



# Data Maturity Assessment – Bewertung der Reife des Datenmanagements für Smart Maintenance

Robert Bernerstätter und Robin Kühnast

Lehrstuhl Wirtschafts- und Betriebswissenschaften, Montanuniversität Leoben, Leoben, Österreich

Eingegangen 8. November 2018; angenommen 14. November 2018; online publiziert 27. November 2018

**Zusammenfassung:** Durch die fortschreitende Digitalisierung in der produzierenden Industrie steigt die Datenmenge in den letzten Jahren kontinuierlich an. Die Auswertung der anfallenden Datenmengen erlaubt einen tieferen Einblick in die Prozesse und fungiert als ein Hebel, um verborgene Effizienzpotenziale zu nutzen. Dadurch steigt der Druck in den Unternehmen, Projekte zu starten, um diese Effizienzpotenziale auszuschöpfen, da diese ein Differenzierungspotenzial darstellen können. Um jedoch ein erfolgreiches Datenanalytikprojekt umzusetzen, müssen die Daten und das Datenmanagement bestimmte Voraussetzungen erfüllen. Es wird daher ein Reifegradmodell vorgestellt, welches dazu dient, die aktuelle Situation in einem Unternehmen im Kontext zur horizontalen und vertikalen Integration zu erheben. Basierend darauf können Verbesserungen aufgezeigt oder passende Business Cases abgeleitet werden. Das Modell wurde in einem Holzverarbeitenden Unternehmen angewandt. Die Ergebnisse und der abgeleitete Business Case werden vorgestellt.

**Schlüsselwörter:** Datenqualität, Reifegradmodell, Assessment, Datenanalytik

## Data Maturity Assessment – Assessing the Maturity of Data Management for Smart Maintenance

**Abstract:** Due to the more intensive progress of digitization in the manufacturing industry, the amount of data has increased considerably in recent years. The analysis of the data volume allows a deeper insight into the processes and acts as a further lever to leverage hidden efficiency potentials. This increases the pressure on companies to start projects in order to exploit these efficiency potentials as these can represent a differentiation potential. In order to implement a successful data analysis project, however,

the data and data management must meet certain requirements. Therefore, a maturity model is presented, which surveys the current situation in a company in the context of industry 4.0 and horizontal and vertical integration. Based on this, improvements can be shown or suitable business cases can be derived. The model was applied in a wood processing company. The results and the derived business case are presented.

**Keywords:** Data quality, Maturity model, Assessment, Data analytics

## 1. Einleitung

Daten gewinnen im Rahmen der Digitalisierung zunehmend an Bedeutung als Ressource. In einer Informationsgesellschaft sind sie wesentlich als Differenzierungsmerkmale und Instrument, um zusätzliche Wertschöpfung in klassischen gesättigten Märkten generieren zu können. Daten gelten als Öl des 21. Jahrhunderts. Ein Mehrwert aus Daten ergibt sich aus Big Data, der schnellen Sammlung großer Daten aus unterschiedlichen Datenquellen und der folgenden Auswertung mittels Analysemethoden des Data-Mining [1].

Um Projekte im Bereich der Datenanalytik umsetzen zu können, ist es nötig, einen Business Case zu definieren. Dieser Business Case ist neben den Anforderungen des Unternehmens auch von den Möglichkeiten abhängig, die im Unternehmen bezüglich der Ressource Daten vorliegen. Neben der Datenqualität, die einen direkten Einfluss auf die Ergebnisse eines solchen Projektes haben, hat auch das Datenmanagement mit dem Einfluss auf die folgende Umsetzung von abgeleiteten automatischen Analyseprozessen Auswirkungen auf Datenanalytikprojekte. Abb. 1 zeigt das Ergebnis einer Umfrage, die speziell diese beiden Einflussfaktoren (ersten beiden Punkte) als vorwiegende Herausforderung zur Umsetzung von Datenanalytikprojekten hervorstreicht.

Es ist daher nötig, die wesentlichen Faktoren zu bewerten, die den Erfolg von Datenanalytikprojekten beeinflus-

Dr. R. Bernerstätter (✉)  
 Lehrstuhl Wirtschafts- und Betriebswissenschaften,  
 Montanuniversität Leoben,  
 Franz-Josef-Straße 18,  
 8700 Leoben, Österreich  
 robert.bernerstaetter@unileoben.ac.at

Abb. 1: Hindernisse bei Daten-analytikprojekten [2]

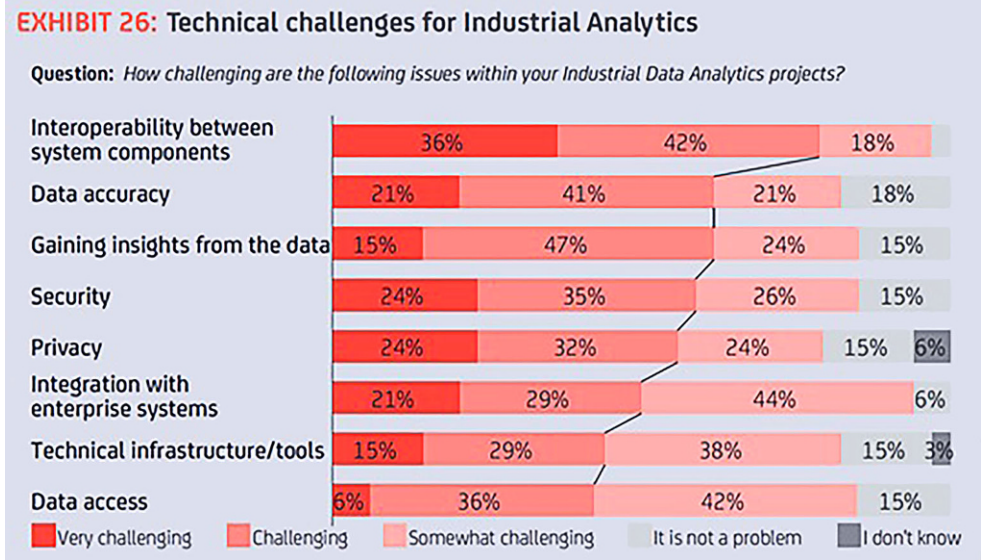
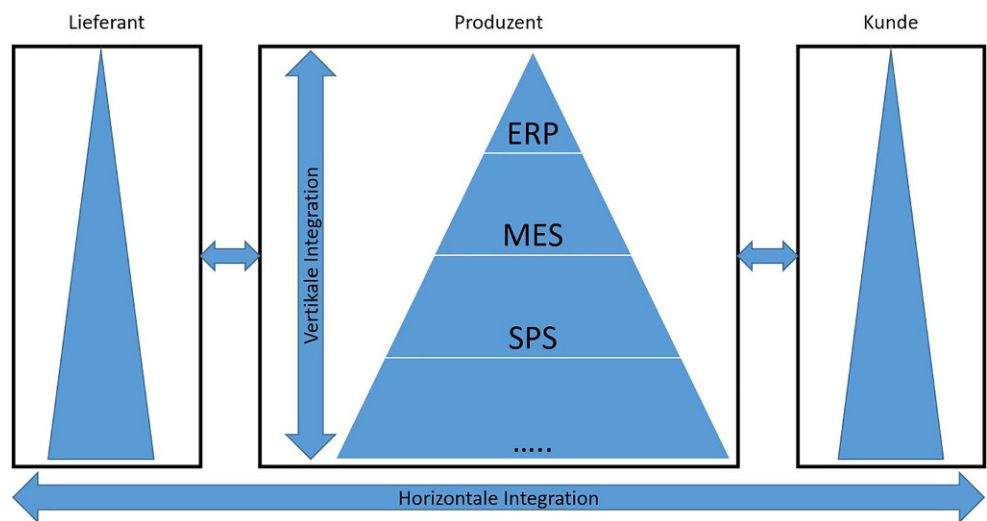


Abb. 2: Horizontale und vertikale Integration



sen. Aus diesem Grund wurde ein Reifegradmodell entwickelt, das die Datenlandschaft vor dem Start eines Projektes umfassend und dabei schnell bewerten kann [3]. Mit den Ausprägungen in den Reifegraden kann eine Aussage über die Erfolgsaussichten eines Projektes getroffen werden und Handlungsempfehlungen abgeleitet werden, um das Unternehmen auf einen Reifegrad zu entwickeln, damit zukünftig Datenanalytikprojekte erfolgreich umgesetzt werden können. Dieses Modell wurde bei einem Unternehmen der Holzverarbeiteten Industrie angewandt, und von dem Ergebnis wurde ein Business Case abgeleitet.

## 2. Datenqualitätsbestimmung

Qualität ist laut ISO „Gesamtheit von Eigenschaften und Merkmalen eines Produktes oder einer Dienstleistung, die sich auf deren Eignung zur Erfüllung festgelegter oder vorausgesetzter Erfordernisse bezieht“ [4].

Um die Qualität von Daten zu bestimmen, steht man vor der Problemstellung der Festlegung von Kriterien, anhand derer die Eignung und die Erfüllung der Erfordernisse gemessen werden kann. Man spricht dabei von fit for use. Je ne Daten, die für die Verarbeitung nötig sind, müssen eine hohe Qualität haben. Wang und Strong haben ein Rahmenwerk geschaffen, welches als Grundlage für eine Vielzahl von Datenqualitätsinitiativen gilt. Tab. 1 zeigt das vorgeschlagene Framework. Es wurden 15 Datenqualitätsdimensionen definiert, die in vier Datenqualitätskategorien unterteilt sind [5].

Die Erfahrung aus Projekten zeigte, dass die Datenqualitätsdimensionen für die industrielle produktionsrelevante Anwendung zu abstrakt sind. Ziel ist es, diese Dimensionen in einem einfachen Bewertungsmodell abzubilden. Es wurde daher ein Reifegradmodell entwickelt, welches die wesentlichen Punkte von Industrie 4.0 aus dem Blickwinkel der Daten abbildet. Da Industrie 4.0 als Begriff sehr weitreichend gefasst ist, wurde der Fokus auf die horizontale und vertikale Integration gelegt.

TABELLE 1  
Datenqualitätsframework [5]

Datenqualität	Inhärente Datenqualität	Zweckabhängige Datenqualität	Darstellungsbezogene Datenqualität	Systemgestützte Datenqualität
Dimensionen	Glaubwürdigkeit (1)	Wertschöpfung (5)	Eindeutige Auslegbarkeit (10)	Zugänglichkeit (14)
	Fehlerfreiheit (2)	Relevanz (6)	Verständlichkeit (11)	Bearbeitbarkeit (15)
	Objektivität (3)	Aktualität (7)	Einheitliche Darstellung (12)	–
	Hohes Ansehen (4)	Vollständigkeit (8)	Übersichtlichkeit (13)	–
	–	Angemessener Umfang (9)	–	–

TABELLE 2  
Reifegradmodell

Reifegradkategorien		Datenerfassung	Datenspeicherung und -übertragung	Datenformate	Datencodierung	Stichprobenumfang	Zeitliche Konsistenz
Reifegrad	5	Durchgängige horizontale und vertikale Integration					
	4	Fast vollständige horizontale und vertikale Integration (Industrie 4.0 mit Schwächen)					
	3	Übergang zur vollständigen horizontalen und vertikalen Integration. Erste Analysen möglich					
	2	Erhebliche Defizite bei der horizontalen und vertikalen Integration. Analysen nur mit manuellem Aufwand					
	1	Keine Ansätze der horizontalen und vertikalen Integration. Höherwertige Analysen (Data Mining) nicht möglich					
		(1), (2), (4), (7)	(4), (14)	(12), (13), (14), (15)	(3), (10), (11), (12)	(6), (9)	(8)
		Zuordnung der Datenqualitätsdimensionen					

Die horizontale Integration beschreibt die Vernetzung von Unternehmen oder Produktionsstufen. Ein idealisiertes Beispiel ist die Vernetzung von Kunden, Produzenten und Lieferanten entlang der Wertschöpfungskette. Die vertikale Integration beschreibt die Vernetzung entlang der Automatisierungspyramide und der Datenverarbeitungssysteme (Abb. 2). Die Herausforderung ist die problemlose Kommunikation von Daten über Systemschnittstellen hinaus [6, 7]. Ein Reifegradmodell bietet die Möglichkeit zur Differenzierung in Reifegradkategorien und über Reifegradlevels.

Es wurden sechs Reifegradkategorien definiert, die die 15 Datenqualitätsdimensionen abbilden sollen. Die ersten beiden Reifegradkategorien bilden die Infrastruktur zur Datenerfassung und die Organisation der Datenspeicherung und -weitergabe ab. Sie haben Einfluss auf die Reifegradausprägung in den anderen Reifegradkategorien. Sie haben auch Einfluss auf die Möglichkeit für die spätere Implementierung eines jeden datenanalytischen Vorgehens. Die letzten drei Kategorien haben eine wesentliche Auswirkung auf die möglichen einsetzbaren Algorithmen für die folgenden Analysen. Sie bestimmen indirekt die Erfolgswahrscheinlichkeit eines Business Case. Die dritte Kategorie beeinflusst wesentlich den Aufbereitungsaufwand sowie die Erfolgswahrscheinlichkeit direkt. Liegen nämlich nicht offen zugängliche proprietäre Formate vor, so kann daran ein Projekt scheitern [8, 9].

Tab. 2 zeigt das Reifegradmodell, wobei die unterste Zeile die Zuordnung der Datenqualitätsdimensionen zu den Reifegradkategorien entsprechend der Zahlen aus Tab. 1

vornimmt. Die Datenqualitätsdimension „Wertschöpfung“ ergibt sich durch einen hohen Reifegrad in allen Kategorien und einem folglich umsetzbaren Projekt. Der höchste Reifegrad 5 zeichnet sich durch die ganzheitliche Realisierung der horizontalen und vertikalen Integration aus. Im niedrigsten Reifegrad 1 wurden in der jeweiligen Kategorie nicht die grundlegendsten Konzepte der Digitalisierung umgesetzt. Für die Reifegradbestimmung wurden für jede Kategorie Fragen definiert, die die Reifegradstufenausprägung erfassen sollen.

### 3. Anwendung des Reifegradmodells

Der Aufwand von Datenanalytik-Projekten verteilt sich ungleichmäßig auf die sechs Phasen des CRISP-DM-Prozessmodells (Abb. 3). Die Datenaufbereitung und -transformation veranschlagen dabei einen Großteil der entstehenden Kosten, wobei der Nutzen erst in der Evaluationsphase fassbar wird. Die Erfolgsaussichten bleiben daher lange Zeit unklar, folglich erhöht sich der Druck, schnelle Ergebnisse zu realisieren. Die Qualität der vorhandenen Daten ist dabei der bestimmende Faktor, sie lässt sich jedoch flächendeckend nur äußerst behäbig verbessern. Die Anwendung der beschriebenen Bestimmung der Datenqualität kann die angesprochenen Probleme aufdecken und teilweise beheben. Im vorliegenden Beispiel werden die Vorgehensweise vorgestellt und das Ergebnis der Anwendung in einem Unternehmen präsentiert. Mit der Determination ergeben sich

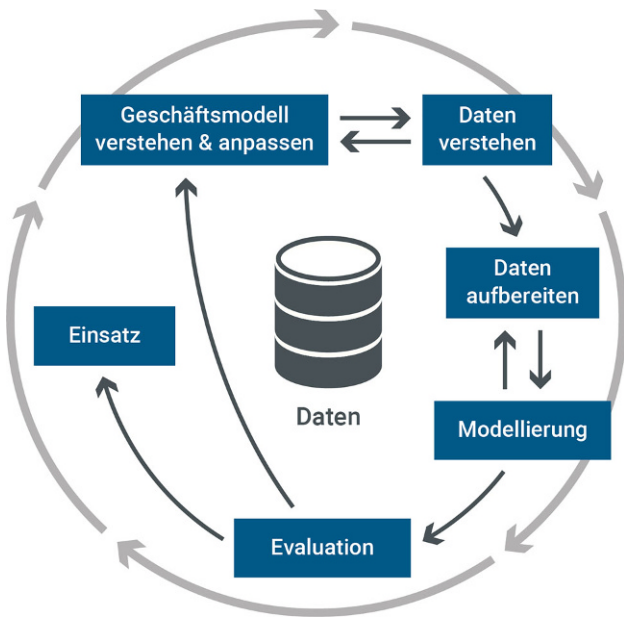


Abb. 3: CRISP-DM

außerdem verschiedene Möglichkeiten zur Modellierung, auf die abschließend eingegangen wird.

### 3.1 Vorgehensweise

Für die Ermittlung der vorherrschenden Datenqualität wurde ein Fragebogen erstellt, der in standardisierten Interviews mit den verantwortlichen Personen für die jeweiligen Datenquellen abgefragt wurde. Der Fragebogen beinhaltet 21 Fragen, wobei zu jeder Frage die Antwortmöglichkeiten bereits ausformuliert vorlagen. Die zu untersuchenden Datenquellen wurden außerdem vorab an die Interviewer übermittelt, um die Relevanz der Frage sowie die Richtigkeit der Antwort sicherzustellen. Nach Abschluss der Be-

fragungen erfolgte die Bewertung anhand der gegebenen Antworten durch die Interviewer.

### 3.2 Ergebnis

Abb. 4 stellt das Ergebnis der beschriebenen Vorgehensweise dar. Die Werte in den jeweiligen Kategorien beziehen sich auf den Durchschnitt der drei relevanten Datenquellen.

Das Ergebnis zeigt deutlich, dass die Reife in dem untersuchten Unternehmen bereits fortgeschritten ist. Eine detaillierte Analyse ermöglicht außerdem, Sofortmaßnahmen einzuleiten, um die den Reifegrad mindernden Faktoren zu beseitigen. So konnten Maßnahmen initiiert werden, die den Reifegrad in Kategorie 1 (Erfassung) und Kategorie 6 (zeitliche Konsistenz) erhöhen und damit bestimmte Hindernisse für die weitere Vorgehensweise reduzieren. Es konnte mit dem durchschnittlichen Ergebnis von 4,1 die Sinnhaftigkeit der Weiterführung eines Datenanalytikprojektes bestätigt werden. Gleichermäßen wurde das Risiko durch die umfassende Betrachtung deutlich reduziert.

### 3.3 Business Case

Neben der verbesserten Transparenz hat die vorgestellte Methodik noch den zusätzlichen Nutzen, den Anwendungsfall zu konkretisieren. In dem beschriebenen Fall erlaubt die hohe zeitliche Konsistenz eine Verknüpfung der Datenbanken miteinander zu einem sogenannten Data Lake. Der Vorteil der Zusammenführung in einer Datenbank wird vor allem in der Ursachenanalyse ersichtlich, bei der der entscheidende Sprung von Korrelation zu Kausalität nur durch einen konsistenten, einheitlichen Zeitstempel realisierbar ist. Für die Schwachstellenanalyse ergeben sich damit neue Möglichkeiten, da der Informationsaustausch der unterschiedlichen Systeme bis dato nur vereinzelt realisiert werden konnte. Durch den hohen Reifegrad besteht die Möglichkeit die Assoziationsanalyse mit einer Sequenzmusteranalyse

Abb. 4: Ergebnis der Reifegradbestimmung

Reifegrad	1	2	3	4	5
Kategorie	„Stahlwerk 1970“	Entstehendes Bewusstsein für Daten	Industrie 4.0-Ansätze	Bereit für Industrie 4.0	Industrie 4.0
				Durchschnitt	
Erfassung			3,8		
Speicherung & Übertragung				4,4	
Datenformate			3,5	4,1	
Datencodierung				4,6	
Datenumfang			4,0		
Zeitliche Konsistenz			4,0		

zu kombinieren und damit eine datengestützte Schwachstellenanalyse umzusetzen [6, 10].

#### 4. Zusammenfassung und Ausblick

Der Einfluss der Datenqualität auf die Erfolgswahrscheinlichkeit von Datenanalytik-Projekten ist vielfach belegt. Das Scheitern erfolgt jedoch meist in einer späten Phase des Projektes und hat damit schon erheblich Ressourcen beansprucht [11, 12].

Um diesen Problemen entgegenzuwirken, wurde ein Reifegradmodell entwickelt, das die Datenlandschaft vor dem Start eines Projektes ganzheitlich und dennoch schnell bewerten kann [3].

Die gewonnene Transparenz unterstützt außerdem die CRISP-DM-Projektphasen „Daten verstehen“ und „Modellierung“ erheblich. Das vorliegende Modell besteht aus sechs Reifegradkategorien, die vier Datenqualitätskategorien und deren 15 Datenqualitätsdimensionen abbilden sollen. Eine vollständige horizontale und vertikale Integration der Systeme und damit auch der Daten kennzeichnet den höchsten Reifegrad.

In dem dargestellten Anwendungsbeispiel wurde mittels Fragebogen und Interviews bei gleichzeitiger Untersuchung der vorhandenen Daten der Reifegrad in den sechs Kategorien für jede der drei Datenquellen bestimmt. Ein durchschnittlicher Wert von 4,1 spricht für eine hohe Erfolgswahrscheinlichkeit eines Datenanalytikprojektes. Zusätzlich konnte die Erkenntnis gewonnen werden, dass die Zusammenführung der Datenbanken in einen Data Lake möglich ist und damit weitreichende, neue Informationen den unterschiedlichen Abteilungen zur Verfügung stehen können. Für die Instandhaltung bedeutet der Zugriff auf Prozess- und Produktionsinformationen einen großen Schritt in Richtung der smart maintenance zu tätigen. Vor allem die erweiterte Schwachstellenanalyse kann sich durch diese Entwicklung zum zentralen Element im Führungssystem der Instandhaltung etablieren [6].

**Funding.** Open access funding provided by Montanuniversität Leoben.

**Open Access** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

#### Literatur

1. Bernerstätter, R.: Daten als Ressource im Kontext von Industrie 4.0 – Kosten und Nutzen von Datenqualität, WINGbusiness, 51 (2018), Nr. 1, S. 36–39
2. Lueth, K. L.; Patsioura, K.; Williams, T. D.; Kermani, Z.Z.: Industrial Analytics 2016/2017 – The current state of data analytics usage in industrial companies, IoT Analytics GmbH, Hamburg, 2016, S. 1–50
3. Bernerstätter, R.: Data Maturity for Smart Factory Applications—An Assessment Model, ACTA Technica Corviniensis-Bulletin of Engineering, XI (2018), Fascicule 1, S. 1–5
4. ÖNORM EN ISO 9000:2015-11-15: Qualitätsmanagementsysteme – Grundlagen und Begriffe, Wien: Österreichisches Norminstitut, 2015
5. Wang, R. Y.; Strong, D. M.: Beyond accuracy—What data quality means to data consumers, Journal of management information systems, 12 (1996), No 4, S. 5–33
6. Biedermann, H.: Lean Smart Maintenance – Controlling: Die Schwachstellenanalyse als zentrales Element im Führungssystem der Instandhaltung, in: Biedermann, H. (Hrsg.): Erfolg durch Lean Smart Maintenance – Bausteine und Wege des Wandels, 1. Aufl., Köln: TÜV-Media, 2017, S. 23–36
7. Geisberger, E.; Broy, M. (Hrsg.): Integrierte Forschungsagenda Cyber-Physical Systems (acatech STUDIE), Heidelberg (u. a.): Springer, 2018
8. Biedermann, H.; Bernerstätter, R.; Kinz, A.: Lean Smart Maintenance – Ressourceneffiziente, wertschöpfungs- und lernorientierte Instandhaltung, in: Horn, G. (Hrsg.): Der Instandhaltungs-Berater, 73. Aufl., Köln: TÜV-Media, 2017, S. 1–56
9. Kinz, A.; Bernerstätter, R.: Instandhaltungsoptimierung mittels Lean Smart Maintenance – Einführung des Lean Smart Maintenance Ansatzes, in: Biedermann, H. (Hrsg.): Lean Smart Maintenance – Konzepte, Instrumente und Anwendungen für eine effiziente und intelligente Instandhaltung, 1. Aufl., Köln: TÜV-Media, 2016, S. 61–100
10. Bernerstätter, R.; Kühnast, R.: Schwachstellenanalyse zur Gewährleistung der Handlungsfähigkeit in komplexen Systemen, in: Biedermann, H. (Hrsg.): Erfolg durch Lean Smart Maintenance – Bausteine und Wege des Wandels, 1. Aufl., Köln: TÜV-Media, 2017, S. 163–183
11. Seyfert, S.; Schlömer, L.; Schiborr, L.A.: Zeit für eine neue Kultur durch Business Intelligence & Advanced Analytics, Sopra Steria SE & BARC, Hamburg, 2018, S. 7
12. Weibl, J.; Hess, T.: Success or Failure of Big Data: Insights of Managerial Challenges from a Technology Assimilation Perspective, Multikonferenz Wirtschaftsinformatik 2018, Lüneburg, 2018, S. 7