

# Industrial Big Data: Tieferes Prozessverständnis durch Daten

Modernes Informationsmanagement in der Produktion

Thomas Thiele, Max Hoffmann und Tobias Meisen,  
Lehrstuhl für Informationsmanagement im Maschinenbau, RWTH Aachen

Längst hat die digitale Transformation auch die industrielle Produktion erreicht. Insbesondere getrieben durch die Erfolge und Fortschritte der vorwiegend amerikanischen IT-Unternehmen ist diese vierte industrielle Revolution, oder Industrie 4.0, wie sie in der Hightech-Strategie der Bundesregierung heißt, in vollem Gange. Eine Vielzahl von Begriffen und Technologien wie Internet der Dinge, Cyber-Physische Systeme und Big Data bilden das Rahmenwerk dieses Wandels. Im vorliegenden Beitrag wird insbesondere der Big Data Begriff adressiert und eine Abgrenzung dieses umfassenden Konzepts hinsichtlich der Definition von Industrial Big Data unternommen. Zudem wird anhand von Fallbeispielen exemplarisch erläutert, welche Mehrwerte die Anwendung dieser Methoden in der Produktion generiert.

## Datenintegration als Schlüsselfaktor

Der Begriff der digitalen Transformation beschreibt weit mehr als die Mechanisierung, Automatisierung und Robotisierung unserer Arbeitswelten. Vielmehr fasst er die durch eine fortschreitende digitale Vernetzung induzierte Veränderung unserer Gesellschaft zusammen. Dabei durchdringt die Digitalisierung zunehmend das tägliche Leben und nimmt Einfluss auf zentrale Lebensbereiche (vgl. hierzu beispielsweise Statistiken in [1] bzgl. der stetig steigenden Nutzung digitaler Medien und Techniken). In der betriebswirtschaftlichen Praxis von Unternehmen impliziert dies eine Auseinandersetzung mit modernem Marktverhalten, Geschäftsmodellen sowie bewährten Lebens- und Arbeitswelten.

Ein wesentlicher Enabler dieses Wandels sind die Möglichkeiten, die uns Big Data eröffnen. Big Data ist gleichzeitig ein Begriff und eine Sammlung von Methoden: Einerseits beschreibt der Ausdruck die stetige Zunahme des Datenvolumens, der -geschwindigkeit und der -vielfältigkeit. Andererseits steht er für die Befähigung technischer Systeme zur Speicherung, Nutzung und Analyse großer Datenmengen – sogenannter Big Data Technologien. [2] Es ist daher wenig verwunderlich, dass die Übertragung der Technologien und Vorgehensmodelle in den industriellen Kontext durch immer mehr produzierende Unternehmen angestrebt wird. Die

produzierende Industrie generiert bereits heute, befähigt durch die vorangegangenen Digitalisierungsbemühungen, große Mengen digitalisierter Daten [3]. Ihre Erschließung und Nutzbarmachung verspricht vielfältige Möglichkeiten hinsichtlich der Optimierung und des Verständnisgewinns zur Gestaltung von Arbeits- und Geschäftsprozessen. Peter Sondergaard, Senior Vice President bei Gartner, bezeichnete den Wert von Informationen treffend mit „Information is the oil of the 21<sup>st</sup> century, and analytics is the combustion engine“ [4].

Diese Übertragung von Big Data auf die industrielle Anwendung – sogenanntes Industrial Big Data – und die Erschließung der Potenziale ist daher für viele Unternehmen eines der strategischen Ziele. Industrial Big Data beschreibt das zugrundeliegende Konzept und die Methoden zum Umgang mit Daten, die in einer fortgeschritten vernetzten, digitalisierten Industrie entstehen [3]. Dabei existieren grundlegende Unterschiede zu Big Data. Industrial Big Data sind in der Regel strukturierter, stark korreliert und zeitabhängig sowie insgesamt ähnlicher in ihrer semantischen Repräsentation. Ein wesentliches Element zum Verständnis dieser Produktionsdaten für Analysten liegt in der Integration der Daten, die sich ohne domänenspezifisches Fachwissen nur schwer realisieren lässt. So haben Sensor- und Prozessdaten

## Industrial Big Data: Data-Driven Process Understanding – Modern Information Management in Production

The digital transformation led to disruptive changes in business models of leading companies. Big Data serves as one of the key enablers in this area. The transfer of this concept in the production domain towards an Industrial Big Data is key challenge for producing companies. Although exemplary key projects exist, no available characterization of structural elements in Industrial Big Data Processes exists. Therefore, this article aims at presenting initial structural elements of Industrial Big Data projects based on exemplary use cases.

### Keywords:

industrial big data, data integration, artificial intelligence, machine learning



Dipl.-Ing. Thomas Thiele leitet die Forschungsgruppe Produktionstechnik am Lehrstuhl für Informationsmanagement im Maschinenbau (IMA) an der RWTH Aachen University.

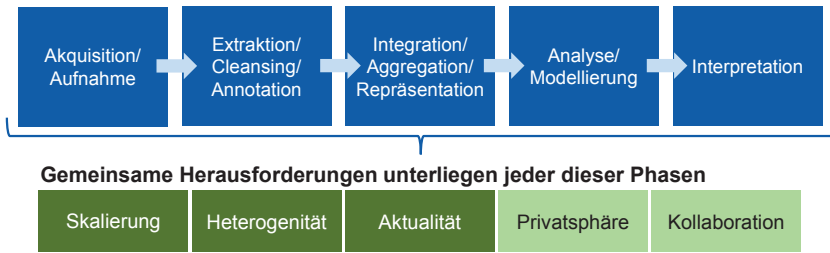


Dipl.-Ing. Max Hoffmann M.B.A. leitet die Forschungsgruppe Industrial Big Data am IMA an der RWTH Aachen University.



Prof. Dr.-Ing. Tobias Meisen ist Geschäftsführer des IMA und Inhaber einer Juniorprofessur zur Interoperabilität von Simulation an der RWTH Aachen University.

thomas.thiele@ima-zlw-ifu.rwth-aachen.de  
www.ima-zlw-ifu.rwth-aachen.de



**Bild 1: Analyseverfahren im Kontext von (Industrial) Big Data.**

eine eindeutige physikalische Bedeutung und ihnen liegt ein deterministischer Interpretationsraum zugrunde, dessen Analyse fachliches Domänenwissen voraussetzt. Es ist somit nicht hinreichend, die Technologien rund um Big Data lediglich anzuwenden, vielmehr ist eine tiefergehende Auseinandersetzung mit selbigen im industriellen Kontext notwendig. Jede Phase des Auswertungsprozesses (Bild 1) bringt hierbei spezifische Herausforderungen mit sich und wird um weitere, gemeinsame Herausforderungen nach beispielsweise Privatsphäre und Skalierung ergänzt.

Zur erfolgreichen Implementierung dieses Prozesses sieht sich die Industrie insbesondere einer zentralen Herausforderung gegenüber. In den letzten 30 Jahren lag der Fokus der Digitalisierung und hiermit verbundener Investitionen in die Automatisierungstechnik auf einer Optimierung spezifischer Kernprozesse. Im Ergebnis dieser Unternehmenspolitik führte dies zur Ausprägung von Insellösungen einzelner Teilprozesse und damit zu verstreuten Datenbeständen. Deren Akquisition, Zusammenführung und Homogenisierung ist daher eine erste Hürde, die es auf dem Weg zu Industrial Big Data zu nehmen gilt. Bisher wurden hierfür sogenannte Business Intelligence Ansätze, die auf Data Warehouses und Extract-Transform-Load (ETL) Integrationsprozessen zur Bereinigung, Aufbereitung und strukturierter Speicherung aufbauen, herangezogen. Für deren Einführung, Inbetriebnahme und Nutzung in den letzten zwei bis drei Jahrzehnten wurden mehrere Millionen Euro investiert [5]. Diese Ansätze setzen aber voraus, dass die Datennutzung, also der Zweck des Einsatzes, bereits vor der Zusammenführung bekannt ist (Schema-on-Write). Andernfalls ist die Definition der Zielstruktur (des Schemas) und des Integrationsprozesses nicht durchführbar. Für moderne Anwendungen ist dieses Vorgehen jedoch häufig nicht praktikabel, da die genaue Struktur der zu integrierenden Daten aufgrund ihrer Komplexität und Heterogenität im Vorhinein nur schwer abzuschätzen ist. Durch eine Vielzahl neuer Datenquellen und Analyseanfragen, die sich erst nach einer initialen Analyse des Betriebs ergeben, unterliegen die benötigten Schemata sowie die erforderliche flexible Zielstruktur ständig dynamischen Anpassungen. Da die Struktur von klas-

sischen Data Warehouses dieser Dynamik nicht gerecht wird, wurden sogenannte Data Lake Konzepte entwickelt, die flexible Zielstrukturen für sämtliche anfallende Prozessdaten bereitstellen. Anstelle von Schema-on-Write setzt der Data Lake auf das sogenannte Schema-on-Read [6]. Unter dieser Methodik werden Strukturen zusammengefasst, die es ermöglichen, Daten in ihrer Rohform abzulegen und das Schema erst bei der Extraktion gemäß ihres Anwendungszwecks zu definieren, indem eine entsprechende Aufbereitung und Zusammenführung heterogener Daten aus verschiedensten Quellen stattfindet. Um dieses Konzept zu ermöglichen, bedarf es jedoch weitaus mehr als der bloßen Idee, sämtliche Daten in einen Data Lake zu speichern. Wie Gartner in [7] aufzeigte, sind Daten ohne ihren Kontext wertlos, da eine entsprechende Zusammenführung eben erst durch den Kontext bewerkstelligt werden kann. Dieser Umstand führt dazu, dass entweder sämtliche Nutzer der verfügbaren Daten über die entsprechenden Kenntnisse und das notwendige domänenspezifische Hintergrundwissen verfügen müssen oder dass der den Daten zugrundeliegende Kontext (repräsentiert durch Metadaten) durch entsprechende technologische Maßnahmen abgespeichert und bei entsprechenden Anfragen an die Daten in geeigneter Form mitgeliefert wird.

Als einer der Vorreiter in der Digitalisierung gestaltet die deutsche Automobilindustrie im Rahmen einiger Vorhaben bereits Konzepte, das Potenzial und die Chancen, die mit derartigen Speicherstrukturen einhergehen, aktiv zu nutzen. Nachfolgendes Fallbeispiel stellt die bisherigen Erkenntnisse und Ergebnisse, die das IMA im Rahmen der Begleitung derartiger Vorhaben erlangt hat, dar. Hiernach werden die Vorteile dieses Vorgehens anhand von Analyseprozessen dargestellt.

### Fallbeispiel: Data Lake Umsetzungen in der Automobilindustrie

Die Erschließung der Potenziale hinsichtlich Prozessverständnis, Automatisierung und Optimierung, die mit Industrial Big Data verbunden sind, stellen zentrale Ziele dar, deren Erreichung von der deutschen Automobilindustrie als Vorreiter dieser Versuche bereits heute angestrebt wird. Innerhalb mehrerer Projektvorhaben haben wir uns daher der Frage angenommen, wie eine erfolgreiche Einführung und flächendeckende, umfassende Nutzbarmachung von Data Lake Architekturen erfolgen kann. Dabei hat sich in allen Vorhaben gezeigt, dass die Umsetzung des sogenannten Distribution Layer, also derjenigen Schicht, die für die durchgängige Versorgung des Data Lake mit Daten aus den Quellsystemen

verantwortlich ist, aufgrund ihrer ausgeprägten Heterogenität eine der entscheidenden Herausforderungen darstellt. Aufgrund der Vielfältigkeit und der Unterschiedlichkeit der Quellsysteme sowie fehlender Standards bei deren initialer Einführung sind individuelle Lösungen (Konnektoren für die Datenintegration) derzeit noch unumgänglich. Dabei zeigt sich jedoch, dass sich wiederkehrende Muster für den Datenabgriff identifizieren und nutzen lassen.

Diese Muster lassen sich durch konfigurierbare Konnektoren in einen generischen Prozess ausgliedern. Hierdurch wiederum ist es möglich, die Anbindung unterschiedlicher Datenquellen auf einige wenige Konnektor-Ansätze zu generalisieren, z.B. unter Berücksichtigung verschiedener Charakteristika standardisierter Datenbanksysteme (z.B. Oracle, IBM oder SAP). Dabei unterstützen die Konnektoren sowohl stream- wie auch batchbasierte (also Anlagen- sowie Datenbank- und Dateisystem-orientierte) Datenquellen. Der derzeit zum Einsatz kommende Technologie-Stack ist so ausgelegt, dass eventuell auftretende Netzausfälle, Überlastungssituationen oder allgemeine Probleme innerhalb der Big-Data-Infrastruktur flexibel ausgleichbar sind.

Zentral innerhalb unseres Ansatzes ist die automatisierte Erstellung von Metadaten und hierauf basierend deren Nutzung zur Anreicherung sämtlicher integrierter Informationen. Aufgrund der generischen Ausgestaltung der zum Einsatz kommenden Konnektoren und der Ableitung von Integrationsmustern war es in der Folge möglich, auf Basis der bestehenden Infrastruktur den Zeitaufwand für die Anbindung neuer Datenquellen systematisch auf wenige Stunden zu reduzieren.

### Erkenntnisgewinn durch Analyse und Künstliche Intelligenz

Der durch die Datenintegration und Metadatenanreicherung gewonnene Datenbestand stellt jedoch nur einen Zwischenschritt dar, um ein ganzheitliches Verständnis der Prozesse zu erlangen, deren Daten zusammengeführt wurden. An diesen Schritt schließen sich Verfahren aus dem Bereich der (kontinuierlichen) Datenanalyse an, welche dazu verwendet werden, Erkenntnisse über Produkte sowie Produktionsprozesse zu generieren. In der Datenanalyse wird, je nach Zielsetzung und Fokus, zwischen deskriptiver, inferenzieller, explorativer und konfirmatorischer Datenanalyse unterschieden [8]. Des Weiteren werden die in dieser Kategorie verorteten Verfahren durch Machine Learning und Verfahren der Künstlichen Intelligenz erweitert.

Der zugrundeliegende Wandel in den Verfahren basiert im Sinne der Big Data Bewegung vor allem auf dem massiven Zuwachs von Rechenleistung an unterschiedlichen Stellen in der Organisation. Der Datenanalyst führt nicht mehr ausschließlich Analysen eines abgeschlossenen Datensatzes durch, dessen Ergebnisse dann als Report zur Verfügung gestellt werden, sondern analysiert kontinuierlich generierte Daten aus dem Produktionsablauf mit nur geringem Zeitverzug.

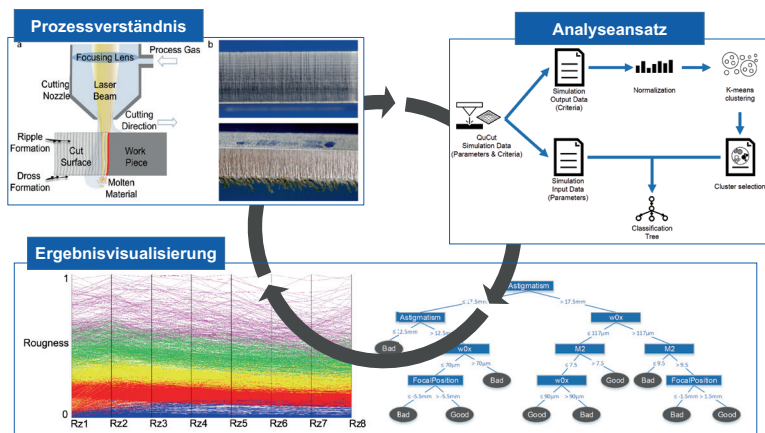
Dieses sogenannte Processing der Daten aus dem Daten-Stream führt zu einer Annäherung von Analyse und Erkenntnisgewinn, die Zeitspannen zwischen Durchführung der Analyse und Umsetzung von hierauf basierenden Maßnahmen zur Optimierung von Prozessen in Echtzeit verringert sich. Die angewandten Verfahren erlauben sowohl eine explorative Analyse großer Datenmengen, um verborgene Strukturen und Erkenntnisse innerhalb der Daten abzuleiten (sog. unsupervised learning), als auch die gezielte Vorhersage von Ereignissen und Zuständen des Untersuchungsgegenstands durch trainierte Modelle (supervised learning). Bei Letzterem werden die Daten und deren darunterliegende Struktur gezielt gelernt. Zu analysierende Ereignisse und Zustände sind hierbei explizit als solche gekennzeichnet, um so dem Algorithmus die Zielvariablen möglicher Vorhersagegrößen zu übergeben. Üblicherweise schließen sich diese Verfahrenskategorien nicht gegenseitig aus, sondern werden meist in Kombination angewandt. So können beispielsweise in großen Datenmengen durch explorative Verfahren Cluster ermittelt werden, die dann durch den Menschen annotiert und in einem überwachten Ansatz als Trainingsdaten für die Ausprägung von Modellansätzen (supervised learning) verwendet werden.

### Fallbeispiel: Analyse von Laserschneidprozessen

In der realen Anwendung finden sich Beispiele insbesondere im Kontext der Vorhersage von Prozess- und Produktqualität. Der Vorhersage geht eine explorative Analyse der Produktionsdaten voran, die die Zusammenhänge zwischen den eigentlichen Produktionsparametern und den Qualitätskriterien erfasst. In [9] wird beispielsweise dieser hybride Ansatz zur Analyse eines simulierten Laserschneidprozesses angewandt. Als Qualitätskriterium dient die Rauigkeit der Kanten im Schnittbild. Als Produktionsparameter werden die Strahlqualität, Strahlversatz und optische Fehler verwendet. Der explorative Teil des hybriden Vorgehensmodells sieht zunächst die Clusterung der Produktionsparame-

### Literatur

- [1] Statista: Internetnutzung in Deutschland, Dossier. 2016. URL: <https://de.statista.com/statistik/studie/id/22540/dokument/internetnutzung-in-deutschland-statista-dossier/>, Abrufdatum 29.05.2017.
- [2] Klein, D.; Tran-Gia, P.; Hartmann, M.: Big Data. In: Informatiklexikon der Gesellschaft für Informatik. URL: <https://www.gi.de/service/informatiklexikon/detailsicht/article/big-data.html>, Abrufdatum 29.05.2017.
- [3] GE Intelligent Platforms: The Rise of Industrial Big Data. 2012.
- [4] Gartner: Gartner Says Worldwide Enterprise IT Spending to Reach \$2.7 Trillion in 2012, 2011. URL: <http://www.gartner.com/newsroom/id/1824919>, Abrufdatum: 29.05.2017.
- [5] Statista: Umsatz mit Business-Intelligence- und Analytics-Software weltweit von 2010 bis 2015 (in Milliarden US-Dollar). 2017. URL: <https://de.statista.com/statistik/daten/studie/259969/umfrage/umsatz-mit-business-intelligence-und-analytics-software-weltweit/>, Abrufdatum 29.05.2017.



**Bild 2: Darstellung des Vorgehens zur Analyse eines Laserschneidprozesses nach [9].**

ter vor, um die Kollektive herauszufiltern, die eine ausreichend geringe Rauigkeit im Schnittbild vorweisen können. Eine Ergebnisvisualisierung in Polarkoordinaten (Bild 2, unten links) zeigt, dass die Prozessparameter im blauen Kollektiv ein vielversprechendes Schnittbild erzeugen. Die in diesem Kollektiv verwendeten Produktionsparameter werden dann mittels eines Klassifikationsverfahrens in einer Baumstruktur dargestellt (Bild 2, unten rechts), um so für den Nutzer die relevanten Bereiche der Produktionsparameter zu visualisieren.

### Fallbeispiel: Tiefziehprozesse in der Automobilindustrie

In der Bearbeitung und Umformung von Blechen von Karosseriebauteilen werden zunehmend intelligente Werkzeuge eingesetzt, die eine aktive Prozessüberwachung und -regelung erlauben. Das Fallbeispiel umfasst die Verbesserung der Produktionsqualität durch die Früherkennung und – im Idealzustand – die hierauf basierende Vermeidung von Rissen. In der Datenanalyse werden die Verläufe der Dehnungsmessstreifen über den Weg hinsichtlich eines funktionalen Bauteils oder eines gerissenen Bauteils klassifiziert. Hierbei wird Expertenwissen in der mathematischen Modellierung formalisiert, sodass ein vollständiges Annotieren des Datensatzes erreicht wird.

Im bisherigen Arbeitsablauf werden die Produktionsparameter nach diversen Testläufen auf Basis des Erfahrungswissens von Domänenexperten manuell festgelegt. Durch die Formalisierung des Expertenwissens und die Möglichkeit zur Klassifikation von Bauteilen vorab wird eine automatisierte Kennwertbestimmung der Produktionsparameter möglich, die eine aktive Früherkennung von Rissen zulässt. Diese wird den Domänenexperten innerhalb ihrer gewohnten Analyseumgebung aufgezeigt, sodass ein direkter Vergleich der manuell eingestellten und automatisiert bestimmten Produktionsparameter ermöglicht wird.

### Fazit und Ausblick

Der Beitrag zeigt auf Basis einer durchgehenden Datenintegrations-Systematik die Bedeutung von Big-Data-basierten Technologien im Kontext der Digitalisierung in der industriellen Produktion auf. Hierbei liegt ein entscheidender Gesichtspunkt in einer Abgrenzung der zum Einsatz kommenden Verfahren im Vergleich zu traditionellen Big-Data-Methoden. Insbesondere der Korrelation heterogener Daten aus bestehenden Informationssystemen kommt hierbei eine hohe Bedeutung für spätere Analysetätigkeiten zu. Neben der Datenintegration und der darauf aufbauenden Analyse der aggregierten Informationen besteht ein weiterer Gesichtspunkt in der erfolgreichen Rückführung von Daten und Analyseergebnissen zum Menschen bzw. Entscheider. So greifen die drei Elemente Datenintegration, Analyse und Rückführung ineinander und sind interdependent zu betrachten, um den Erfolg eines Industrial Big Data Projekts herbeizuführen. Dieser Erfolg reicht hierbei vom tieferen Prozessverständnis bis hin zum Konzept einer Anlagenregelung, die Grenzen der Verfahren sind dabei noch bei weitem nicht ausgeschöpft und sind derzeit Gegenstand von Forschungsvorhaben. Insbesondere Verfahren wie beispielsweise Deep Learning erweisen sich in vielen Bereichen zunehmend als vielversprechender Ansatz, da v.a. die Analyseergebnisse aus komplexen hochdimensionalen Daten von hoher Relevanz für industrielle Anwendungsfälle sind. Beispiele aus Anwendungsbereichen wie der Bild- und Spracherkennung geben hierbei Anlass zu hohen Erwartungen auch für den industriellen Bereich.

Die vorgestellten Fallbeispiele basieren auf realen Daten. Somit wird die Bedeutung einer engen Zusammenarbeit von Forschung und Industrie deutlich. Während in der Forschung die Kompetenzen zur Konzeption von Industrial Big Data-Verfahren bereits vorliegen, kann nur eine aktive Anwendung dieser Verfahren durch die Industrie die erforderlichen Szenarien und im Besonderen reale Prozessdaten generieren. Somit erfordert die digitale Transformation die Neuausrichtung von Kooperationen hin zu einem synergetischen Zusammenwachsen der Forschungsbestrebungen und industrieller Anwendung.

**Schlüsselwörter:** Industrial Big Data, Datenintegration, Künstliche Intelligenz, Machine Learning

*Die Autoren bedanken sich bei der Deutschen Forschungsgemeinschaft DFG für die freundliche Unterstützung der Forschungsarbeiten im Rahmen des Exzellenzclusters „Integrative Produktionstechnik für Hochlohnländer“.*

[6] Tamara Dull: Data Lake vs Data Warehouse: Key Differences. 2015. URL: <http://www.kdnuggets.com/2015/09/data-lake-vs-data-warehouse-key-differences.html>, Abrufdatum 29.05.2017.

[7] Gartner: Gartner Says Beware of the Data Lake Fallacy 2014. URL: <http://www.gartner.com/newsroom/id/2809117>, Abrufdatum: 29.05.2017.

[8] Kamps, U.: Datenanalyse. In: Gabler Wirtschaftslexikon. URL: <http://wirtschaftslexikon.gabler.de/Archiv/1823/datenanalyse-v10.html>, Abrufdatum 29.05.2017.

[9] Tercan, H.; Al Khawli, T.; Eppelt, U.; Büscher, C.; Meisen, T.; Jeschke, S.: Improving the laser cutting process design by machine learning techniques. In: Production engineering 11 (2017) 2, S. 195-203.