

# Big Data in der Logistik

Ein ganzheitlicher Ansatz für die datengetriebene Logistikplanung, -überwachung und -steuerung

Norman Spangenberg, Martin Roth, Stefan Mutke und Bogdan Franczyk, Universität Leipzig

Die Bedeutung der Logistik hat sich in den vergangenen Jahrzehnten stark verändert. Während diese früher zu den Kernfunktionen der meisten Unternehmen zählte, werden Logistikdienstleistungen heutzutage häufig an Logistikdienstleister ausgelagert. Diese Verlagerung führt zu neuen organisatorischen Strukturen und ermöglicht ebenso die Umsetzung neuer innovativer Geschäftsmodelle. Durch die Digitalisierung der Logistik steigt der notwendige Integrations- und Koordinationsaufwand exponentiell an und kann nur durch den intelligenten Einsatz von IT beherrschbar gestaltet werden kann. Dieser Beitrag beleuchtet das Spannungsfeld der Logistik und IT und zeigt auf, welchen Herausforderungen die Logistik unterliegt und wie diese durch den adäquaten Einsatz von Big Data Technologien bewältigt werden können.

Die Logistik sichert als Querschnittsfunktion die Versorgung der Haushalte sowie Unternehmen mit logistischen Gütern ab und stellt für Unternehmen zunehmend einen entscheidenden Wirtschafts- und Wettbewerbsfaktor dar. Dabei unterliegt die Logistik ständigen Veränderungen, sowohl aus organisatorischer als auch technischer Sicht. Insbesondere die zunehmende Digitalisierung führt zu neuen Herausforderungen, bietet aber auch Potenziale für heutige als auch zukünftige Geschäftsmodelle.

Die mit der Digitalisierung einhergehende Steigerung der Datenmengen, welche immer schneller und aus einer Vielzahl an Quellen entspringen (Big Data), führt zu einem Paradigmenwechsel innerhalb der Logistik. Big Data bezeichnet dabei Technologien, Methoden und Algorithmen, die Unternehmen in die Lage versetzen, ihre Daten zu beherrschen und nutzbar zu machen, um Effizienz und Wettbewerbsfähigkeit zu steigern [1]. Big Data ermöglicht zunächst die anfallenden Datenmengen aus verschiedensten Systemen, Prozessen und Sensoren aufzunehmen, zu verarbeiten und zu speichern. Insbesondere bieten Big Data-Konzepte die Möglichkeit, über Analysen zusätzliche Informationen zu generieren, die in Planungs- und Entscheidungsprozessen verwendet werden, um Mehrwerte in der Logistik zu generieren.

Entscheidend für das Gelingen von Big Data in der Logistik ist das Zusammenspiel von Logistik- und Entscheidungsprozessen, Big Data-Methoden und Daten. Nur durch die Übertragung von Big Data-Methoden können die komplexer werdenden Prozesse in der Logistik und angrenzenden Gebieten handhabbar gestaltet und verwaltet werden. Dabei spielen auf operativer Ebene Entscheidungen in Echtzeit eine immer wichtigere Rolle, um in Logistikprozesse steuernd eingreifen zu können, um so die Resilienz dieser zu erhöhen.

Ziel des Beitrags ist es, aufzuzeigen, wie aktuelle Logistiksysteme von datengetriebenen Verfahren profitieren können und welche Rolle diese bei der Etablierung zukünftiger Geschäftsmodelle besitzen. Weiterhin werden Querverweise zwischen dem Logistiklebenszyklus und geeigneten Big Data-Tools aufgezeigt und erklärt, welchen Einfluss diese auf Logistikprozesse besitzen.

Herausforderungen und Potenziale datengetriebener Logistik

Wettbewerbsdruck, unvorhersehbare Märkte sowie dynamische Änderungen von Regularien beeinflussen maßgeblich die Prozessgestaltung und -durchführung von Unternehmen.

## Big Data in Logistics – A Holistic Approach for Data-Driven Logistics Planning, Monitoring and Management

Over the last years, the importance of logistics has changed significantly. While logistics used to be a core function of most companies, logistics services nowadays are often outsourced to service providers. This leads to new organizational structures and enables innovative business models. Caused by the digitalization of logistics, efforts for integration and coordination rise and can only be made controllable by intelligent use of IT. This contribution examines the field of tension of logistics and IT. It shows which challenges to face and how to overcome these by using Big Data technologies.

### Keywords:

logistics lifecycle, big data, logistics planning, logistics monitoring

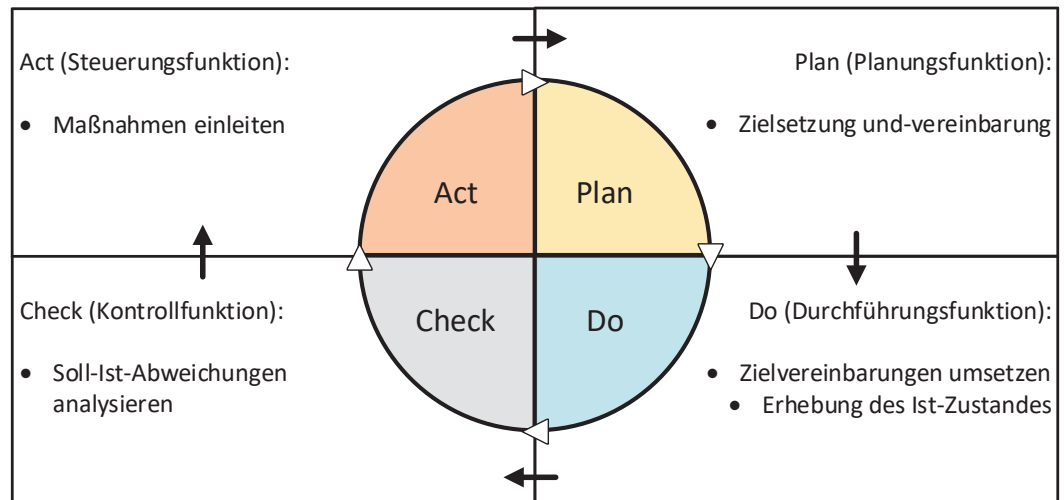
Norman Spangenberg arbeitet als wissenschaftlicher Mitarbeiter am Lehrstuhl für Informationsmanagement der Universität Leipzig sowie im Big Data-Kompetenzzentrum ScaDS Dresden/Leipzig.

Martin Roth arbeitet als wissenschaftlicher Mitarbeiter am Lehrstuhl für Informationsmanagement der Universität Leipzig. Er leitet die Forschungsgruppe „Funkbasierte Anwendungen“.

Stefan Mutke arbeitet als wissenschaftlicher Mitarbeiter am Lehrstuhl für Informationssysteme in der Logistik der Universität Leipzig. Er leitet die Forschungsgruppe „Logistik“.

Prof. Dr.-Ing. Bogdan Franczyk ist Inhaber der Lehrstühle Informationsmanagement am Institut für Wirtschaftsinformatik der Universität Leipzig und Wirtschaftsinformatik der University of Economics in Wrocław.

franczyk@wifa.uni-leipzig.de  
www.wifa.uni-leipzig.de/  
iwi/im.html



**Bild 1: Deming-Kreis mit Logistikfunktionen.**

Um diesen Herausforderungen entgegenzutreten, lagern viele Unternehmen interne Aktivitäten an externe Dienstleister aus, um die Effizienz zu erhöhen. Diese Entwicklung ist auch in der Logistik zu beobachten und führt zu neuen Geschäftsmodellen wie dem 4th Party Logistics (4PL) Provider. Im Vergleich zum 3rd Party Logistics Provider (3PL) gestaltet, plant, überwacht und steuert dieser logistische Dienstleistungen, welche durch eine Vielzahl an externen Logistikdienstleistern (LDL) gemeinsam erbracht werden, ohne dabei über eigene physische Logistikressourcen zu verfügen [2]. Somit bedient sich der 4PL an mehreren externen Dienstleistern, welche gemeinsam diesen Ablauf realisieren, wodurch die Gesamtzielgrößen auf die einzelnen Dienstleister heruntergebrochen werden müssen. Aufgrund des zunehmenden Netzwerkcharakters steigt die Komplexität enorm an, wodurch neue Lösungsansätze für unterstützende IT-Systeme erforderlich werden. Dies wird noch verschärft, da nicht nur klassische Logistikfunktionen ausgelagert werden, sondern auch Mehrwertdienstleistungen wie Verzollung. Dadurch müssen zum einen die Besonderheiten und Vielzahl an externen Dienstleistern in der Prozessplanung berücksichtigt werden. Zum anderen müssen Überwachungsdaten der Prozessdurchführung schnellstmöglich zur Verfügung stehen, damit bei Teilprozessabweichungen steuernd eingegriffen werden kann. Eine weitere Möglichkeit, dem Konkurrenzdruck entgegenzuwirken, ist Synchronmodalität [3]. Ziel dabei ist, den Transportablauf zu flexibilisieren, indem die Informations- und Warenflüsse miteinander verknüpft werden, wodurch sich Lieferketten während des Transports flexibel gemäß den Rahmenbedingungen anpassen lassen. Der synchronmodale Transport minimiert Pufferzeiten, unterstützt die Bündelung von Waren und

Gütern und ermöglicht die schnelle Anpassung der Transportmodi [4]. Um diese Flexibilität zu erreichen, müssen prozessbegleitende Daten in Echtzeit zur Verfügung stehen.

Auch die Umsetzung des Sharing Economy-Prinzips eröffnet der Logistik neue Potenziale [5]. So ist es möglich, dass Spediteure ihre Ladungsflächen teilen, um dadurch Lagerkosten, Umwege oder weitere Ressourcen zu sparen. Durch die Umsetzung des Sharing-Gedanken in der Logistik wird die intelligente Vernetzung aller logistischen Prozesse durch die neuen Aufgaben von Industrie 4.0 vorangetrieben. Um dies zu erreichen, ist es notwendig, dass die Erfassung, Aufbereitung und Verfügbarkeit von unternehmensübergreifenden Daten rollenspezifisch und anwendergerecht echtzeitnah durchgeführt wird.

Die zentrale Herausforderung für heutige, aber auch zukünftige Logistiksysteme besteht in der intelligenten Nutzung von Daten, damit ein Höchstmaß an Datentransparenz und -verfügbarkeit erreicht wird. Dies bedingt auch die phasenübergreifende Nutzung und Anreicherung von Daten durch Big Data Technologien und deren Zusammenwirken in verschiedenen Planungs- und Steuerungsprozessen. Dabei muss insbesondere die Datensicherheit in allen teilnehmenden Systemen umfassend berücksichtigt werden [6]. Insgesamt betrachtet stellt die Logistik aufgrund ihrer Querschnittsfunktion einen kritischen Erfolgsfaktor für die erfolgreiche Umsetzung von Industrie 4.0 dar [7].

Logistiklebenszyklus im LSEM-Ansatz

Basierend auf dem Deming-Zyklus [8] werden nachfolgend die typischen Funktionen innerhalb eines Lebenszyklus logistischer Dienst-

leistungen erläutert (Bild 1). Der nachfolgend beschriebene „Logistik Service Engineering und Management“ (LSEM)-Ansatz zeigt auf, wie Daten phasenübergreifend erhoben und für Entscheidungs- und Planungsprozesse genutzt werden [9, 10].

Beginnend mit der Planungsfunktion werden die Ziele einer Dienstleistungsdurchführung zwischen Kunde und Logistikdienstleister vereinbart. Dies beinhaltet die Spezifikation von zeitlichen und zustandsbezogenen Aspekten. Wird dies im Sinne des 4PL durchgeführt, teilt der Koordinator die Zielgrößen der Gesamtdienstleistung auf die Aktivitäten der teilnehmenden externen Dienstleister auf und vereinbart mit diesen ebenso Ziele und hält diese in Form von Verträgen fest.

Gegenstand der Durchführungsfunktion ist zum einen die Umsetzung der zuvor definierten Ziele und zum anderen die Erhebung des Ist-Zustands. Letzteres hat zum Ziel, die Zustände des real laufenden Prozesses zu erheben und virtuell abzubilden.

Basierend auf diesen Daten wird anschließend die Kontrollfunktion durchgeführt. Es wird überprüft, ob die vereinbarten Ziele (Soll-Zustand) durch die Prozessdurchführung (Ist-Zustand) erreicht wurden. Sollten Abweichungen aufgetreten sein, werden die spezifischen Daten an die nächste Phase weitergeleitet.

Innerhalb der Steuerungsfunktion werden die zuvor generierten Analysedaten genutzt, um ggf. Maßnahmen für die jeweilige Instanz, zumindest jedoch für die nächste Durchführung, einzuleiten.

Aufbauend auf dieser Beschreibung wird nachfolgend der „Logistik Service Engineering und Management“ (LSEM)-Ansatz beschrieben und aufgezeigt, wie Daten phasenübergreifend erhoben und für Entscheidungs- und Planungsprozesse genutzt werden können (Bild 2).

Im LSEM-Ansatz erfolgt die Planung logistischer Dienstleistungen durch die Nutzung verschiedener Modelle wie Prozess-, Simulationsmodelle oder Profile, um frühzeitig ein Höchstmaß an Planungssicherheit zu erreichen. Die Simulation als zentraler Bestandteil der Planung nutzt hierfür Instanzdaten (LDL-Profil) bereits abgeschlossener Prozesse sowie externe Datenquellen. Korrekte und robuste Daten sowie Informationen stellen daher einen wichtigen Erfolgsfaktor innerhalb der Simulation dar. Sobald die Simulation erfolgreich durchlaufen wurde, wird das daraus entstehende gültige Modell instanziiert.

Während der Durchführung einer logistischen Dienstleistung erfolgt die kontinuierliche Erhebung des Ist-Zustandes. Dieser ständig anfallende Datenstrom wird den zuvor definierten Zuständigkeiten der Logistikdienstleister (LDL)

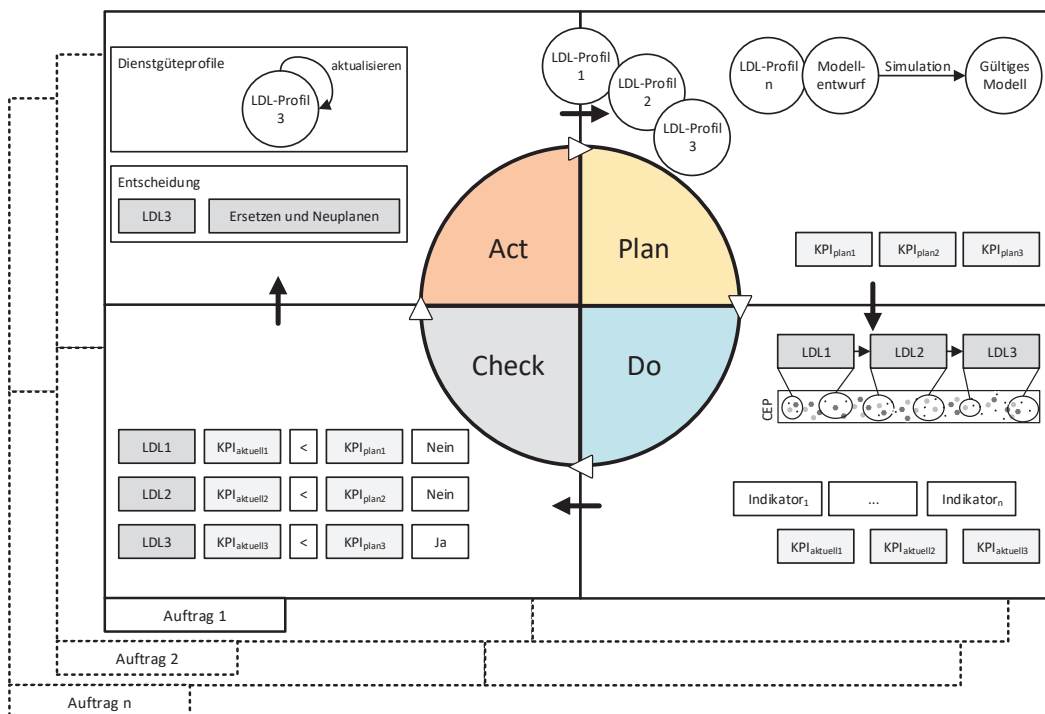


Bild 2: Ablauf des „Logistik Service Engineering und Management“-Ansatzes.

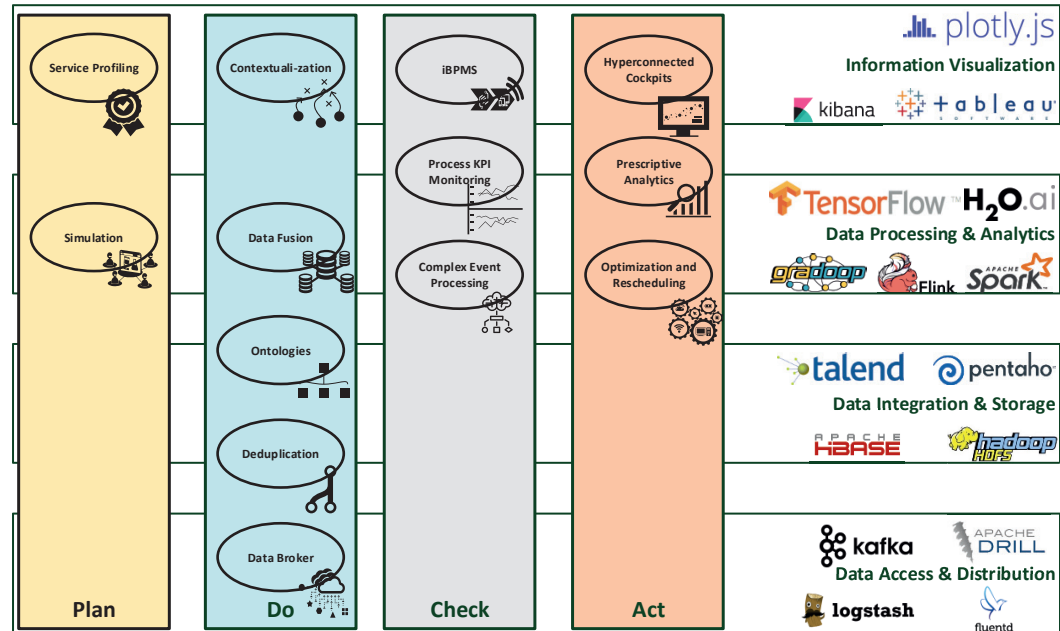


Bild 3: Anwendung von Big Data im Logistiklebenszyklus.

zugeordnet. Dazu werden Methoden des Complex Event Processing (CEP) genutzt, um Ereignisse, während sie generiert werden, zu verarbeiten und Beziehungen zwischen diesen zu untersuchen bzw. prozessinterne mit externen Daten anzureichern [11].

Diese Datenbasis wird in der Check-Phase genutzt, um einen kontinuierlichen Abgleich zwischen Soll- und Ist-Zustand durchzuführen. Dabei werden diverse Cloud-Services eingebunden, um zu überprüfen, ob geplante Prozessschritte vertragskonform durchgeführt werden. Weiterhin werden statistische Verfahren genutzt, um zustandsbezogene Daten zu analysieren und Prognosen über die Einhaltung der Schwellwerte zu berechnen. Aufgrund der Echtzeitfähigkeit wird in dieser Phase u. a. ermittelt, dass LDL 3 nicht mehr sein Prozessziel erreichen kann (Bild 2). Dem Disponenten bleibt dadurch genug Zeit, steuernd in den Prozess einzugreifen, um die Einhaltung der Gesamtdienstleistung sicherzustellen.

In der abschließenden Phase Act werden entsprechende Maßnahmen eingeleitet, wie bspw. der Austausch von LDL 3 und die Aktualisierung dessen Dienstleisterprofils. Diese feingranulare Datenbasis steht zeitnah zur Verfügung und wird bei der Planung neuer Aufträge genutzt. Schwachstellen bzw. Optimierungspotenziale können durch die Nutzung von Data-Mining-Methoden und Vorgehen, wie bspw. CRISP-DM oder ASUM-DM [12,13], aufgedeckt werden. Dadurch arbeitet die Planung auf einer aktuellen und zuverlässigen

Datenbasis, wodurch belastbarere Aussagen möglich werden.

### Big Data-Technologien für datengetriebene Logistik

In diesem Abschnitt wird die konzeptionelle Anreicherung und Integration des Logistiklebenszyklus mit Methoden und Technologien aus den Bereichen Big Data und Datenanalyse vorgestellt. Dabei lassen sich folgende Ebenen unterscheiden:

- Data Access & Distribution (Zugriff auf Datenquellen und deren Verteilung)
- Data Integration & Storage (Einbinden der Datenquellen, Vorverarbeitung und Speicherung)
- Data Processing & Analytics (Verarbeitung und Durchführung von Analysen)
- Information Visualization (Darstellung der gewonnenen Informationen)

Insbesondere zwei Big Data-Entwicklungen sichern die Planung von logistischen Dienstleistungen umfassend ab. Zum einen sind dies neue Datenquellen, die zu detaillierteren Prozess- und Simulationsmodellen führen können [14]. Durch die Anreicherung der Planungsdaten mit bspw. Geo- und Fahrplandaten, aber auch durch Fortschritte im Enterprise Data Warehousing lassen sich die Planungsmodelle erweitern. Ferner lässt sich die bisher anzureichende manuelle Erstellung und Pflege der Dienstleistungsprofile auf Basis interner Daten durch den Einbezug externer Datenanbieter sowie der Nutzung von Klassifikationsverfahren

#### Literatur

- [1] Gartner Research: IT glossary – „Big Data“. URL: <http://www.gartner.com/it-glossary/big-data>, Abrufdatum 01.06.2017.
- [2] Hausladen, I.: IT-gestützte Logistik: Systeme-Prozesse-Anwendungen. Wiesbaden 2014.
- [3] Tavasszy, L. A.; Behdani, B.; Konings, R.: Intermodality and synchronomodality. 2015.
- [4] van Stijn, E.; Hesketh, D.; Tan, Y.-H.; Klievink, B.; Overbeek, S.; Heijmann, F.; Pikart, M.; Butterly, T.: The Data Pipeline. In: Proceedings of the Global Trade Facilitation Conference 2011. New York 2013.
- [5] Gesing, B.: Sharing Economy Logistics - Rethinking logistics with access over ownership, DHL Customer Solutions & Innovation, 2017.
- [6] Bousonville, T.: Logistik 4.0: Die digitale Transformation der Wertschöpfungskette. Wiesbaden 2016.
- [7] Zillmann, M.: Keine Industrie 4.0 ohne Digitalisierung der Supply Chain Intelligente Logistikdienstleistungen für die Fertigungsindustrie. Mindelheim, 2016.
- [8] Deming, W. E.: Out of the Crisis. Cambridge 2000.



ren vereinfachen und zu erheblichen Teilen automatisieren. Zum anderen können skalierbare und verteilte Simulationsarchitekturen eingesetzt werden, um dem immensen Bedarf an Rechenleistung gerecht zu werden. Dadurch sind komplexere Modelle der Simulation möglich, die zusätzliche Granularitäten, Zielgrößen und Rahmenbedingungen erlauben, bspw. die Simulation mehrerer Prozessvarianten.

In der Do-Phase besteht das Hauptziel in der Datenbeschaffung. Um zukünftige Logistikkonzepte wie Synchromodalität umzusetzen, müssen Transportmittel, aber auch Transportgüter selbst aktuelle und detaillierte Informationen liefern. Für die Zustandserfassung lassen sich Sensoren sowie das Internet of Things für deren Vernetzung nutzen [15]. Weiter ist auch die Datenverteilung innerhalb der Do-Phase durch Big Data zu unterstützen. Für die Datenquellenanbindung hat sich als Standard-Framework Apache Kafka herausgebildet. Für die Vorverarbeitung der Daten sind Talend und Pentaho führende Anbieter für ETL-Prozesse wie Data Cleaning, Deduplication und Matching. Weiter werden diese gemeinsam mit semantischen Methoden wie Kontextualisierung, Ontologien und Data Fusion verwendet, um dynamisch Informationen in der Logistikplanung zu integrieren.

Aufgrund der großen Datenmenge und Komplexität bieten sich für das Überwachen (Check) der Logistiknetzwerke neue Möglichkeiten, um die Informationsmenge zu reduzieren und für Logistik-Manager überschaubar zu machen. Data-Processing-Frameworks wie Apache Spark [16] und Apache Flink [17] erlauben die Aggregation der Daten zu aussagekräftigen Informationen. Darüber hinaus lassen sich durch Zustandsdaten und Methoden zur Mustererkennung relevante Ereignisse erkennen. Damit lassen sich intelligente Geschäftsprozessmanagement-Lösungen (iBPMS) implementieren, die analytische Methoden wie Predictive Analytics einbeziehen [18].

Eine enge Verknüpfung zwischen diesen Informationen und Analysen besteht zu den Aufgaben der Act-Phase. Durch Big Data-Methoden und Decision Support System-Komponenten lassen sich operative Entscheidungsprozesse optimieren. Weiterhin wird an Konzepten wie Hyperconnected Cockpits geforscht [19]. Diese sollen Entscheidungs- und Kollaborationsprozesse vereinfachen, indem Informationen und Analysen integriert werden, um so zu Operational Intelligence zu führen. Dazu werden Echtzeitdaten, aggregierte Informationen und Ereignisse mit Entscheidungsmodellen und Er-

fahrungswissen der Manager kombiniert. Als Ergebnis entstehen Handlungsempfehlungen, bspw. für die Umplanung laufender Logistikprozesse. Durch Prescriptive Analytics lassen sich diese Empfehlungen bewerten, indem die konkreten Auswirkungen der Alternativen prognostiziert werden. Für die Implementierung dieser Konzepte sind skalierbare Systeme für large-scale testing und Optimierungsrechnungen notwendig.

Bild 3 zeigt zusammenfassend die Anwendung von Big Data im Logistiklebenszyklus.

### Zusammenfassung

Dieser Beitrag beleuchtete neue Stoßrichtungen der Logistik und zeigte neben den Potenzialen dieser Trends auch deren Herausforderungen auf. Die steigende Bedeutung der adäquaten Verwendung von Daten wurde in den nachfolgenden Kapiteln aufgegriffen und auf Basis dessen der LSEM-Ansatz vorgestellt, welcher die gewonnenen Daten aus dem Lebenszyklus nutzt und einen Rahmen für die Datenverwendung definiert. Die funktionale Ausführung wird durch die Vorstellung geeigneter Werkzeuge und Methoden in den unterschiedlichen Schichten einer typischen Big Data-Architektur angereichert. Das vorgestellte Konzept stellt einen ganzheitlichen Ansatz dar und kann Logistikunternehmen als Basis zur Umsetzung einer datengetriebenen Logistik dienen. Hierdurch kann die notwendige Datentransparenz und -verfügbarkeit erhöht und die daraus resultierende Prozesstransparenz zielgrößenoptimierend für bestehende Logistiksysteme genutzt werden. Der diskutierte Ansatz wird derzeit in der Praxis untersucht, wodurch sich Anpassungen ergeben könnten. Bspw. wurden die Aspekte der Datensicherheit bisher noch nicht ausreichend betrachtet.

### Schlüsselwörter:

Logistiklebenszyklus, Big Data, Logistikplanung, Logistiküberwachung

*Der Beitrag basiert auf praktischen Erfahrungen aus den Projekten „ScaDS“ und „LSEM“. Diese sind durch das Bildungsministerium für Bildung und Forschung gefördert (ScaDS - 01IS14014B; LSEM - 03IPT504X). Das Big Data-Kompetenzzentrum ScaDS Dresden/Leipzig forscht u. a. in den Bereichen Graph-Datenverarbeitung und High-Performance Computing sowie im Bereich Business Data (Lehrstuhl für Informationsmanagement der Universität Leipzig).*

- [9] Mutke, S.; Roth, M.; Ludwig, A.; Franczyk, B.: Towards Real-time Data Acquisition for Simulation of Logistics Service Systems. In: International Conference on Computational Logistics (2013), S. 242-256.
- [10] Roth, M.; Mutke, S.; Klarmann, A.; Franczyk, B.; Ludwig, A.: Continuous Quality Improvement in Logistics Service Provisioning. In: 17th International Conference on Business Information Systems (2014), S. 253-264.
- [11] Bruns, R.; Dunkel, J.: Complex Event Processing: Komplexe Analyse von massiven Datenströmen mit CEP. Wiesbaden 2015
- [12] Kalgotra, P.; Sharda, R.: Progression analysis of signals: Extending CRISP-DM to stream analytics. In: IEEE International Conference on Big Data (2016), S. 2880-2885.
- [13] Tuovinen, L.: A Conceptual Model of Actors and Interactions for the Knowledge Discovery Process. 2016.
- [14] Talend: Big Data & Logistics: 7 Current Trends to Watch. URL: [www.talend.com/blog/2016/02/18/big-data-logistics-7-current-trends-to-watch/](http://www.talend.com/blog/2016/02/18/big-data-logistics-7-current-trends-to-watch/), Abrufdatum 01.06.2017.
- [15] Lu, M.; Borbon-Galvez, Y.: Advanced logistics and supply chain management for intelligent and sustainable transport. Proceedings of the 19th ITS World Congress (2012), S. 1-9.
- [16] Apache Foundation: Apache Spark. URL: <http://spark.apache.org/>, Abrufdatum 01.06.2017.
- [17] Apache Foundation: Apache Flink. URL: <http://flink.apache.org/>, Abrufdatum 01.06.2017.
- [18] Grambow, G.; Oberhauser, R.; Reichert, M.: On the Fundamentals of Intelligent Process-Aware Information Systems. In: Grambow, G.; Oberhauser, R.; Reichert, M. (Hrsg): Advances in Intelligent Process-Aware Information Systems. Cham 2017, S. 1-13.
- [19] Bearingpoint: Connected Supply Chain - Cockpit. URL: [www.bearingpoint.com/de-at/unsere-expertise/branchen/automotive/-connected-supply-chain-cockpit/](http://www.bearingpoint.com/de-at/unsere-expertise/branchen/automotive/-connected-supply-chain-cockpit/), Abrufdatum 01.06.2017.