Datenanalyse	14
3.6.1	Deskriptive Analyse.....	14
3.6.2	Verfahrensauswahl	14
3.6.3	Verfahrensanwendung.....	15
3.6.4	Ergebnisüberprüfung.....	15
3.7	Datennutzbarkeitsprüfung.....	15
3.7.1	Abschließende Datenvalidierung	15
3.7.2	Nutzbarkeitsbewertung	16
4	Konzept der integrierten Methodennutzung	16
5	Ausblick.....	18
6	Literatur	18

1 Einleitung

Bei der modellgestützten Analyse von Logistiksystemen umfasst der Aufwand für die Informations- und Datenbeschaffungsaufgaben eine nicht zu unterschätzende Größenordnung. Untersuchungen wie beispielsweise [RHe01] weisen für Simulationsstudien in Produktion und Logistik darauf hin, dass innerhalb einer Simulationsstudie über 30 % des Aufwandes allein auf die Datenbeschaffung entfallen. Andere Autoren (vgl. [KGr95]) geben sogar eine Größenordnung von bis zu 50% des gesamten Projektaufwandes als Aufwand für die Datenbeschaffung und -aufbereitung an. Hinzu kommt, dass die Simulationsexperten heute kaum methodische Unterstützung zur Sicherstellung der Qualität der zu verwendenden Daten sowie zur Bestimmung ihrer richtigen Quantität und Granularität erhalten. Entsprechende Checklisten, Regelwerke und Vorgehensmodelle zur gezielten Informationsbeschaffung stehen meist gar nicht oder nur für spezifische Anwendungen zur Verfügung. Diese Situation wird bei der Modellierung von Logistiknetzen noch dadurch erschwert, dass sich die Datenbeschaffung nicht nur auf ein Unternehmen, sondern auf mehrere Unternehmen mit zwangsläufig heterogenen Datenbeständen beziehen muss.

Ein erstes Vorgehensmodell zur Informationsgewinnung für die Modellierung und Simulation (M&S) von Logistiknetzen haben Wenzel und Bernhard bereits in [WBe03] vorgestellt. In diesem Beitrag wurde auch thematisiert, ob die in der Literatur i. d. R. benannte Datenbeschaffung die eigentliche Aufgabe unter methodischen Gesichtspunkten inhaltlich hinreichend umfassend charakterisiert. Die Autoren verweisen in diesem Zusammenhang auf die heute übliche Differenzierung zwischen Wissen, Information und Daten (vgl. [NTa97]) und fordern in Analogie zur Modellierung von Entscheidungsprozessen innerhalb des unternehmensinternen Informationsmanagements eine informationstheoretische Betrachtungsweise auch für M&S von logistischen Systemen. Detaillierte Ausführungen zur Differenzierung der Begriffe Wissen, Information und Daten sind dem Technical Report 05001 [BDW05] zu entnehmen.

Informationsgewinnung beinhaltet im Gegensatz zum Datenmanagement eine veränderte Sichtweise. Im Fokus steht die notwendige, richtige und gültige Information bezüglich einer Aufgabenstellung; das Datum selbst, d. h. die Codierung der Information spielt eine untergeordnete Rolle. Die Informationsgewinnung erweitert damit das Aufgabenspektrum des Datenmanagements um Aufgaben der Informationsbeschaffung und -bewertung. Nicht mehr die Frage: „Welche Daten werden benötigt?“ steht im Vordergrund, sondern es muss zunächst primär diskutiert werden, wie sich der Informationsbedarf für die Aufgabenstellung unter Berücksichtigung der gewählten Modellierungsmethode darstellt. Diese Herangehensweise erlaubt eine Separation zwischen dem Informationsbedarf und der jeweiligen Codierung der Information in Daten und lässt damit letztendlich eine verbesserte Datenbereitstellung zu.

2 Integration der Informationsgewinnung in ein übergeordnetes Simulationsvorgehensmodell

Um die Informationsgewinnung in die modellgestützte Analyse einbinden zu können, ist zunächst das jeweilige übergeordnete Vorgehensmodell zu bestimmen, in das die Informationsgewinnung einzuordnen ist. Die übergeordneten Vorgehensmodelle können sich je nach verwendeter Analyseverfahren unterscheiden und müssen hinsichtlich der Einordnung eines Informationsgewinnungsprozesses geprüft und ggf. ergänzt werden.

Im Folgenden wird die Informationsgewinnung im Kontext der ereignisdiskreten Ablaufsimulation (discrete event simulation - DES) näher betrachtet. Eine Übersicht zu Vorgehensmodellen im Kontext der Simulation und Software-Entwicklung ist in [RSW07] gegeben.

Für die Informationsgewinnung in der ereignisdiskreten Simulation lehnen sich Wenzel und Bernhard bereits in ihrem Beitrag [WBe03] an das Simulationsvorgehensmodell der

VDI 3633, Blatt 1 (vgl. [VDI00]; in neuerer Fassung in [VDI07]) an und ordnen den Prozessschritt der Informationsgewinnung in eine Folge von iterativ durchzuführenden Prozessschritten ein. Die Informationsgewinnung wird damit als ein Teilschritt des übergeordneten Vorgehens verstanden, der die bisherige Datenbeschaffung (Daten erheben, aufbereiten, abstimmen) ersetzt und auf Basis der formulierten Aufgabe und damit auch des Untersuchungsgegenstands sowie des festgelegten Untersuchungsziels (Aufgabe und Ziel formulieren, Experimentplan aufstellen, Aufwand abschätzen) durchgeführt wird.

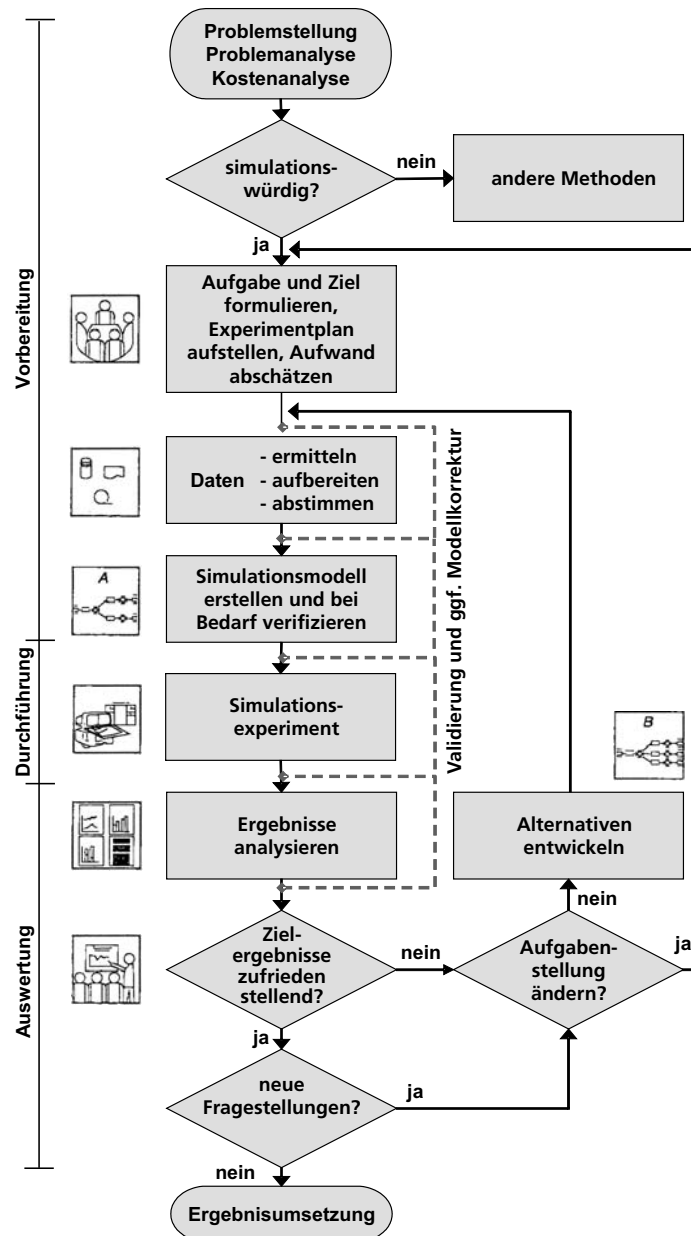


Abbildung 1: Simulationsvorgehensmodell ([VDI07], Bild 7)

Entscheidend ist, dass bereits eine detaillierte Aufgabenbeschreibung, die die Systemgrenzen und die geplante Detaillierung festlegt, vorhanden ist, um eine zielgerichtete und auf das umzusetzende Modell eingeschränkte Informationsgewinnung durchführen zu können. Die Informationsgewinnung wiederum liefert die konkreten Eingangsdaten für das ausführbare, experimentierbare Modell. Innerhalb des Vorgehens ist sie iterativ mit dem Konzept-

modell (auch: konzeptuelles Modell), dem ausführbaren, experimentierbaren Modell, der Validierung und der Ergebnisanalyse verknüpft.

Eine Analyse durchgeführter Simulationsstudien zeigt allerdings, dass die Schritte der Informations- und Datenbeschaffung nicht in einer Prozessfolge der Modellbildung, sondern eher differenziert von der Modellbildung betrachtet werden müssen. Diese Differenzierung schlägt sich auch in einem von der Arbeitsgruppe „Validierung“ der ASIM-Fachgruppe „Simulation in Produktion und Logistik“ entworfenem Simulationsvorgehensmodell wieder [RSW07], das sich in seinen Grundzügen an das Simulationsvorgehensmodell der VDI 3633 anlehnt.

Das Simulationsvorgehensmodell ist insbesondere durch die konsequente Einführung von Phasenergebnissen als Basis für die durchgängige und iterative Verifikation und Validierung (V&V) (vgl. hierzu im Detail [RSW07]) sowie die gesonderte Behandlung von Modell und Daten gekennzeichnet. Das Vorgehen bei der Modellbildung ist in die Phasen Aufgaben- definition, Systemanalyse, Modellformalisierung, Implementierung sowie Experimente und Analyse gegliedert, die Phasen der Datenbeschaffung und -aufbereitung mit den Phasenergebnissen Rohdaten und aufbereitete Daten sind aus der Reihenfolge der Modellierungsschritte herausgenommen, da sie inhaltlich, zeitlich sowie bezüglich der einzubindenden Personen z. T. unabhängig von der Modellierung erfolgen können. Die grafische Einordnung des Phasenergebnisses Rohdaten im Vorgehensmodell ist chronologisch nicht zwingend. Jedoch müssen die aufbereiteten Daten spätestens mit der Fertigstellung des ausführbaren Modells vollständig vorliegen. Für eine detaillierte Beschreibung des Simulationsvorgehensmodells sei auf [RSW07] verwiesen.

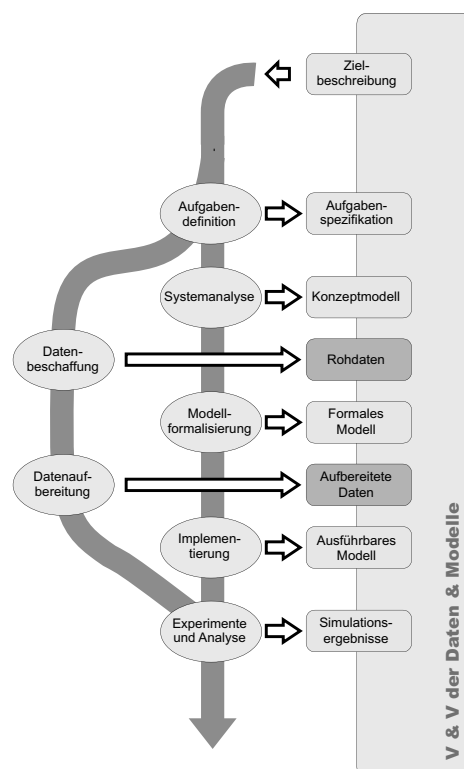


Abbildung 2: Simulationsvorgehensmodell [RSW07]

Eine differenzierte Betrachtung der Datenbeschaffung und -aufbereitung nach dem in Kapitel 1 motivierten Verständnis einer Informationsgewinnung ist jedoch auch in diesem Simulationsvorgehensmodell noch nicht erfolgt. Es erweitert allerdings das Vorgehensmodell der VDI 3633 um eine Trennung in Datenbereitstellung sowie Modellbildung und stellt somit eine weitaus bessere Basis für die Integration eines Informationsgewinnungsprozesses dar.

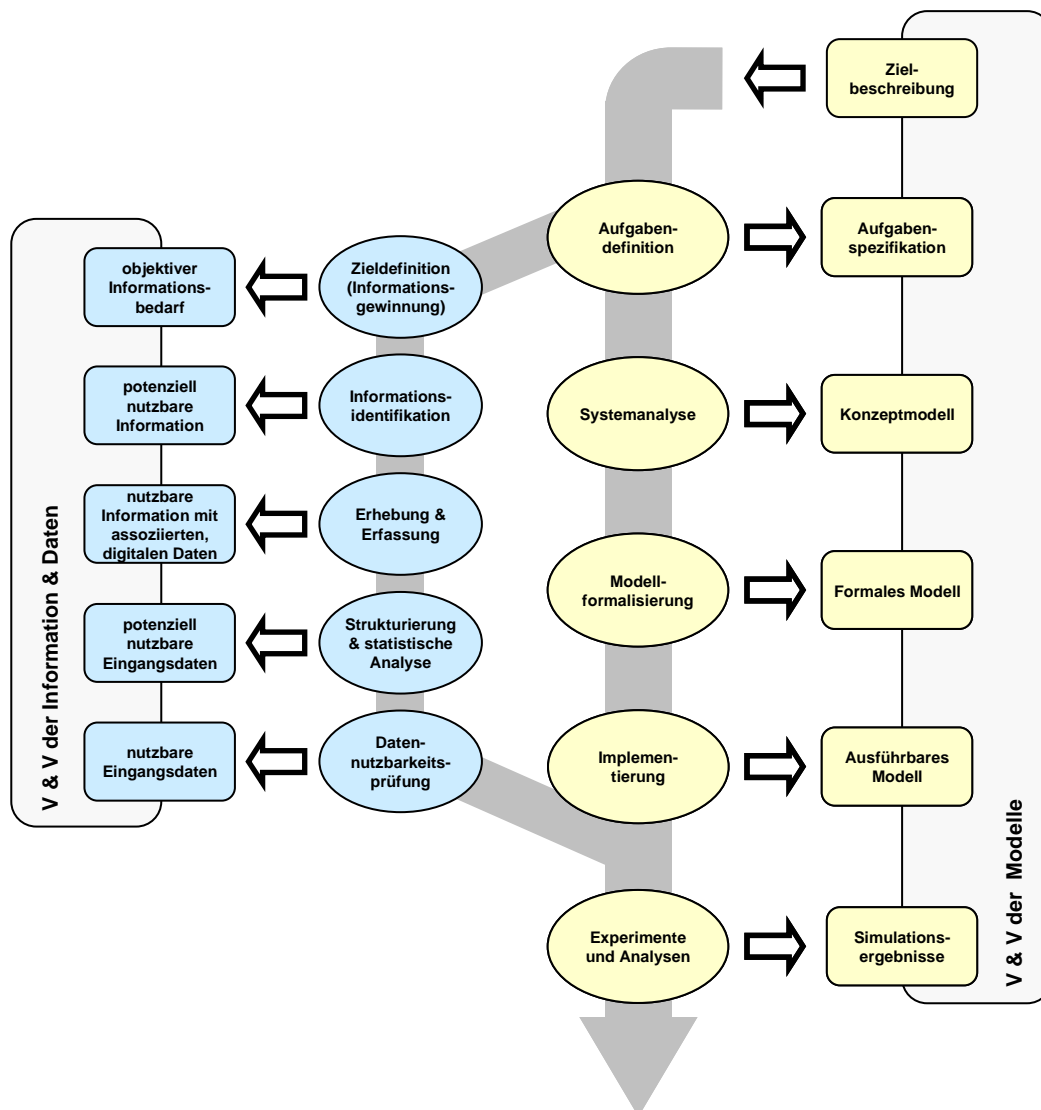


Abbildung 3: Integration der Informationsgewinnung in das Simulationsvorgehensmodell

Mit der Prämisse, dass die für die Modellbildung notwendige Information im Vordergrund steht, müssen die beiden Phasen der Datenbeschaffung und -aufbereitung aus Abbildung 2 ausgehend von dem Informationsbedarf für die Modellbildung weiter detailliert werden (vgl. Abbildung 3). In diesem Zusammenhang ist das Ziel der Informationsgewinnung über die Beschreibung eines objektiv notwendigen Informationsbedarfs zu konkretisieren (Zieldefinition) und die innerhalb des / der Unternehmen potenziell nutzbare Information zu identifizieren (Informationsidentifikation). Die eigentlichen Datenbeschaffung aus dem Simulationsvorgehensmodell detailliert sich in eine Informations- und Datenerhebung sowie eine Datenerfassung. Der Prozess der Datenaufbereitung umfasst eine Datenstrukturierung sowie Aufgaben der statistischen Analyse. Eine abschließende Datennutzbarkeitsprüfung erlaubt einen korrekten Abschluss der zeitlich parallel zur Modellbildung liegenden Informationsgewinnung, stellt die ordnungsgemäße Bereitstellung und Verwendung der Daten im Kontext der Modellbildung sicher und führt damit die beiden parallelen Vorgehen wieder zusammen.

Abbildung 3 stellt das erweiterte Simulationsvorgehensmodell vor, in dem die oben aufgeführte Detaillierung der Informationsgewinnung integriert ist. In Analogie zum Simulationsvorgehensmodell aus Abbildung 2 werden auch hier die Ergebnisse jeder einzelnen Phase einer Verifikation und Validierung (V&V) unterzogen. Allerdings wird in Ergänzung zur durchgängigen V&V die Datennutzbarkeitsprüfung als letzte Phase der Informationsgewinnung aus der V&V herausgenommen und explizit aufgeführt, um die Vernetzung zur Modellbildung zu dokumentieren.

3 Prozesse der Informationsgewinnung

Die Beschreibung der Informationsgewinnung in einem zeitlich strukturierten Kontext erfolgt anhand eines prozessorientierten Vorgehensmodells in Anlehnung an das im SFB 559 verwendeten Prozesskettenparadigma. In diesem Modell wird durch die explizite Differenzierung in eine Informations- und Datensicht innerhalb der Prozesse die Möglichkeit geschaffen, den gesamten Handlungsspielraum von der Zieldefinition über die Ermittlung des notwendigen Informationsbedarfs bis hin zu den tatsächlich nutzbaren Eingangsdaten aufzuspinnen. Als Ressourcen eines jeden Prozesses stehen hierbei vor allem die Methoden der Datenerhebung [HJL04], [JMa05], Statistik [Fen02] und Visualisierung [BJW03], [BWe04] sowie deren kombinierte Anwendung aufgrund kausaler Abhängigkeiten im Mittelpunkt. Somit erlaubt das Vorgehensmodell (siehe Abbildung 4) neben der eindeutigen Anordnung der jeweiligen Prozessschritte und deren strukturiertem Ablauf vor allem die integrative und zielorientierte Nutzung der Methoden (siehe u. a. [WBe03]; [BWe05]).

In den folgenden Unterkapiteln werden die einzelnen Prozesse aus Abbildung 4 detailliert. Die in Abbildung 3 verdeutlichte durchgängige Verifikation und Validierung der Information und Daten werden in jedem Prozess aufgegriffen und prozessweise erläutert.

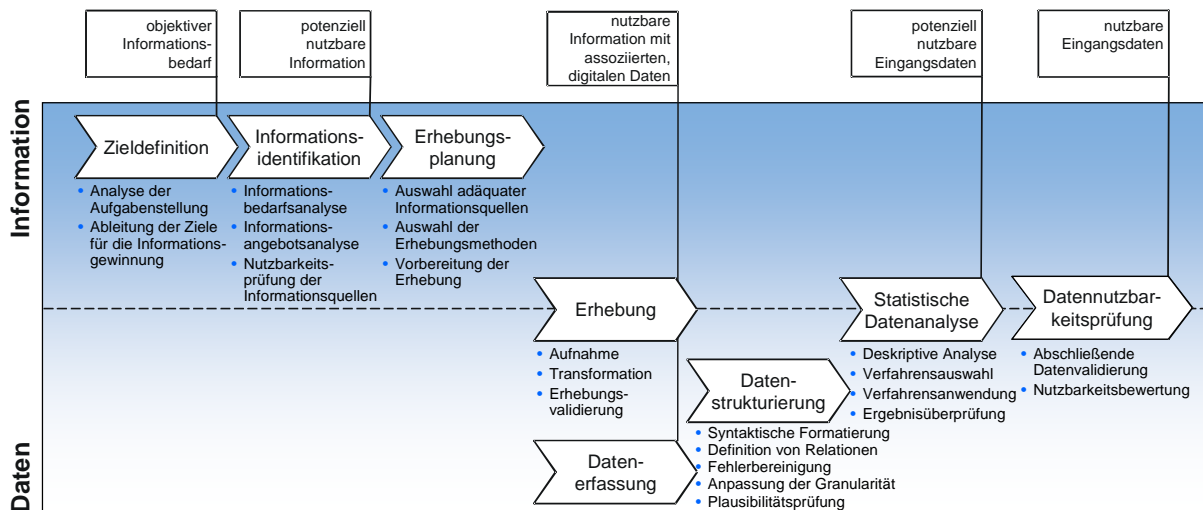


Abbildung 4: Prozessorientiertes Vorgehensmodell zur Informationsgewinnung

3.1 Zieldefinition

Der Prozess der Informationsgewinnung beginnt mit einer Analyse der gegebenen Aufgabenstellung. Unter Berücksichtigung des Untersuchungsgegenstandes, der Studienziele und der ausgewählten Analyse- und Modellierungsmethode wird der notwendige objektive, theoretische Informationsbedarf hergeleitet (siehe Abbildung 5).

3.1.1 Analyse der Aufgabenstellung

Die Analyse der Aufgabenstellung beinhaltet eine Detaillierung der Aufgabenstellung [HJM07], die sich dazu auf Basisinformationen wie vorhandene Kundeninformationen, die Kundenanfrage bzw. eine Ausschreibung, das Angebot, ein Lasten- und Pflichtenheft ([VDI97], [CRW+07]) stützt. Diese Informationen dienen dem Anwender zur Schaffung der Klarheit bzgl. der Aufgabenstellung. Unter Berücksichtigung der übergeordneten Aufgabenstellung, der Systembeschreibung sowie der gewählten Analyse- und Modellierungsmethode erfolgt eine weitere Detaillierung der Aufgabenstellung in Bezug auf die richtige Granularität, die Abbildungsform sowie die Relevanz der Information in Bezug auf das spätere Analysemodell. Mit einer Vorauswahl der zur Beschreibung des Untersuchungsgegenstandes zu verwendenden Standardprozesse [HHN+07] kann im nächsten Schritt eine effektive Informationsidentifikation erfolgen.

3.1.2 Ableitung der Ziele für die Informationsgewinnung

Die Ableitung der Ziele für die Informationsgewinnung stellt die Festlegung eines objektiven Informationsbedarfs für die gegebene Aufgabenstellung bei besonderer Berücksichtigung der Analyse- und der Systembeschreibung dar. Hierbei handelt es sich um einen theoretischen Bedarf, der idealerweise aufgrund der Vorgaben zu erfüllen ist. Zu seiner Ermittlung dienen neben der zuvor gewonnenen Klarheit bzgl. der Aufgabenstellung in Form eines erweiterten Lasten- und Pflichtenheftes insbesondere die zu verwendenden Standardprozesse und der damit verbundene Logistikdatensatz, der projiziert auf den jeweiligen logistischen Standardprozess den dafür typischen Informationsbedarf liefert. Unter dem Logistikdatensatz wird die Gesamtheit der zur Beschreibung logistischer Prozesse notwendigen Daten verstanden. Somit können konkrete Ziele der Informationsgewinnung formuliert, nach Haupt- und Unterzielen strukturiert und gegebenenfalls in einem nochmals erweiterten Pflichtenheft ergänzt werden. Letztlich führt die Zieldefinition zur Festlegung eines objektiven Informationsbedarfs für die gegebene Aufgabenstellung.

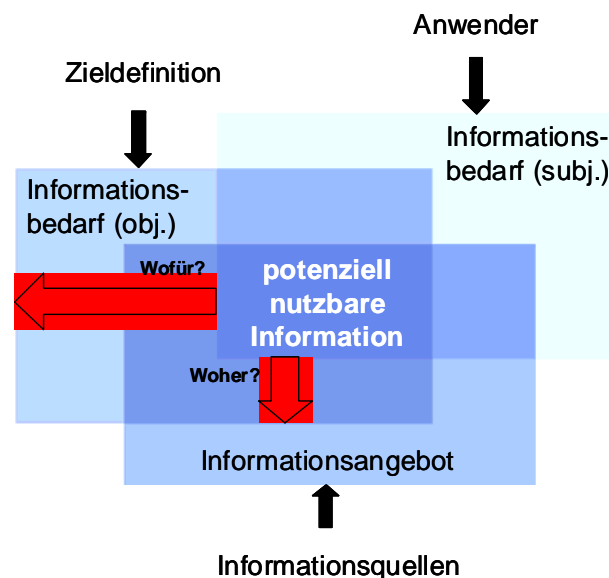


Abbildung 5: Informationsraum in Anlehnung an [Wey01]

3.2 Informationsidentifikation

Zentraler vorbereitender Schritt zur Anwendung der Methoden der Informationsgewinnung ist neben der Zieldefinition die Informationsidentifikation. Dieser setzt sich zusammen aus der Informationsbedarfsanalyse, der Informationsangebotsanalyse und -bewertung sowie einer Nutzbarkeitsprüfung der Informationsquellen, wobei im Folgenden explizit Lösungsansätze, Methoden und Verfahren für die Informationsbedarfsanalyse vorgestellt werden.

3.2.1 Informationsbedarfsanalyse

Die Informationsbedarfsanalyse stellt die Herleitung des subjektiven Informationsbedarfs aus der objektiv für eine Aufgabenstellung und Zieldefinition notwendigen Information dar (siehe Abbildung 5). Der subjektive Informationsbedarf entspricht dabei dem Bedarf, der durch den Planer offensichtlich als notwendig erkannt wurde, was je nach Erfahrung und Qualifikation zu einer nicht vollständigen Überdeckung mit dem objektiven Informationsbedarf führen kann. Dies geschieht durch eine methodische und strukturierte Analyse des objektiven Informationsbedarfs in Zusammenhang mit der problembezogenen Detaillierung und Konkretisierung der Aufgabenstellung und Zieldefinition, z. B. unter Nutzung empirischer Methoden der Informationsbedarfsanalyse und der Markt- und Benutzerforschung [Klu97]. Mit der Projektion des generellen Logistikdatensatzes [BHJ05] auf die zu verwendenden Standardprozesse [HHN+07] unter Berücksichtigung der Aufgabenstellung, des konzeptuellen Modells sowie der Analyse- und der Systembeschreibung können Mindestanforderungen bzgl. der

Granularität, der Darstellungsform sowie der Relevanz der Information benannt werden. Hieraus resultiert eine Spezifizierung des benötigten subjektiven Informationsbedarfs.

3.2.2 Informationsangebotsanalyse

In der Informationsangebotsanalyse und -bewertung wird anhand des subjektiven Informationsbedarfs eines Anwenders sowie des erweiterten Pflichtenhefts eine Sichtung der bzw. aller bekannter Informationsquellen, die möglicherweise die benötigte Information liefern können, durchgeführt. Diese Quellen werden anschließend bzgl. der Verfügbarkeit, der Zugriffsmöglichkeit, des Erhebungsaufwands und der Qualität der Quelle sowie der zu erwartenden Information (z. B. Aktualität, Granularität, Verlässlichkeit, Güte, etc.) bewertet.

3.2.3 Nutzbarkeitsprüfung der Informationsquellen

Die Nutzbarkeitsprüfung der Informationsquellen beinhaltet den Abgleich des Informationsangebots und -bedarfs unter besonderer Berücksichtigung der Qualität der Information bzw. Informationsquellen mit dem subjektiven Informationsbedarf. Basis dieser Überprüfung sind die zuvor identifizierten Informationsquellen und das daraus ermittelte und bewertete Informationsangebot zusammen mit dem subjektiven Informationsbedarf. Daraufhin wird eine Überprüfung der Quellenverfügbarkeit bzgl. des Erhebungsaufwands durchgeführt. Letztlich führt der Gesamtprozess der Informationsidentifikation zu einer Festlegung der potenziell nutzbaren Information mit Bezug zu ihrer Herkunft (Informationsquellen), der zu erwartenden Qualität [Epp06] sowie des damit verknüpften Erhebungsaufwands.

3.3 Erhebungsplanung

In der Erhebungsplanung wird die eigentliche Erhebung von Informationen und / oder Daten korrespondierend mit den Ergebnissen der Informationsidentifikation soweit organisatorisch vorbereitet, dass der Anwender in der Lage ist, ohne zusätzliches Expertenwissen die notwendigen Informationen / Daten zu detektieren und sie mit geeigneten Methoden zu erheben. Dazu erfolgt eine Auswahl adäquater Informationsquellen sowie dazu passender Erhebungsmethoden, die der Vorbereitung der Erhebung vorangehen.

3.3.1 Auswahl adäquater Informationsquellen

Basierend auf den in der Informationsidentifikation aufgestellten Bewertungen findet in diesem Schritt die Selektion adäquater Informationsquellen mit der im Hinblick auf die Erhebung bestmöglichen Effizienz statt. Die Effizienz entspricht hierbei im Sinne einer Nutzwertanalyse [Zan70] einer Relation des Erhebungsaufwandes zu den damit gewonnenen Ergebnissen der Erhebung in Bezug auf die in der Informationsangebotsanalyse (Kap. 3.2.2) genannten Faktoren. Zu diesem Zweck werden die Quellen konkret gesichtet und im Hinblick auf mögliche Erhebungsmethoden untersucht. Als Ergebnis werden dann die vor dem Hintergrund des subjektiven Informationsbedarfs geeigneten Informationsquellen für die Erhebung mitsamt ihrer Priorisierung ausgewiesen.

3.3.2 Auswahl der Erhebungsmethode

Auf Basis der ausgewählten Informationsquellen werden die entsprechenden Erhebungsmethoden aus der Methodensystematik [HJL04], [JMa05] bestimmt. Dazu werden zum einen entsprechend der Informationsquellen zur Primär- und Sekundärerhebung und zum anderen gemäß der Projektrandbedingungen (z. B. Ziele, Budget, Zeit, verfügbare Ressourcen) die relevanten Methoden ausgewählt, wobei verschiedene Kriterien die Auswahl unterstützen. Diese Kriterien sind zum Beispiel der geschätzte Zeitverbrauch, Personalaufwand, Kostenaufwand, die mögliche Beeinflussung betrieblicher Abläufe oder die Fehleranfälligkeit des Verfahrens. Ergebnis dieses Prozessschrittes ist eine Rangliste mit einsetzbaren Erhebungsmethoden bezogen auf die jeweilige Informationsquelle.

3.3.3 Vorbereitung der Erhebung

Während der Erhebungsvorbereitung sind für die jeweils ausgewählten Erhebungsmethoden entsprechende Erhebungsunterlagen zu erstellen, Genehmigungen einzuholen und Pretests durchzuführen. Je nach Methode müssen hier z. B. Fragebögen entwickelt, Personal geschult, Messaufbauten errichtet und geprüft oder Sekundärmaterial beschafft werden. Weiterhin sind an den Erhebungen zu beteiligende Personen zu informieren, Termine zu vereinbaren und Genehmigungen bei Betriebsräten, Behörden, usw. einzuholen. In einer ausführlichen Erhebungsanleitung werden Informationen wie die Vorgehensweise bei der zu verwendenden Methode und eine Beschreibung der Informationsquelle hinterlegt.

Weiterhin ist bei der Vorbereitung der Erhebung eine Voruntersuchung, der sogenannte Pretest [Voß04], durchzuführen, bei dem eine Überprüfung der erstellten Erhebungsinstrumente wie z. B. des Fragebogens im Hinblick darauf stattfindet, ob sie valide (gültige) und reliable (zuverlässige) Erhebungsergebnisse ermöglichen.

3.4 Erhebung und Datenerfassung

Bei der Datenerhebung geht es darum, die in der Informationsbedarfsanalyse geforderten Daten und Informationen wie in der Erhebungsplanung beschrieben und vorbereitet zum einen aus ausgewählten Quellen zu extrahieren und sie zum anderen für den nachfolgenden Prozessschritt der Analyse gezielt als Daten bereitzustellen. Hierzu sind entweder die separate Aufnahme der Informationen inklusive einer ggf. notwendigen Transformation und anschließender Datenerfassung oder eine integrierte Datenerfassung durchzuführen, deren Output in einer den Prozess der Erhebung abschließenden Erhebungsvalidierung geprüft wird.

3.4.1 Aufnahme

Je nach verwendeter Methode werden Information und / oder Daten aufgenommen und protokolliert. Zur Durchführung kann die in [JMa05] vorgestellte Methoden- und Vorgehensbeschreibung herangezogen werden. Die sogenannten automatischen Methoden [Kro06], [Kuc05] sind in der Regel in der Lage, die Daten direkt elektronisch zu speichern (protokollieren). Bei den manuellen Methoden müssen die erhobenen Daten oder Informationen, die beispielsweise auf Papier protokolliert sind, noch mittels einer Datenerfassung in elektronische Daten gewandelt werden. In diesem Fall ist die Datenerfassung ein gesonderter Schritt, der in dieser Systematik parallel zur Erhebung dargestellt ist. Zusätzlich ist die durchgeführte Aufnahme zu protokollieren [Sch05], so dass das Ergebnis dieses Prozessschrittes erhobene Informationen und / oder Daten und die dazugehörige Erhebungsdokumentation ist.

3.4.2 Transformation

Der Schritt der Transformation wird notwendig, falls die Erhebung von Informationen nicht in geschlossener, sondern in offener Form (vgl. [HJL04]), z. B. über einen nicht standardisierten Fragebogen oder über ein Interview, erfolgt und somit keine standardisierten Antworten erhoben werden. In diesem Fall muss die erhobene Information erst noch interpretiert werden, um sie in die für die Aufgabe tatsächlich zu verwendende Information zu wandeln. Erfolgt die Transformation zeitgleich mit einer Datenerfassung, werden die Informationen direkt in digitale Daten überführt.

Sind Informationen geschlossen, also standardisiert, erhoben oder bereits Daten aufgenommen, entfällt dieser Schritt.

3.4.3 Datenerfassung

Der Prozess der Datenerfassung schließt alle manuellen und automatisierten Operationen ein, um Daten in eine maschinenlesbare Struktur zu übertragen und sie auf einem

Datenträger zu speichern. Die Daten können hierbei manuell, halbautomatisch oder automatisch erfasst werden [JMa05]. Innerhalb der manuellen Erfassung werden die Daten z. B. über eine Tastatur in ein Computersystem eingegeben. Ein halbautomatischer Aufnahmeprozess wird durch einen expliziten Startbefehl für jede Beobachtung eingeleitet. Ein Beispiel für diese Methode ist der Gebrauch eines Handscanners für das Ablesen eines Barcode-Labels. Hier wird der Leseprozess manuell eingeleitet und daraufhin automatisch durchgeführt. Ein anderes Beispiel ist die manuelle Zusammenführung von automatisch erfassten Teildaten. Eine automatische Aufnahmemethode liegt vor, wenn der Aufnahmeprozess nicht direkt beeinflusst werden kann. Dies geschieht z. B. beim Einsatz von Fotosensoren, um den Durchsatz von Einzelförderern in einem Materialflusssystem zu ermitteln. Nach dem Prozess der Datenerfassung sind die Daten in digitaler Form vorhanden. Die Datenerfassung kann vorgehensabhängig zu verschiedenen Zeitpunkten der Erhebung erfolgen: gleich zu Beginn z. B. bei automatischen Erhebungsmethoden, zwischendurch z. B. im Verlauf der Transformation oder zum Abschluss z. B. nach einer Multimomentanalyse.

3.4.4 Erhebungsvalidierung

Bei der Erhebungsvalidierung werden die erhobenen Daten mit dem ursprünglichen subjektiven Informationsbedarf abgeglichen. Dabei werden sowohl der Erhebungsvorgang als auch die Daten im Hinblick auf Qualität und Vollständigkeit überprüft. Mit diesem die Erhebung abschließenden Prozessschritt stehen dem Anwender dann nutzbare Informationen mit assoziierten digitalen Daten, die dazugehörigen Informationen über Wertebereiche von Variablen sowie die endgültige Erhebungsdokumentation zur Verfügung.

3.5 Datenstrukturierung

Durch die Einzelprozesse der Datenstrukturierung werden die mit Ende der Erhebung digital vorliegenden Daten zunächst in eine analysefähige Form gebracht (syntaktische Formatierung). Anschließend werden Beziehungen zwischen den Variablen und zwischen möglichen Tabellen charakterisiert (Definition von Relationen), falsche Einträge bereinigt und verdächtige, möglicherweise falsche Beobachtungen bzw. Datensätze untersucht (Fehlerbereinigung) und die Daten in die nötige Granularitätsform gebracht (Anpassung der Granularität). Diese Reihenfolge ist dabei optional und kann situationsspezifisch geändert werden. Abschließend erfolgt mit einer Plausibilitätsprüfung eine Kontrolle der dann vorliegenden Datenbank hinsichtlich der Richtigkeit der vorgenommenen Änderungen und bezüglich der Weiterverwendbarkeit in den folgenden Schritten.

3.5.1 Syntaktische Formatierung

Mit Abschluss der Erhebungsvalidierung sind die erhobenen Daten zwar digital, aber nicht zwangsläufig in der für die weitere Analyse notwendigen bzw. gewünschten Form verfügbar. Im Formatierungsschritt werden sie in ein später leicht les- und verwendbares, einheitliches Dateiformat konvertiert. Hierzu zählt z. B. die Umwandlung von String-Einträgen in charaktertypische Datenformate wie Boolean, Integer, Datum, o. ä. Anschließend werden die Einzeldaten in einem jeweiligen Datenbank- oder Tabellenformat geschlossen dargestellt. Schließlich wird die Richtigkeit der Formatierung überprüft. Im Folgenden wird die Gesamtheit der so formatierten Daten als die vorliegende Datenbank bezeichnet.

3.5.2 Definition von Relationen

Dieser Prozessschritt erklärt den Zusammenhang und den gegenseitigen Bezug zwischen den verschiedenen Variablen, Beobachtungen der erhobenen Daten oder ggf. zusammenhängenden Teilen der Datenbank.

Zunächst werden zwischen den Variablen bestehende, „natürliche“ Abhängigkeiten oder Hierarchien beschrieben. So können manche Variablen nur dann beobachtbar sein, wenn eine andere Variable einen bestimmten Wert annimmt. Ebenso kann der Wertebereich einer Variablen von der Ausprägung einer anderen Variable abhängen. Als Beispiel betrachte man

die Lagermengen von Stückgütern in verschiedenen Lagerstandorten: Die Standorte selbst stellen eine Variable dar, ohne die die Angabe einer Lagermenge nicht aussagekräftig ist. Damit sind die Lagermengen eine vom Standort abhängige Variable. Sollten sich dann die Lager durch ihr Lagervolumen unterscheiden, hat dies Einfluss auf die möglichen Lagermengen. Diese Ausprägungen sind also vom jeweiligen Lagerstandort abhängig. Mit Hilfe von kontextbezogenem Hintergrundwissen lässt sich weiterhin festzustellen, ob eine Abhängigkeit unmöglich, möglich, unbekannt oder bekannt ist.

Darüber hinaus können weitere Relationen zwischen Variablen bestehen, wenn diese eine gemeinsame „Dimension“ besitzen wie z. B. zeitliche und räumliche Eigenschaften.

Sind mit der Erhebung Daten in verschiedenen Tabellen erhoben worden, müssen diese aggregiert werden: Verschiedene Variablen, die am selben Merkmalsträger beobachtet wurden, sollten einander zugeordnet werden. Umfassen die Daten darüber hinaus mehrere gleich aufgebaute Tabellen, ist es häufig sinnvoll, diese zu einer Gesamttabelle zusammenzufügen.

3.5.3 Fehlerbereinigung

In diesem Schritt wird die Datenbank auf Eintragungen untersucht, die im logischen Widerspruch zu der erhobenen Information steht. Dabei werden die in der Erhebung gewonnenen Kenntnisse über die Wertebereiche, Variablen sowie die Erhebungsdokumentation verwendet. Als allgemeine, einführende Referenzen können z. B. [Dav96] oder [DJo03] herangezogen werden.

Zunächst wird untersucht, welche Eintragungen in der Datenbank aufgrund der Eigenschaft der zugehörigen Variablen als falsch anzusehen sind. Durch den Typ und Wertebereich ergeben sich überprüfbare Eigenschaften der Variableneinträge. So dürfen z. B. numerische Variablen keine Textangaben enthalten oder dichotome Variablen nicht mehr als zwei verschiedene Ausprägungen aufweisen. Variableneinträge, die außerhalb des Definitions- bzw. Wertebereiches liegen, wie z. B. nicht-ganzzahlige Stückzahlen, müssen demnach ebenfalls als Fehler angesehen werden. Außerdem ist zu prüfen, ob bestimmte Kombinationen von Variablenausprägungen nicht möglich sind (z. B. bei den Variablen Monat und Tag die Kombination „Februar“, „30“). Die Information über die Variablenart und ihren Wertebereich kann sich aufgrund logischer Eigenschaften ergeben oder im Zuge der Erhebungsvalidierung (Prozess 4: Erhebung und Erfassung) ermittelt worden sein und liegt damit in jedem Fall vor.

Außer derart einfach auffindbaren Fehlern umfasst dieser Prozessschritt auch das Auffinden und Untersuchen von Ausreißern, d. h. Beobachtungen, die aufgrund ihrer relativen Position zu anderen Einträgen auffällig sind. Da es sich bei diesen sowohl um mögliche Fehler wie auch um einzelne, möglicherweise unwahrscheinliche aber reale Beobachtungen handeln kann, ist eine gesonderte Analyse erforderlich. Erschwert wird dies durch die Tatsache, dass der Begriff des Ausreißers i. d. R. nicht eindeutig und nur relativ zu bestimmten Modellannahmen definiert werden kann (vgl. [BLe94], [GKP03]). Als erstes ist mit einer eindimensionalen Betrachtung der jeweiligen Variablen zu beginnen. Handelt es sich um numerische Variablen, können hier Visualisierungsverfahren oder deskriptive Verfahren wie Stabdiagramme, Boxplots, Lage- und Schiefemaße verwendet werden (vgl. Prozessschritt Deskriptive Analyse in Kapitel 3.6.1). Hierbei ist die Variablenstruktur, z. B. bei zeitlichen oder räumlichen Daten, zu berücksichtigen. Ein Vergleich der Ergebnisse von sogenannten robusten Verfahren mit nichtrobusten kann einen ersten Einblick in die Ausreißerstruktur der Daten geben.

Nach dem Auffinden von fehlerhaften Daten muss die Datenbank um diese bereinigt werden. Als tatsächlich fehlerhaft erkannte Eintragungen sind zu löschen und als fehlender Wert (missing value) zu markieren. Nur in den Fällen, in denen sich die wahre Ausprägung logisch und eindeutig aus dem Vergleich mit den übrigen Daten ergibt (z. B. die Tageszeit aus der Uhrzeit), kann ein korrigierter Wert eingesetzt werden. Anschließend werden Beobachtungen, die aufgrund ihrer besonderen Position im Vergleich zu anderen Datensätzen als

Ausreißer ermittelt wurden, in einer begleitenden Dokumentation als solche bezeichnet. Stehen Informationen aus ähnlichen Datenbanken zur Verfügung, können diese nach Möglichkeit benutzt werden, um die Ausreißer noch genauer zu untersuchen. Die Beurteilung der Ausreißer sollte zurückhaltend erfolgen. Verdächtige Beobachtungen dürfen in keinem Fall automatisch als Fehler deklariert und gelöscht oder ersetzt werden. Vielmehr müssen die späteren statistischen Analyseverfahren auf ausreißerbehaftete Daten anwendbar sein.

Die Datendokumentation wird zusammen mit der fehlerbereinigten Datenbank an die folgenden Prozessschritte übergeben.

3.5.4 Anpassung der Granularität

Die Einträge in der Datenbank werden zunächst für jede Variable in jeweils einheitliche Maßeinheiten umgewandelt. Dies ist insbesondere notwendig, wenn das Datenmaterial in verschiedenen Tabellen vorliegt, die möglicherweise noch durch verschiedene Erhebungsverfahren erstellt wurden. Für jede Variable muss dazu festgelegt werden, in welcher Einheit die Ausprägungen fortan gespeichert werden sollen und mit welcher Genauigkeit. Als Vorgaben gehen daher der subjektive Informationsbedarf und insbesondere die Mindestanforderungen bzgl. der Granularität und des Detaillierungsgrads ein. Diese Anforderungen können sich auf die räumliche und zeitliche Auflösung (wie viele Beobachtungen pro Untersuchungseinheit und Zeitintervall) aber auch auf die Genauigkeit der Beobachtungen selbst beziehen (z. B. Anzahl an Dezimalstellen bei numerischen Variablen oder maximaler Umfang der Zeichenkette bei Text-Angaben).

Nach dieser Standardisierung können je nach Anforderung weitere granularitätsanpassende Schritte erfolgen. Hierzu zählen beispielsweise die (Neu-)Kodierung von Variablen (wie z. B. das Umwandeln dichotomer Ausprägungen in 0 / 1 - Werte oder die eindeutige Umwandlung von Texten in Zahlenwerte) sowie die Bildung von übergreifenden Klassen oder Kategorien. Im letztgenannten Fall sind statistische Klassierungsverfahren einzusetzen, wenn die durch den subjektiven Informationsbedarf gegebenen Anforderungen Freiräume zur Gestaltung der Klassen lassen (wie Klassenzahl oder -breite). Beispiele sind hier Histogrammschätzung und Clusteranalyse (z. B. [Rom04], [Rud99]).

Es ist zu berücksichtigen, dass durch die Anpassung der Granularität die Genauigkeit und damit der Informationsgehalt der Daten in keinem Fall erhöht werden kann.

Die vorgenommenen Änderungen werden in der Datendokumentation protokolliert und zusammen mit der nun granularitätsangepassten Datenbank an den nächsten Prozessschritt übergeben.

3.5.5 Plausibilitätsprüfung

Abschließend erfolgt die Prüfung der Datenbank hinsichtlich Vollständigkeit, Widerspruchsfreiheit und denjenigen Anforderungen, die durch den subjektiven Informationsbedarf gestellt werden. Hierzu werden die Daten- und die Erhebungsdokumentation genutzt. Dies umfasst Aspekte der Datenverifikation sowie der Datenvalidierung.

Zunächst muss durch eine Datenverifikation die inhaltliche Richtigkeit und Widerspruchsfreiheit der Daten überprüft werden. Sofern sich aus theoretischem Wissen oder experimentellen Ergebnissen bestimmte Eigenschaften der Daten ergeben, müssen diese nachgewiesen werden. Hierunter fallen beispielsweise Verteilungseigenschaften von Variablen (wie exponential verteilte Zwischenankunftszeiten). Insbesondere ist nachzuweisen, dass durch die zuvor definierten Relationen und Abhängigkeiten innerhalb der Daten keine Unstimmigkeiten entstehen. Die Anzahl und Art der in der Fehlerbereinigung ermittelten fehlenden Werte und Ausreißer müssen ebenfalls bewertet werden. Dabei ist zu berücksichtigen, welche Qualitätsanforderungen realistischer Weise an die Datenbank gestellt werden können, wozu der Vergleich mit ähnlichen, historischen Studien hilfreich ist.

Positive Ergebnisse dieser Untersuchungen sind die notwendige Voraussetzung, um anschließend durch eine Datenvalidierung zu entscheiden, ob der subjektive Informations-

bedarf durch das Datenmaterial befriedigt werden kann. Dafür ist zu zeigen, dass der Umfang der Datenbank genügend groß ist und dass die Repräsentativität der Daten (vgl. z. B. [RWA85] oder [Kre93]) bzgl. des Untersuchungsziels trotz der ggf. festgestellten Qualitätseinbußen durch Fehler oder Ausreißer weiterhin gewährleistet ist.

Ist das Ergebnis der Plausibilitätsprüfung insofern negativ, dass die Daten nicht im erforderlichen Ausmaß verifiziert oder validiert werden konnten, so muss zwangsläufig eine neue Erhebung oder Nacherhebung stattfinden, wofür je nach Situation evtl. neue Informationsquellen erschlossen werden müssen. In jedem Fall sind die Ergebnisse der Plausibilitätsprüfung in einer die Datendokumentation ergänzende Qualitätsbewertung festzuhalten.

3.6 Statistische Datenanalyse

Der Prozess der statistischen Datenanalyse erstellt auf Grundlage der Datenbank, die die bereinigten, strukturierten Datensätze enthält, Eingangsdaten für die später folgende modellbasierte Analyse. Nach einer allgemeinen deskriptiven Analyse werden statistische Verfahren ausgewählt, die den gegebenen Informationsbedarf am besten erfüllen. Nach Durchführung der Verfahren und Interpretation der Ergebnisse werden diese abschließend hinsichtlich ihrer weiteren Verwendbarkeit überprüft.

3.6.1 Deskriptive Analyse

Die deskriptive Analyse dient der Beschreibung allgemeiner statistischer Eigenschaften der Beobachtungen in der Datenbank. Hierzu zählen die Berechnung von Verteilungskennzahlen wie z. B. Lage- und Streuungsparameter von Variablen sowie die grafische Veranschaulichung der empirischen Verteilungen z. B. durch Histogramme oder Boxplots. Zusammenhänge zwischen Variablen, die durch Definition von Relationen aufgedeckt werden, können ebenfalls durch Kennzahlen wie z. B. Korrelationskoeffizienten oder durch Visualisierungsverfahren wie Scatterplots charakterisiert werden [BDW05]. Die Auswahl geeigneter deskriptiver Verfahren sollte sich am subjektiven Informationsbedarf orientieren. Damit erlauben die Ergebnisse einen zwar groben aber schnell verständlichen Überblick für wesentliche Eigenschaften der Beobachtungen. Eine Übersicht über deskriptive Methoden enthalten z. B. [BCK04], [HEK05] oder [TSW06].

3.6.2 Verfahrensauswahl

Dieser Prozessschritt stellt die zentrale Verknüpfung zwischen der Ausgangsaufgabenstellung, die den subjektiven Informationsbedarf erzeugt hat, und den späteren Eingangsdaten für das weitere Modell dar. Die Auswahl der statistischen Verfahren muss damit sicherstellen, dass alle Anforderungen, die durch den subjektiven Informationsbedarf gegeben sind, durch die Ergebnisse der verschiedenen Verfahren abgedeckt werden.

Zu Beginn muss unter Verwendung des subjektiven Informationsbedarfs die statistische Fragestellung ermittelt werden. Daraus lassen sich Anforderungen an die zu verwendenden statistischen Verfahren hinsichtlich der gewünschte Art und Struktur der Ergebnisse ableiten. Weitere Nebenbedingungen an Verfahren bezüglich der Art der Eingangsdaten ergeben sich durch die Eigenschaften des Datenmaterials, die aus der Datendokumentation und vor allem aus der deskriptiven Analyse abgeleitet werden können. Um geeignete Verfahren zu ermitteln, ist die Verwendung eines Methodenbaukastens hilfreich. Dieser liefert für den gegebenen subjektiven Informationsbedarf und die Datenstruktur, wie sie durch die Schritte Datendokumentation und Deskriptive Analyse beschreibbar ist, alle verwendbaren Verfahren, die dort in Methodenkategorien hinterlegt sind. Derartige Methodenkategorien werden für die in diesem Prozessschritt relevanten statistischen und Visualisierungsverfahren in [FKK05] und [BDW05] vorgestellt. Gibt es kein einzelnes Verfahren, muss überlegt werden, durch welche Kombinationen von Verfahren die Zielvorgaben erfüllt werden können. Nach einem Vergleich der möglichen Verfahren wird anschließend eines davon oder eine Abfolge verschiedener gewählt, mit dem die folgende Analyse durchzuführen ist. Hierbei sind Kriterien wie Verfügbarkeit, Programmierbarkeit oder

Umsetzbarkeit, Kosten und Dauer der Durchführung sowie erwartete Genauigkeit des Ergebnisses kontextbezogen zu berücksichtigen.

An den folgenden Prozessschritt werden schließlich die ausgewählten Verfahren mit Angabe über die Reihenfolge ihrer Durchführung sowie die dafür notwendigen Teile der Datenbank weitergegeben.

3.6.3 Verfahrensanwendung

Die im vorigen Prozessschritt ausgewählten Verfahren werden unter Verwendung statistischer Software durchgeführt. Alle Ergebnisse sowie mögliche auftretende Fehler oder Warnmeldungen werden protokolliert. Abschließend erfolgt eine Interpretation der Ergebnisse hinsichtlich ihrer statistischen Eigenschaften.

3.6.4 Ergebnisüberprüfung

Die Ergebnisse und ihre Interpretation werden zunächst auf Vollständigkeit und dann auf ihre Richtigkeit hin überprüft (Datenverifikation): Hierunter fallen Aspekte der inhaltlich-logischen und statistischen Richtigkeit. Inhaltlich-logische Richtigkeit kann durch Plausibilitätsuntersuchungen überprüft werden, in denen die Resultate und Folgerungen der Analyse auf ihre inhaltliche Widerspruchsfreiheit hin untersucht werden. Zur Überprüfung der statistischen Richtigkeit muss vor allem die Gültigkeit der jeweiligen Modellannahmen, z. B. im Fall von Regressionsmodellen durch Residualanalyse, nachgewiesen werden. Außerdem muss die Fehlerfreiheit verwendeter Software oder Algorithmen nachgewiesen werden.

Mit den Ergebnissen der statistischen Verfahren werden anschließend die potenziell nutzbaren Eingangsdaten erzeugt, sofern sie nicht bereits Ergebnis des vorigen Prozessschrittes sind, und dokumentiert.

Ist die Richtigkeit der Ergebnisse nicht gewährleistet oder fraglich, sind an den jeweiligen Stellen Korrekturen notwendig. Je nach Ursache kann daher an dieser Stelle eine Neudurchführung der Prozesse Erhebung und/oder Statistische Datenanalyse notwendig sein.

3.7 Datennutzbarkeitsprüfung

Der Prozessschritt der Datennutzbarkeitsprüfung umfasst die zwei Schritte abschließende Datenvalidierung sowie Nutzbarkeitsbewertung und führt zu nutzbaren Eingangsdaten für das übergeordnete, ausführbare Simulationsmodell.

3.7.1 Abschließende Datenvalidierung

Der Prozessschritt der abschließenden Datenvalidierung ist eng verzahnt mit den vorher stattfindenden Prozessschritten zur statistischen Datenanalyse. Im Gegensatz zur Datenverifikation (Sind die Daten richtig?) während der Datenstrukturierung und -analyse übernimmt die Datenvalidierung (Sind es die richtigen Daten?) die Überprüfung, ob die potenziell nutzbaren Eingangsdaten dem subjektiven Informationsbedarf hinsichtlich der Anforderungen Plausibilität, Vollständigkeit, Glaubwürdigkeit und Richtigkeit für die Anwendung genügen (vgl. [Bal98a], [Bal03]). Ein Abgleich mit dem ursprünglichen subjektiven Informationsbedarf sowie die Interpretation der statistischen Analyseergebnisse sind bei dieser Prüfung zwingend notwendig. Zu den in diesem Kontext einzusetzenden Methoden (vgl. beispielsweise [Epp06]) gehören Checking Order, Plausibilitätstests, vergleichende Second Source Tests, Sensitivitätsanalysen, Reviewing oder Ratings aber auch Visualisierungsverfahren zur Darstellung vergleichender Analysen (vgl. [BJW03]). Letztendlich stellt die Datenvalidierung sicher, dass die erzeugten potenziell nutzbaren Daten für den subjektiven Informationsbedarf zweckdienlich sind (Sind dies die richtigen Daten zur Erfüllung des Informationsbedarfs?).

3.7.2 Nutzbarkeitsbewertung

Der zweite Prüfschritt, die Nutzbarkeitsbewertung, umfasst die Analyse der validierten potenziell nutzbaren Eingangsdaten im Kontext des implementierten Simulationsmodells. In diesem Zusammenhang müssen die Daten mit dem Simulationsmodell gemeinsam hinsichtlich der Korrektheit und Angemessenheit überprüft werden. Diese Schritte sind äquivalent zu den Schritten der Verifikation und Validierung (V&V) für die Simulation. Während sich die Verifikation primär auf die Beantwortung der Fragen nach der Korrektheit z. B. in Bezug auf die formale Nutzbarkeit der Daten und Datenquellen bezieht, beantwortet die Validierung in erster Linie die Frage nach der hinreichenden Abbildung des Systems durch das Simulationsmodell einschließlich der Daten in Bezug auf den Untersuchungszweck und das beabsichtigte Systemverhalten (Angemessenheit). Zu untersuchende Kriterien sind u. a. die Eignung und Genauigkeit der Daten beispielsweise in Bezug auf ihre Passgenauigkeit zum Simulationsmodell (Detaillierungsgrad) oder die Eignung für die hinreichende Nachbildung des Systemverhaltens. Klassische V&V-Techniken wie Historical Data Validation, Sensitivity Analysis, Visualization Methods (z. B. Animation), Trace Analysis, Extreme Condition Tests finden hier u. a. ihren Einsatz. Für weitere V&V-Techniken sei auf [Bal98b], für Vorgehensmodelle zu V&V auf [BBH+02] und [RSW07] verwiesen.

Ergebnisse dieses Prozessschrittes sind validierte, im Kontext eines (Simulations-)Modells nutzbare Eingangsdaten. Aus der Dokumentation der Eingangsdaten, der statistischen Analyseergebnisse, der bisherigen Qualitätsbewertung der Datensätze und Informationsquellen sowie der V&V-Ergebnisse wird darüber hinaus eine abschließende Bewertung der für die Simulation nutzbaren Eingangsdaten abgeleitet. Hierzu werden Angaben zu Qualitätskriterien wie Ursprung und Aktualität der Daten sowie Angaben zur Qualität der Informationsquellen zusammengestellt, um für den Anwender Hinweise zur Informationsgüte geben zu können. In diesem Zusammenhang werden auch Regeln zur richtigen Verwendung der Daten im Kontext des (Simulations-)Modells festgelegt.

4 Konzept der integrierten Methodennutzung

Die in den Prozessen des Vorgehensmodells einzusetzenden Methoden entstammen verschiedenen Disziplinen. Hierbei ist es häufig so, dass auch innerhalb eines Prozesses verschiedene Methoden und deren Verfahren z. B. aus der Datenerhebung, Visualisierung oder statistischen Datenanalyse zum Einsatz kommen. Schon aufgrund ihrer unterschiedlichen Herkunft muss daher beachtet werden, dass der Output eines Verfahrens als Input eines nachfolgenden Verfahrens verwendbar ist. Das Prinzip der integrierten Methodennutzung verlangt daher eine präzise Charakterisierung der zur Verfügung stehenden Methoden und wird idealerweise durch Methodenkategorien in einem Methodenbaukasten bereitgestellt. Stellvertretend für die oben genannten Disziplinen sei diesbezüglich auf [HJL04], [JMa05], [BDW05] und [FKK05] verwiesen.

Die Kategorisierung erfolgt dabei sinnvollerweise gemäß einer (möglicherweise hierarchischen) Taxonomie und berücksichtigt insbesondere die Eigenschaften von Input und Output der jeweiligen Methoden. Diese Eigenschaften sind wiederum sowohl auf Daten- wie auch auf Informationsebene zu spezifizieren. Eine umfassende Methodenkategorisierung hilft also Methoden auszuwählen, die geeignet aufeinander aufbauen können. Dazu müssen Input und Output aller Verfahren genau charakterisiert werden. Soll ein konkretes Verfahren zur Lösung einer bestimmten Aufgabenstellung eingesetzt werden, muss also zunächst überlegt werden, ob der nötige Input durch Output der vorangehenden Verfahren hinsichtlich Information und Daten erzeugt wird. Der Input kann dabei durch Output verschiedener Verfahren vorhanden sein, die zu einem beliebigen Zeitpunkt vorher verwendet wurden. Mit Hilfe dieser Überlegungen lässt sich auch planen, ob die Verkettung mehrerer Methoden möglich ist und welches Ergebnis resultieren wird.

Beispielhaft sei folgende Situation betrachtet: Für ein Simulationsmodell, das als Eingangsdaten Zwischenankunftszeiten von Flugzeugen auf einem Luftfrachtknotenpunkt benötigt, sind im Prozessschritt der Erhebung (siehe Abbildung 4) bereits über einen Zeitraum von sechs Wochen ganztägig Zwischenankunftszeiten beobachtet worden und liegen nach der Datenerfassung und Datenstrukturierung in einer plausibilitätsgeprüften Datenbank vor. Nach der deskriptiven Analyse wurde im Prozessschritt der Verfahrensauswahl entschieden, für jede halbe Stunde eines Tages die Verteilung der zugehörigen Zwischenankunftszeiten zu schätzen (Maximum-Likelihood-Schätzung des Lokationsparameters einer Exponentialverteilung). Der Output dieses Verfahrens enthält die Information „Exponentialverteilungen für 48 halbe Stunden, Maximum-Likelihood-Schätzung“ und als Daten 48 Parameterschätzer für jede halbe Stunde sowie die ursprünglichen Beobachtungen. Im anschließenden Prozessschritt der Ergebnisüberprüfung muss die inhaltliche Richtigkeit des Ergebnisses überprüft werden. Es ist möglich, zunächst mit einem Visualisierungsverfahren wie dem Liniendiagramm eine erste grobe Beurteilung zu erhalten. Durch den Informationsoutput der Schätzmethode werden zwei Informationskomponenten (Zeit, Häufigkeit) und zwei Informationsobjekte (beobachtete und geschätzte Funktion) bereitgestellt. Entsprechend werden zwei Achsen für die Informationskomponenten und die beiden darzustellenden Objekte im Diagramm abgebildet. Exemplarisch werden die Resultate in Abbildung 6 für die 5. und 13. halbe Stunde dargestellt.

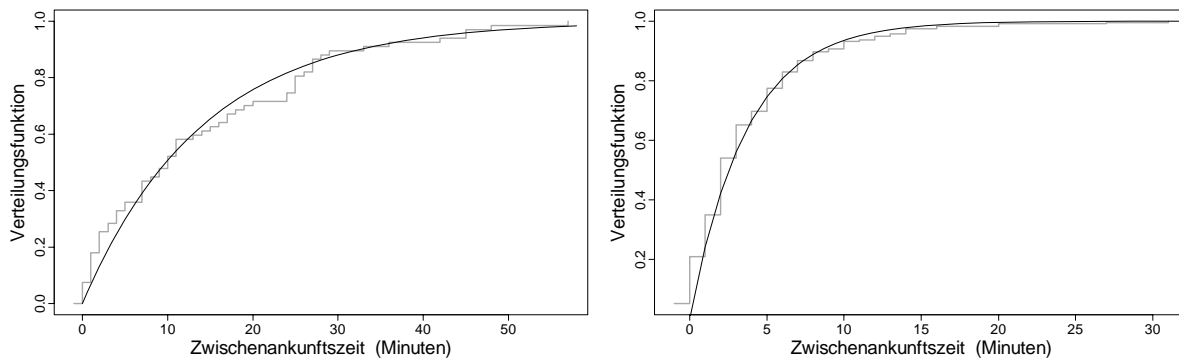


Abbildung 6: Empirische Verteilungsfunktionen (grau) und geschätzte Exponentialverteilungen (schwarz) für die Zwischenankunftszeiten eingehender Flugzeuge in der 5. (links) bzw. 13. (rechts) halben Stunde eines Tages.

Es ist nun einerseits möglich, die Unterschiede zwischen Beobachtungen und Schätzung visuell zu erfassen und zu bewerten, andererseits wird man aus Gründen der Objektivität zusätzlich eine statistische Methode verwenden wollen, um diese Bewertung durchführen zu lassen. Zwar sieht die Anpassung für die 13. halbe Stunde zufriedenstellend aus, aber das Ergebnis für die fünfte halbe Stunde ist nicht eindeutig einzuordnen.

Aufbauend auf dieser Vorgabe kann nun mit dem Ergebnis der Visualisierung der Kolmogoroff-Smirnov Test (vgl. [BTr94]) durchgeführt werden. Dieser nutzt den maximalen vertikalen Abstand zwischen zwei Verteilungsfunktionen, um die Hypothese der Gleichheit einer empirischen mit einer theoretischen Verteilungsfunktion zu überprüfen. Die nötigen Eingangsdaten sind damit direkt aus dem Visualisierungsergebnis ableitbar (Länge des maximalen Abstands und Umfang der Beobachtungen). Auf Informationsebene erfolgen als Input z. B. lediglich die Angaben „Empirische Verteilungsfunktion und Exponentialverteilung mit gegebenem Parameter“ und „Testniveau 5 %“. Auf Datenebene werden die Beobachtungen und der im vorherigen Verfahren geschätzte Parameterwert übergeben. Das Testverfahren berechnet nun auf Basis des Inputs eine Teststatistik, die durch einen Vergleich mit dem durch das vorgegebene Testniveau gegebenen kritischen Wert das Ergebnis produziert: Auf Informationsebene beinhaltet der Output „Testentscheidung zum Niveau 5 %“ und auf Datenebene je nach Resultat entweder „Hypothese ablehnen“ oder

„Hypothese nicht ablehnen“. Im obigen Beispiel der Halbstundenverteilungen ist der Output bei der fünften wie auch bei der dreizehnten halben Stunde „nicht ablehnen“.

5 Ausblick

Für die Informationsgewinnung innerhalb eines übergeordneten Simulationsvorgehensmodells wurde in dieser Arbeit ein detailliertes, prozessorientiertes Vorgehensmodell entwickelt. Im Rahmen der Beschreibung der einzelnen Prozessschritte wurden jeweils Hinweise auf die potentiell einzusetzenden Methoden aus den Bereichen der Datenerhebung, statistischen Datenanalyse und Visualisierung gegeben. Nicht diskutiert wurde die Frage, wer die einzelnen Schritte ausführt. Als Akteure denkbar sind neben dem mit dem substanzwissenschaftlichen Problem vertrauten Anwender, Experten z.B. auf den Gebieten der Simulation, Datenerhebung, Statistik und Visualisierung. Innerhalb der weiteren Diskussion dieses Vorgehensmodells zur Informationsgewinnung sollten die Rollen, die diese Akteure innerhalb der einzelnen Prozessschritte einnehmen, spezifiziert werden.

Ein wichtiges Merkmal des Vorgehensmodells ist die durchgängige Verifikation und Validierung der Zwischenergebnisse am Ende jeden Prozessschrittes. Hier setzt die Weiterentwicklung von geeigneten Kriterien zur Bewertung der Güte der erzielten Informationen an. Als Qualitätskriterien werden üblicherweise z. B. Zugänglichkeit, Bedeutsamkeit, Objektivität und Korrektheit der Informationen herangezogen [KSW02], [Epp06]. Aber auch qualitätsändernde Einflüsse aufgrund der Anwendung von speziellen Methoden der Datenerhebung und Statistik müssen berücksichtigt werden. Erst eine durchgängige Berücksichtigung des Faktors Qualität ermöglicht eindeutige Aussagen bezüglich der Güte der letztendlich gewonnenen nutzbaren Eingangsdaten.

Das hier vorgestellte Vorgehensmodell zur Informationsgewinnung lässt sich in der vorliegenden Form bereits innerhalb von Simulationsstudien einsetzen. Zukünftige Entwicklungen hinsichtlich der leichteren Anwendbarkeit und Bedienbarkeit durch integrierte Methodenbaukästen und Softwareunterstützung im gesamten Informationsbeschaffungsprozess werden die Akzeptanz und Verbreitung der Ergebnisse weiter erhöhen.

6 Literatur

- [Bal98a] Balci, O.: Verification, Validation and Accreditation. In: Medeiros, D. J.; Watson, E. F.; Carson, J. S.; Manivannan, M. S. (Eds.): Proceedings of the 1998 Winter Simulation Conference. Washington, 13.-16. Dezember 1998, S. 41-44.
- [Bal98b] Balci, O.: Verification, validation and testing. In: Banks, J. (Hrsg.): Handbook of simulation. John Wiley, New York, 1998, S. 335-393.
- [Bal03] Balci, O.: Verification, Validation, and Certification of Modeling and Simulation Applications. In: Chick, S.; Sánchez, P.J.; Ferrin, D.; Morrice, D.J. (Eds.): Proceedings of the 2003 Winter Simulation Conference. New Orleans, 7.-10. Dezember 2003, S. 150-158.
- [BBH+02] Berchtold, C.; Brade, D.; Hofmann, M.; Köster, A.; Krieger, T.; Lehmann, A.: Verifikation, Validierung und Akkreditierung von Modellen und Simulationen. Abschlussbericht zur Studie, Kennziffer M/GSPO/Z0076/Z9976, 2002.
- [BCK04] Burkschat, M; Cramer, E; Kamps, U.: Beschreibende Statistik - Grundlegende Methoden. Springer, Berlin, 2004.
- [BDW05] Bernhard, J.; Dragan, M.; Wenzel, S.: Evaluation und Erweiterung der Kriterien zur Klassifizierung von Visualisierungsverfahren für GNL. Technical Report – Sonderforschungsbereich 559 „Modellierung großer Netze in der Logistik“ 05001, 2005.

- [BHJ05] Bernhard, J.; Hömberg, K.; Jodin, D.: Standardprozesse als Grundlage für die Informationsbedarfsanalyse zur Modellierung von Großen Netzen der Logistik. In: Magdeburger Schriftenreihe zur Logistik – Logistikprozesse entwerfen, führen, bewerten. In: Wissenschaftliche Themenhefte des Lehrstuhls für Logistik der Universität Magdeburg, Heft 21, 2005, S. 3-14.
- [BJW03] Bernhard, J.; Jessen, U.; Wenzel, S.: A Taxonomy of Visualization Techniques for Simulation in Production and Logistics. In: Chick, S.; Sánchez, P. J.; Ferrin, D.; Morrice, D. J. (Eds.): Proceedings of the 2003 Winter Simulation Conference. New Orleans, 7.-10. Dezember 2003, S. 729-736.
- [BLe94] Barnett, V.; Lewis, T.: Outliers in Statistical Data. 3rd edition, John Wiley & Sons, Chichester, 1994.
- [BTr94] Büning, H.; Trenkler, G.: Nichtparametrische statistische Methoden. 2. Auflage, deGruyter, Berlin, 1994.
- [BWe04] Bernhard, J.; Wenzel, S.: Eine Taxonomie für Visualisierungsverfahren zur Anwendung in der Simulation in Produktion und Logistik. In: Schulze, T.; Schlechteweger, S.; Hinz, V. (Hrsg.): Simulation und Visualisierung 2004. Proceedings der Tagung "Simulation und Visualisierung 2004", SCS-European Publishing House, Erlangen San Diego, 2004, S. 85-96.
- [BWe05] Bernhard, J.; Wenzel, S.: Information Acquisition for Model-based Analysis of Large Logistics Networks. In: Proceedings of SCS-ESM 2005 – 19th European Simulation Multiconference. SCS-European Publishing House, Riga, Latvia, 1.-4. June 2005.
- [CRW+07] Collisi-Böhmer, S.; Rose, O.; Weiß, M.; Wenzel, S.: Qualitätskriterien für die Simulation in Produktion und Logistik – Planung und Durchführung von Simulationsstudien. Erscheint in: Springer, Berlin, 2007.
- [Dav96] Davidson, F.: Principles of Statistical Data Handling. Sage Publications, Thousand Oaks (CA), 1996.
- [DJo03] Dasu, T.; Johnson, T.: Exploratory Data Mining and Data Cleaning. Wiley, New York, 2003.
- [Epp06] Eppler, M. J.: Managing Information Quality. Springer, Berlin, 2006.
- [Fen02] Fender, T.: Statistical Methods for Information Management in Large Logistics Networks. Joint Statistical Meetings (JSM), New York, USA, 2002.
- [FKK05] Fender, T.; Krampe, A.; Kuhnt, S.: Kriterien für die Kategorisierung statistischer Methoden im Rahmen eines Methodennutzungsmodells zur Informationsgewinnung in GNL. Technical Report – Sonderforschungsbereich 559 „Modellierung großer Netze in der Logistik“ 05005, 2005.
- [GKP03] Gather, U., Kuhnt, S., Pawlitschko, J.: Concepts of Outlyingness for various data structures. In: Misra, J.C. (Eds.): Industrial Mathematics and Statistics, Narosa Publishing House, New Delhi, 2003.
- [HEK05] Hartung, J.; Elpelt, B.; Klösener, K.-H.: Statistik. 14. Aufl., Oldenbourg, München, 2005.
- [HHN+07] Hömberg, K.; Hustadt, J.; Nagel, L.; Jodin, D.; Kochsiek, J.; Riha, I.: Basisprozesse für die Modellierung in großen Netzen der Logistik. Technical Report – Sonderforschungsbereich 559 „Modellierung großer Netze in der Logistik“ 07004, 2007.
- [HJL04] Hömberg, K.; Jodin, D.; Leppin, M.: Methoden der Datenerhebung. Technical Report – Sonderforschungsbereich 559 „Modellierung großer Netze in der Logistik“ 04002, 2004.

- [HJM07] Holzmüller, H. H.; Jockisch, M.; Matijevic, M.: Herausforderung bei der Erhebung von Kundenanforderungen an intralogistische Systeme. In: Sonderforschungsbereich 696 „Forderungsgerechte Auslegung von intralogistischen Systemen“. Verlag Praxiswissen, Dortmund, 2007, S. 27-47.
- [JMa05] Jodin, D.; Mayer, A.: Automatisierte Methoden und Systeme der Datenerhebung. Technical Report – Sonderforschungsbereich 559 „Modellierung großer Netze in der Logistik“ 05004, 2005.
- [KGr95] Kosturiak, J.; Gregor, M.: Simulation von Produktionssystemen. Springer, Wien, 1995.
- [Klu97] Kluck, M.: Methoden der Informationsanalyse. Eine Einführung in die empirischen Methoden der Informationsbedarfsanalyse und der Markt- und Benutzerforschung. In: Buder, M.; Rehfeld, W.; Seeger, T.; Strauch, D. (Hrsg.): Grundlagen der praktischen Information und Dokumentation. Ein Handbuch zur Einführung in die fachliche Informationsarbeit. Saur, München, 1997, S. 795-821.
- [Kre93] Kreienbrock, L.: Einführung in Stichprobenverfahren. 2. Aufl., Oldenbourg, München, 1993.
- [Kro06] Kromrey, H.: Empirische Sozialforschung – Modelle und Methoden der standardisierten Datenerhebung und Datenauswertung. Bd. 1040: UTB für Wissenschaft, Uni-Taschenbücher. 11. Aufl., Lucius & Lucius Verlagsgesellschaft, Stuttgart, 2006.
- [KSW02] Kahn, B.; Strong, D.; Wang, R.: Information Quality Benchmarks. In: COMMUNICATIONS OF THE ACM. No. 4ve, 2002, Nr. 45, S. 184-192.
- [Kuc05] Kuckartz, U.: Einführung in die computergestützte Analyse qualitativer Daten. Lehrbuch VS Verlag für Sozialwiss., Wiesbaden, 2005.
- [NTa97] Nonaka, I.; Takeuchi, H.: Die Organisation des Wissens. Campus Verlag, Frankfurt New York, 1997.
- [Rhe01] Rabe, M.; Hellingrath, B.: Handlungsanleitung Simulation in Produktion und Logistik. SCS Verlag, San Diego, 2001.
- [Rom04] Romesburg, H. C.: Cluster Analysis for Researchers. Lulu Press, North Carolina, 2004.
- [RSW07] Rabe, M.; Spieckermann, S.; Wenzel, S.: Verifikation und Validierung für die Simulation in Produktion und Logistik – Vorgehensmodelle und Techniken. Erscheint in: Springer, Berlin, 2007.
- [Rud99] Rudolph, A.: Data Mining in Action – Statistische Verfahren der Klassifikation. Shaker Verlag, Aachen, 1999.
- [RWA85] Rossi, P. H.; Wright, J. P.; Anderson, A. B.: Handbook of Survey Research. Academic Press, New York, 1985.
- [Sch05] Schulte-Zurhausen, M.: Organisation. 4. Aufl., Verlag Vahlen, München, 2005.
- [TSW06] Toutenburg, H.; Schomaker, M.; Wissmann, M.: Arbeitsbuch zur deskriptiven und induktiven Statistik. Springer, Berlin, 2006.
- [VDI97] VDI 3633 Blatt 2: Lastenheft/Pflichtenheft und Leistungsbeschreibung für die Simulationsstudie. Verein Deutscher Ingenieure, Düsseldorf, Beuth Verlag, Berlin, 1997.
- [VDI00] VDI 3633 Blatt 1: Simulation von Logistik-, Materialfluss- und Produktionssystemen, Entwurf. Verein Deutscher Ingenieure, Düsseldorf, Beuth Verlag, Berlin, 2000.

- [VDI07] VDI 3633 Blatt 1: Simulation von Logistik-, Materialfluss- und Produktionssystemen. Verein Deutscher Ingenieure, Düsseldorf, Beuth Verlag, Berlin, 2007.
- [Voß04] Voß, W.: Taschenbuch der Statistik. Fachbuchverlag Leipzig im Carl Hanser Verlag, München, 2004.
- [WBe03] Wenzel, S.; Bernhard, J.: Vorgehensmodell zur Informationsgewinnung für die Modellierung von Logistiksystemen. In: Hohmann, R. (Hrsg.): Simulationstechnik. Tagungsband zum 17. Symposium in Magdeburg. Reihe: Frontiers in Simulation, FS 13, SCS-Europe BVBA, Ghent, 2003, S. 379-384.
- [Wey01] Weyerke, M.: Entwicklung eines rechnergestützten, benutzerorientierten Systems zur Informationsgewinnung in Produktionsunternehmen. Fortschritt-Berichte VDI Reihe 16, Nr. 134, VDI Verlag, Düsseldorf, 2001.
- [Zan70] Zangemeister, C.: Nutzwertanalyse in der Systemtechnik – Eine Methodik zur multidimensionalen Bewertung und Auswahl von Projektalternativen. 4. Auflage, Wittenmannsche Buchhandlung, München, 1976.

Sonderforschungsbereich 559

Bisher erschienene Technical Reports

- 05002 Bernd Hellingrath, Sana Mehicic-Eberhardt, Markus Witthaut: Entwicklung eines Anaylserahmens für die Untersuchung organisatorischer Aspekte in der Supply Chain
- 05003 Dennis Müller, Mathias Stöber, Axel Thümmler: Einsatz der Response Surface Methode zur Optimierung komplexer Simulationsmodelle
- 05004 Dirk Jodin, Andreas Mayer: Automatisierte Methoden und Systeme der Datenerhebung
- 05005 Thomas Fender, Anne Krampe, Sonja Kuhnt: Kriterien für die Kategorisierung statistischer Methoden im Rahmen eines Methodennutzungsmodells zur Informationsgewinnung in GNL
- 05006 Kay Hömberg, Dirk Jodin, Maik Langenbach, Christian Kellner: Konzept einer logistischen Informationsbedarfsanalyse mit Hilfe von Basisprozessen und standardisierten Logistikdaten
- 05007 Hans-Werner Graf: Festlegung der Abfahrts- und Ankunftszeiten (Fahrplangestaltung)
- 06001 Iwo Riha: Grundlagen des Cost-Benefit-Sharing
- 06002 Jens Finzel, Michael Hierweck, Andreas van Almsick, Jan Sören Kriege, Mathias Schwenke: ProC/B-Editor – Handbuch
- 06003 Mirko Eickhoff, Michael Hierweck, Mathias Schwenke: Hands On ProC/B-Tools – Eine beispielorientierte Einführung in die Anwendung der ProC/B-Tools
- 06004 Doris Blutner, Stephan Cramer, Tobias Haertel: Der Mensch in der Logistik: Planer, Operateur und Problemlöser
- 06005 Tobias Haertel: UsersAward: Ein Beitrag zur optimalen Gestaltung von Mensch-Maschine-Systemen in der Logistik
- 06006 Falko Bause, Tim Geißen, Anne Meinke, Veye Tatah, Marcus Völker: Performance Evaluation for Cost Calculation of Business Processes
- 06007 Peter Kemper, Carsten Tepper: Trace Analysis – Gain Insight through Modelchecking and Cycle Reduction
- 06008 Jochen Bernhard, Dirk Jodin, Kay Hömberg, Sonja Kuhnt, Christoph Schürmann, Sigrid Wenzel: Vorgehensmodell zur Informationsgewinnung – Prozessschritte und Methodennutzung -

Alle Technical Reports können im Internet unter
<http://www.sfb559.uni-dortmund.de/>
abgerufen werden. Für eine Druckversion wenden Sie
sich bitte an die SFB-Geschäftsstelle
e-mail: andrea.grossecappenberg@iml.fraunhofer.de