

<p>VEREIN DEUTSCHER INGENIEURE</p> <p>VERBAND DER ELEKTROTECHNIK ELEKTRONIK INFORMATIONSTECHNIK</p>	<p>Implementierung und Betrieb von Big-Data-Anwendungen in der produzierenden Industrie Validierung von Modellen</p> <p>Implementation and operation of big data applications in the manufacturing industry Validation of models</p>	<p>VDI/VDE 3714</p> <p>Blatt 6 / Part 6</p> <p>Ausgabe deutsch/englisch Issue German/English</p>
---	--	--

Die deutsche Version dieser Richtlinie ist verbindlich.

The German version of this standard shall be taken as authoritative. No guarantee can be given with respect to the English translation.

Inhalt	Seite	Contents	Page
Vorbemerkung .....	2	Preliminary note.....	2
Einleitung.....	2	Introduction.....	2
<b>1 Anwendungsbereich.....</b>	<b>3</b>	<b>1 Scope.....</b>	<b>3</b>
<b>2 Normative Verweise.....</b>	<b>4</b>	<b>2 Normative references.....</b>	<b>4</b>
<b>3 Begriffe.....</b>	<b>5</b>	<b>3 Terms and definitions.....</b>	<b>5</b>
<b>4 Abkürzungen.....</b>	<b>5</b>	<b>4 Abbreviations.....</b>	<b>5</b>
<b>5 Validierung von Modellen.....</b>	<b>5</b>	<b>5 Validation of models.....</b>	<b>5</b>
5.1 Allgemeine Anmerkungen.....	5	5.1 General remarks.....	5
5.2 Teststrategien – Auswahl unabhängiger Stichproben.....	9	5.2 Testing strategies – Selection of independent samples.....	9
5.3 Methoden zur Modellgütemessung.....	11	5.3 Methods for model quality measurement.....	11
Schrifttum .....	21	Bibliography.....	21

Former edition: 04/21 Draft, in German only  
Frühere Ausgabe: 04.21 Entwurf, deutsch

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

VDI-Handbuch Informationstechnik, Band 1: Angewandte Informationstechnik  
VDI/VDE-Handbuch Automatisierungstechnik

## Vorbemerkung

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

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

Eine Liste der aktuell verfügbaren und in Bearbeitung befindlichen Blätter dieser Richtlinienreihe sowie gegebenenfalls zusätzliche Informationen sind im Internet abrufbar unter [www.vdi.de/3714](http://www.vdi.de/3714).

## Einleitung

Der Fachausschuss „Big Data“ der VDI/VDE-Gesellschaft 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 und welche Einschränkungen und Hindernisse bestehen. Praktikern und Praktikerinnen 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

Blatt 4 Analyseverfahrensklassen

Blatt 5 Modellierungsverfahren

**Blatt 6** Validierung von Modellen

Blatt 7 Online-Anwendung von datengetriebenen Modellen

Die Richtlinienreihe VDI/VDE 3714 ist im Fachausschuss 7.24 „Big Data“ des Fachbereich 7 „Digitale Transformation“ der VDI/VDE-Gesellschaft

## Preliminary note

The content of this standard has been developed in strict accordance with the requirements and recommendations of the standard VDI 1000.

We wish to express our gratitude to all honorary contributors to this standard.

A catalogue of all available parts of this series of standards and those in preparation as well as further information, if applicable, can be accessed on the internet at [www.vdi.de/3714](http://www.vdi.de/3714).

## Introduction

The “Big Data” Technical Committee of the VDI/VDE Society Measurement and Automatic Control has taken on the task of drawing up this standard to demonstrate the economic and ecological benefits of big data, to improve the transfer of knowledge across different industries and sectors, and to promote and standardize the implementation and operation of big data applications in the manufacturing industry.

This series of standards is intended to provide orientation on required measures for big data analysis and to show which methods are suitable for target-oriented work and which limitations and obstacles exist. Practitioners should be given advice on which methods and considerations are necessary for the success of a big data project regarding its use and sustainable operation.

The series of standards VDI/VDE 3714 comprises the parts:

Part 1 Implementation of Big Data projects

Part 2 Data quality

Part 3 Data management

Part 4 Analysis process classes

Part 5 Modelling procedures

**Part 6** Validation of models

Part 7 Online application of data-driven models

The series of standards VDI/VDE 3714 is published in the Technical Committee 7.24 “Big Data” of the Technical Division 7 “Digital Transformation” of

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), die 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 der Datenmenge, ihrer Struktur und Integrität wird keine Annahme getroffen. Zur Diskussion und Charakterisierung der Daten helfen die sogenannten „fünf Vs“, die die einzelnen Dimensionen von Big Data bezeichnen. Die Daten werden charakterisiert durch Umfang (*Volume*), Unterschiedlichkeit (*Variety*) und ihre zeitliche Taktung (*Velocity*). Insbesondere bei industriellen Anwendungen sind die Qualität der Daten (*Validity*) und der unternehmerische Mehrwert (*Value*) relevant.

Auf weitere grundsätzliche technische Regeln sei hier hingewiesen, insbesondere im Umfeld von Industrie 4.0:

- VDI 2222 Blatt 1
- VDI 3517
- VDI/VDE 4000
- VDI 4010
- DIN EN ISO 9000
- DIN EN ISO 9001
- DIN EN ISO 9004
- ISO 13053

## 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

the VDI/VDE Society Measurement and Automatic Control (GMA). Thus, production as well as measurement and automation technology represent the focal points. In production processes, for example, large amounts of data are collected for control and regulation tasks or for quality assurance, which can be used for further process and business improvements by means of data analysis. This series of standards provides a general orientation as well as indications of potential difficulties and hurdles in the implementation of big data applications, from development through commissioning to sustainable operation.

Adjacent to big data are topics such as the Internet of things (IoT), the networking of devices (smart devices), or the increasing “computer omnipresence” (ubiquitous computing), as well as terms such as business intelligence, data analytics, advanced analytics, data mining, smart data, and data warehouse systems that generally address the use of data.

The series of standards assumes a general availability of all required data. No assumption is made regarding the amount of data, its structure and integrity. For the discussion and characterization of data, the so-called “five Vs”, which denote the individual dimensions of big data, are helpful. The data is characterized by *volume*, *variety*, and *velocity*. The quality of the data (*validity*) and the added business value (*value*) are particularly relevant for industrial applications.

Further fundamental technical rules should be pointed out here, especially in the environment of Industry 4.0:

- VDI 2222 Part 1
- VDI 3517
- VDI/VDE 4000
- VDI 4010
- DIN EN ISO 9000
- DIN EN ISO 9001
- DIN EN ISO 9004
- ISO 13053

## 1 Scope

Although the term “big data” has been used for several years, it continues to be associated with very different topics and aspects and is accordingly discussed in a differentiated manner in the social debate. The ever-advancing digital communication,