

VEREIN
DEUTSCHER
INGENIEURE

VERBAND DER
ELEKTROTECHNIK
ELEKTRONIK
INFORMATIONSTECHNIK

Implementierung und Betrieb von
Big-Data-Anwendungen in der
produzierenden Industrie
Datenqualität

VDI/VDE 3714
Blatt 2
Entwurf

Implementation and operation of big data applications in the manufacturing industry – Data quality

Einsprüche bis 2020-03-31

- vorzugsweise über das VDI-Richtlinien-Einspruchportal <http://www.vdi.de/3714-2>
- in Papierform an VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik Fachbereich Digitale Transformation Postfach 10 11 39 40002 Düsseldorf

Inhalt	Seite
Vorbemerkung	2
Einleitung	2
1 Anwendungsbereich	3
2 Normative Verweise	4
3 Begriffe	4
4 Datenqualität	4
4.1 Struktur zur Beschreibung der Datenqualität	4
4.2 Dimensionen der Datenqualität	5
4.3 Quantitative Bewertung der Datenqualität (Metrik)	11
Schrifttum	12

VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik (GMA)
Fachbereich Digitale Transformation

Vorbemerkung

Der Inhalt dieser Richtlinie ist entstanden unter Beachtung der Vorgaben und Empfehlungen der Richtlinie VDI 1000.

Alle Rechte, insbesondere die des Nachdrucks, der Fotokopie, der elektronischen Verwendung und der Übersetzung, jeweils auszugsweise oder vollständig, sind vorbehalten.

Die Nutzung dieser Richtlinie ist unter Wahrung des Urheberrechts und unter Beachtung der Lizenzbedingungen (www.vdi.de/richtlinien), die in den VDI-Merkblättern geregelt sind, möglich.

An der Erarbeitung dieser Richtlinie waren beteiligt:

Assoc. Prof. Dr. habil. *Martin Atzmüller*, Tilburg

Dipl.-Ing. *Sven Behrendt*, Worms

Dr. Dipl.-Chem. *Raphael Benner*, Heilbronn

Dipl.-Inf. *Rolf Blumenthal*, Lüneburg

Dr.-Ing. *Markus Feist*, Ludwigshafen

Dr. rer. nat. *Eric Frauendorfer*, Burghausen

Dipl.-Ing. *Thomas Froese*, Willich

Prof. Dr. *Jochen Garcke*, Sankt Augustin

Dr. *Jörg Gebhardt*, Hahnstätten

Dr.-Ing. *Ingo Gräf*, Frankfurt

Dipl.-Ing. *Herbert Grieb*, Karlsruhe

Dr.-Ing. *Gunnar Hoffmann*, Essen

Jakob Kinghorst M.Sc., Böblingen

Dipl.-Ing. *Christoph Kugler*, Würzburg

Dr. Dipl.-Chem. *Heiko Kulinna*, Lampertheim

Dr.-Ing. *Silke Müller*, Ratingen

Dr.-Ing. *Dirk Ortloff*, Peterhausen

Dipl.-Math. *Roland Rosen*, München

Prof. Dr.-Ing. *Heiko Tapken*, Osnabrück

Hasan Tercan M.Sc., Aachen

Iris Weiß M.Sc., München

Prof. Dr.-Ing. *Winfried Wilke*, Schweinfurt

Allen, die ehrenamtlich an der Erarbeitung dieser Richtlinie mitgewirkt haben, sei gedankt.

Einleitung

Der Fachausschuss „Big Data“ der VDI-Gesellschaft für Mess- und Automatisierungstechnik hat sich mit der Erstellung dieser Richtlinie der Aufgabe angenommen, den ökonomischen und ökologischen Nutzen von Big Data aufzuzeigen, den Wissenstransfer über verschiedene Industrien und Branchen hinweg zu verbessern und die Implementierung und den Betrieb von Big-Data-Anwendungen in der produzierenden Industrie voranzutreiben und zu vereinheitlichen.

Diese Richtlinienreihe soll eine Orientierung über erforderliche Maßnahmen zur Big-Data-Analyse geben und aufzeigen, welche Methoden für eine zielführende Arbeit geeignet sind bzw. welche Einschränkungen und Hindernisse bestehen. Dem Praktiker sollen Hinweise gegeben werden, welche Methoden und Betrachtungen für den Erfolg eines Big-Data-Projekts hinsichtlich des Einsatzes und des nachhaltigen Betriebs notwendig sind.

Die Richtlinienreihe VDI/VDE 3714 umfasst die Blätter:

Blatt 1 Durchführung von Big-Data-Projekten

Blatt 2 Datenqualität

Blatt 3 Datenbewirtschaftung (in Vorbereitung)

Blatt 4 Analyseverfahrensklassen (in Vorbereitung)

Blatt 5 Modellierungsverfahren (in Vorbereitung)

Blatt 6 Validierung von Modellen (in Vorbereitung)

Blatt 7 Online-Anwendung von datengetriebenen Modellen (in Vorbereitung)

Das vorliegende Blatt 2 betrachtet die Qualitätsdimensionen technischer Daten im Hinblick auf die Generierung, Transfer, Speicherung und Nutzung in Big-Data-Anwendungen.

Eine Liste der aktuell verfügbaren Blätter dieser Richtlinienreihe ist im Internet abrufbar unter www.vdi.de/3714.

Die Richtlinienreihe VDI/VDE 3714 ist im Fachausschuss 7.24 „Big Data“ des Fachbereich 7 „Anwendungsfelder der Automation“ der VDI/VDE-Gesellschaft für Mess- und Automatisierungstechnik (GMA) entstanden. Damit stellen die Produktion sowie die Mess- und Automatisierungstechnik die Schwerpunkte dar. In den Produktionsprozessen werden beispielsweise für Steuerungs- und Regelungsaufgaben oder für die Qualitätssicherung große Datenmengen erhoben, die mittels Datenanalyse für weitere Prozess- und Geschäftsverbesserungen genutzt werden können. Diese Richtlinienreihe gibt eine generelle Orientierung sowie Hinweise auf potenzielle Schwierigkeiten und Hürden bei der Durchführung von Big-Data-Anwendungen von der Entwicklung über die Inbetriebnahme bis zum nachhaltigen Betrieb.

Benachbart zu Big Data finden sich Themen wie das Internet der Dinge (IoT), Vernetzung von Geräten (Smart Devices) oder die zunehmende „Rechnerallgegenwart“ (Ubiquitous Computing) sowie Begriffe wie Business Intelligence, Data Analytics, Advanced Analytics, Data Mining, Smart Data und Data Warehouse Systeme, die generell die Nutzung von Daten adressieren.

Die Richtlinienreihe geht von einer generellen Verfügbarkeit aller benötigten Daten aus. Bezüglich Datenmenge, ihrer Struktur und Integrität wird keine Annahme getroffen. Zur Diskussion und Charakterisierung der Daten helfen die fünf sogenannten „V“, die die einzelnen Dimensionen von Big Data bezeichnen. Die Daten werden charakterisiert durch Umfang (*Volume*), Unterschiedlichkeit (*Variety*) und ihre Schnelligkeit (*Velocity*). Insbesondere bei industriellen Anwendungen sind die Qualität der Daten (*Validity*) und der unternehmerische Mehrwert (*Value*) relevant.

In Big-Data-Anwendungen werden sehr viele Daten aus den unterschiedlichsten Quellen vernetzt, um daraus Informationen und dann letztendlich Prozesswissen zu generieren. Damit stellen die Daten die Basis der wertschöpfenden Modellbildung in Big-Data-Projekten dar. In Bild 1 ist der Prozess der Datennutzung zur Erhöhung der Wettbewerbsfähigkeit in Form einer „Wissenstreppe“ schematisch in Anlehnung an North [1] dargestellt.

In diesem Zusammenhang wird deutlich, dass bei Big-Data-Projekten die Erhöhung der Wettbewerbsfähigkeit und damit der Wertschöpfung immer von den Daten abhängt. Also hat die Qualität der Daten und der daraus generierten Informationen einen erheblichen Einfluss auf die Güte (oder Quantität) der Wertschöpfung von Datenprojekten. Da datengetriebene Modelle mit Trainingsdaten formuliert und mit Validierungsdaten überprüft werden, hängt die Güte des Modells wesentlich von der Qualität der Trainings- und Validierungsdaten ab. Dabei beschreibt der Begriff „Datenqua-

lität“ (siehe auch Abschnitt 4) die Ausprägung von Eigenschaften eines Datenbestands, die dessen Eignung ausmachen, festgelegte und vorausgesetzte Anforderungen zu erfüllen. Daher kann die Datenqualität nur im Zusammenhang mit der Datennutzung und der Datenerzeugung beurteilt werden, die in VDI/VDE 3714 Blatt 3 (in Vorbereitung) ausführlich behandelt werden.

1 Anwendungsbereich

Mit dem Begriff „Big Data“ werden – obwohl er bereits seit einigen Jahren verwendet wird – unverändert sehr unterschiedliche Themen und Aspekte assoziiert und entsprechend in der gesellschaftlichen Diskussion differenziert diskutiert. Die immer weiter voranschreitende digitale Kommunikation, der in der Umsetzung befindliche Breitbandausbau und die überall mögliche Verarbeitungsmöglichkeit von Daten beflügeln diese Diskussion sowohl in der Öffentlichkeit als auch in der Fachwelt. Die Themen reichen von Datenschutz über Datensicherheit bis hin zu generellen Strategien für die digitale Wertschöpfung bei klein- und mittelständischen Unternehmen und auch für Großunternehmen.

Im Kontext dieser Richtlinie geht es bei Big Data um Technologien zur Datenanalyse. Entsprechende Algorithmen und Werkzeuge können Erkenntnisse über betriebliche Abläufe liefern und zu deren Optimierung beitragen. Hierzu bedarf es der Umsetzung dieser Methoden und Werkzeuge zur Verarbeitung, Analyse und Interpretation von umfangreichen und komplexen Daten in Big-Data-Anwendungen.

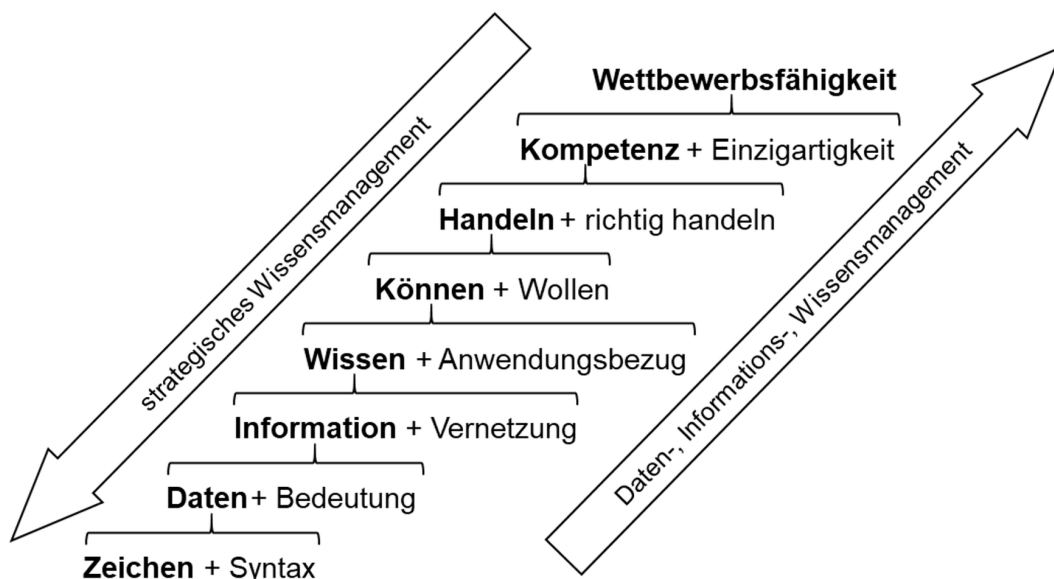


Bild 1. Wissenstreppe nach North

Die Richtlinienreihe unterstützt die Ersteller und Nutzer bei der Vorbereitung, Entwicklung, Inbetriebnahme dieser Anwendungen und ihrem nachhaltigen Einsatz. Letztendlich sollen diese Big-Data-Anwendungen verlässlichere Entscheidungsgrundlagen schaffen, um Produkte und Produktionsprozesse ökonomisch, ökologisch und technisch zu verbessern.

Die Richtlinienreihe soll dazu beitragen, die Vielfalt der in den letzten Jahren durch Forschungs-, Entwicklungs- und Praxisarbeiten entstandenen Erkenntnisse aufzubereiten, die Entwicklung und den Einsatz von Big-Data-Anwendungen in produzierenden Industrien sowie deren Nutzung im regulären Betrieb zu unterstützen.

Zur Zielgruppe gehören alle Stakeholder, von den Praktikern bis zu den Entscheidern, von der Fertigungs- bis zur Prozessindustrie. Die Richtlinienreihe wendet sich dabei an die Nutzer und die Ersteller von Big-Data-Anwendungen in der produzierenden Industrie, unabhängig und übergreifend für alle Führungs- und Fachaufgaben.

2 Normative Verweise

Die folgenden zitierten Dokumente sind für die Anwendung dieser Richtlinie erforderlich:

VDI 2222 Blatt 1:1997-06 Konstruktionsmethodik; Methodisches Entwickeln von Lösungsprinzipien

VDI/VDE 3517 Validierung in der Prozessleittechnik

VDI/VDE 3417 Blatt 1:2019-12 Implementierung und Betrieb von Big-Data-Anwendungen in der produzierenden Industrie; Durchführung von Big-Data-Projekten

VDI 4010:2017-10 Zuverlässigkeitsverbesserung in der Produktnutzungsphase

DIN EN ISO 9000:2015-11 Qualitätsmanagementsysteme; Grundlagen und Begriffe

ISO 13053 Quantitative Verfahren zur Prozessverbesserung; Six Sigma