
Joachim Dorschel
Herausgeber

Praxishandbuch Big Data

Wirtschaft – Recht – Technik

 Springer Gabler

Inhaltsverzeichnis

1	Einführung	1
	Werner Dorschel und Joachim Dorschel	
1.1	Keynote: „Die Zeit ist reif für Big Data“	1
1.2	Einführung und Überblick	5
1.2.1	Definitionen	6
1.2.1.1	Volume	7
1.2.1.2	Velocity	7
1.2.1.3	Variety	8
1.2.1.4	Veracity	8
1.2.2	Perspektiven	8
1.2.2.1	Gesellschafts- und rechtspolitische Sicht	9
1.2.2.2	Ökonomische Sicht	9
1.2.2.3	Technische Sichtweise	10
1.2.3	Gegenstand dieses Handbuchs	11
	Literatur	13
2	Wirtschaft	15
	Joachim Dorschel, Werner Dorschel, Ulrich Föhl, Wilhelmus van Geenen, Dieter Hertweck, Martin Kinitzki, Philipp Küller, Carsten Lanquillon, Hauke Mallow, Lothar März, Fouad Omri, Sigurd Schacht, Alphonse Stremler und Elke Theobald	
2.1	Datenorientierung statt Bauchentscheidung: Führungs- und Organisationskultur in der datenorientierten Unternehmung	15
2.1.1	Unternehmerische Sinnhaftigkeit von Big Data Entscheidungen ..	17
2.1.2	Fakten erhöhen die Verantwortlichkeit der einzelnen Akteure ..	22
2.1.3	Kreativität der Mitarbeiter als Teil einer Big Data freundlichen Unternehmenskultur	23
2.1.4	Informations- und Kommunikationskompetenz und Verände- rungskompetenz als Basis schneller Reaktionszeiten	26

2.1.5	Führung wird komplexer und bedarf der Unternehmensmodellierung, sowie des aktiven Managements der Unternehmensarchitektur	28
2.1.6	Zusammenfassung: Tipps für Entscheider, die es bei der Einführung einer Datengetriebenen Entscheidungskultur zu beachten gibt	30
2.2	Enterprise Architecture Management und Big Data	32
2.2.1	Enterprise Architecture Management und Big Data	32
2.2.1.1	EAM ein kurzer Überblick	33
2.2.1.2	Competitive Advantage durch Big Data	38
2.2.2	EAM als Ausgangspunkt für die Etablierung von Big Data im Unternehmen	41
2.2.2.1	Einführung und Entwicklung einer Unternehmensarchitektur	42
2.2.2.2	Einführung von Big Data unter besonderer Beachtung der Unternehmensarchitektur	51
2.2.3	Fazit	54
2.3	Advanced Analytics mit Big Data	55
2.3.1	Begriffsdefinitionen und Varianten	55
2.3.1.1	Analyse und Analytics	55
2.3.1.2	Analytics-Varianten	56
2.3.1.3	Analytics trifft auf Big Data	63
2.3.2	Analyseaufgaben	63
2.3.2.1	Prädiktive Analyseaufgaben	64
2.3.2.2	Beschreibende Analyseaufgaben	66
2.3.3	CRISP-DM: Ein Prozessmodell für Analyseprozesse	68
2.3.3.1	Business Understanding	70
2.3.3.2	Data Understanding	71
2.3.3.3	Data Preparation	71
2.3.3.4	Modeling	71
2.3.3.5	Evaluation	73
2.3.3.6	Deployment	73
2.3.4	Big Data Analytics: Was ist anders?	74
2.3.4.1	Einfluss der Daten auf den Analyseprozess	74
2.3.4.2	Technologische Aspekte	83
2.3.4.3	Der Mensch im Unternehmen als Einflussfaktor	86
2.3.5	Zusammenfassung und Ausblick	88
2.4	Simulation: Neue Einsatzfelder durch Big Data	89
2.4.1	Einführung	89
2.4.2	Planungsablauf in der Fahrzeugindustrie	90
2.4.3	Herausforderungen an die Planung	93
2.4.3.1	Erhöhung der Planungsgenauigkeit	93

2.4.3.2	Einsatz der Simulation in der Planung	94
2.4.3.3	Simulationsgestützte Planung	96
2.4.3.4	Erhöhte Datenanforderungen	97
2.4.4	Praxisbeispiel Automobilendmontage	98
2.4.4.1	Zielsetzung der Anwendung	98
2.4.4.2	Ablauf einer Anwendung	100
2.4.4.3	Datenanforderungen	101
2.4.5	Fazit und Ausblick	103
2.5	Big Data-Analysen: Anwendungsszenarien und Trends	104
2.5.1	Big Data-Analysen: Anwendungsszenarien	105
2.5.1.1	Marketing und Vertrieb	105
2.5.1.2	Forschung und Entwicklung	106
2.5.1.3	Kundenservice	107
2.5.1.4	Produktion	107
2.5.1.5	Logistik	107
2.5.1.6	IT	109
2.5.1.7	Risikomanagement	110
2.5.2	Big Data-Analysen: Trends	110
2.5.2.1	Trends im Rechtswesen	110
2.5.2.2	Trends im Transportwesen	111
2.5.3	Trends im Sozialen Sektor	111
2.5.4	Trends im Gesundheitswesen	112
2.6	Big Data wird zu Smart Data – Big Data in der Marktforschung	112
2.6.1	Big Data in der Marktforschung – Goldgrube oder Datengrab?	112
2.6.2	Der Marktforschungsprozess bei Big Data	114
2.6.2.1	Die Forschungsfrage	114
2.6.2.2	Das Forschungsdesign	115
2.6.2.3	Die Erhebungsphase: Die Nadel im Heuhaufen	116
2.6.3	Aktuelle Herausforderungen für den Big Data Einsatz in der Marktforschung	117
2.6.3.1	Datenzugänglichkeit und Repräsentativität	118
2.6.3.2	Herausforderung Text Mining und Social-Media-Analyse	118
2.6.3.3	Pluralität der Meinungen	119
2.6.3.4	Interpretation multimedialer Daten	120
2.6.3.5	Der Kontext macht den Unterschied	120
2.6.3.6	Von Korrelationen und Kausalitäten	121
2.6.3.7	Topaktuell und doch Schnee von gestern	121
2.6.4	Die Zukunft von Big Data in der Marktforschung	122
2.7	Big Data und Electronic Commerce – Neue Erkenntnisse zur Customer Journey	123
2.7.1	Einleitung	123
2.7.2	Aktuelle Themen im E-Commerce	123

2.7.3	Daten und Datenstrukturen	125
2.7.4	Umfassende Verhaltensanalyse im Rahmen der Customer Journey	127
2.7.4.1	Bedarfs-/Mangelerkennung	128
2.7.4.2	Suche	130
2.7.4.3	Bewertung	130
2.7.4.4	Kauf und Nachkaufphase	131
2.7.5	Wie aus „Big Data“ „Smart Data“ wird	132
2.8	Big Data in der Kreditwirtschaft	134
2.8.1	IT in der Kreditwirtschaft	134
2.8.1.1	Abgrenzung	134
2.8.1.2	Mainframe, Batch, Dialog und Multichannel	134
2.8.1.3	Legacy-Systeme und Standardisierung	135
2.8.1.4	Core-Banking-Systeme und Fachanwendungen	135
2.8.1.5	Datenverwaltung, IDV und Business Intelligence	135
2.8.1.6	Aktuelle Herausforderungen	136
2.8.2	Big Data bewegt die Bank-IT	137
2.8.2.1	Digitalisierung der Kundenbeziehung	137
2.8.2.2	Transparenzanforderungen durch die Bankenaufsicht	138
2.8.3	Einzelne Geschäftsbereiche	139
2.8.3.1	Zahlungsverkehr	139
2.8.3.2	Handel	141
2.8.3.3	Kreditgeschäft	142
2.8.3.4	Gesamtbanksteuerung	144
2.8.3.5	Vertrieb und Multichannel Services	145
2.8.4	Big Data, Outsourcing und Cloud Computing	146
2.8.4.1	Gefahr der Datendesintegration	146
2.8.4.2	Managed Services für Big Data in der Cloud	147
2.8.5	Fazit	147
2.9	Chancen und Herausforderungen von Big Data in der Industrie	148
2.9.1	Unternehmerische Ziele zur Erhöhung der Wertschöpfung	148
2.9.1.1	Anforderungen in Produktion und Logistik	148
2.9.2	Effizienzsteigerung durch integriertes Realtime-Informations- und Datenmanagement in der integrierten Supply Chain	149
2.9.3	Ein Modell der Produktion	150
2.9.4	Leistungssteuerung in Echtzeit für maximale Reaktivität der Supply Chain	150
2.9.5	Ebenen und Stufen der Planung	152
2.9.6	Daten als Schlüsselfaktor des unternehmerischen Erfolges	154
2.9.6.1	Kundenindividuelle Produkte und Leistungen konfigurieren	154
2.9.6.2	Transparenz schaffen	154
2.9.6.3	Reaktionsfähigkeit erhöhen	155

2.9.6.4	Entscheidungen durch Lösungsvorschläge unterstützen	156
2.9.6.5	Neue Produktionskonfigurationen und Produkt- einführungen durch Szenarien absichern	156
2.9.7	Erfolgsfaktoren zum Ausschöpfen der Potenziale von Big Data	158
2.9.7.1	Umgang mit Daten	158
2.9.7.2	Technologien	158
2.9.7.3	Analysetechniken und Algorithmen	159
2.9.7.4	Datenzugriff	160
2.9.7.5	Organisationale Transformation und Führung	160
2.9.8	Fazit	160
	Literatur	161
3	Recht	167
	Michael Bartsch, Olaf Botzem, Thorsten Culmsee, Joachim Dorschel, Jenny Hubertus, Carsten Ulbricht und Thorsten Walter	
3.1	Datenschutz	167
3.1.1	Prinzipien des Datenschutzrechts	167
3.1.1.1	Einleitung	167
3.1.1.2	Prinzipien des Datenschutzrechts	168
3.1.1.3	Fazit	173
3.1.2	Gesetzliche Erlaubnistatbestände und Interessenabwägung	174
3.1.2.1	Anwendungsbereiche und Abgrenzungen von TMG, TKG und BDSG	175
3.1.2.2	Der Legitimationstatbestand der Einwilligung	175
3.1.2.3	Weitere Befugnisse zur Datenverarbeitung nach TMG und TKG	176
3.1.2.4	Weitere Befugnisse zur Datenverarbeitung nach dem TKG	177
3.1.2.5	Weitere Befugnisse zur Datenverarbeitung nach dem BDSG	178
3.1.3	Anonymisierung und Pseudonymisierung; Verschlüsselung	185
3.1.3.1	Anonymisierung	186
3.1.3.2	Pseudonymisierung	187
3.1.3.3	Verschlüsselung	188
3.1.4	Technologien zur Umsetzung datenschutzrechtlicher Anforderungen	190
3.1.5	Zulässigkeit einzelner Phasen von Big Data-Analysen	190
3.1.5.1	Erhebung von Big Data	191
3.1.5.2	Speichern von Big Data	191
3.1.5.3	Personenbezogene Auswertung von Big Data	191
3.1.5.4	Auswertung von Big Data	192
3.1.5.5	Veröffentlichen von Big Data	192

3.1.5.6	Zusammenfassung	193
3.1.6	Betroffenenrechte	193
3.1.6.1	Benachrichtigung des Betroffenen	194
3.1.6.2	Benachrichtigungspflicht bei Web-Crawling und Screen-Scraping?	195
3.1.6.3	Auskunftsanspruch des Betroffenen	196
3.1.6.4	Korrekturrechte	197
3.1.6.5	Das „Recht auf vergessen werden“	198
3.1.7	Internationale Datenverarbeitung	199
3.1.7.1	Anwendbares Recht	199
3.1.7.2	Voraussetzungen für die rechtskonforme Daten- verarbeitung in der EU	201
3.1.7.3	Voraussetzungen für die rechtskonforme Daten- verarbeitung in Drittstaaten	201
3.1.7.4	Praxisfall Cloud Computing	203
3.1.7.5	Zusammenfassung	204
3.1.8	Big Data in der Personalabteilung	205
3.1.8.1	Einführung	205
3.1.8.2	Daten, Daten und noch mehr Daten	205
3.1.8.3	Problemstellung	206
3.1.8.4	Zusammenfassung	210
3.1.9	Automatisierte Entscheidungen und Scoring	211
3.1.9.1	Automatisierte Einzelentscheidungen	211
3.1.9.2	Scoring	212
3.2	Leistungsschutz	213
3.2.1	Urheberrecht an Daten	213
3.2.1.1	Internationales Urheberrecht	214
3.2.1.2	Urheberrechtliche Schutzfähigkeit von Informationen und Daten	214
3.2.1.3	Urheberrechtlicher Schutz der Einzeldaten	216
3.2.1.4	Urheberrechtlicher Schutz von computergenerierten Werken	216
3.2.1.5	Urheberrechtlicher Schutz von Sammel- oder Datenbank- werken	216
3.2.2	Schutz des Datenbankherstellers	217
3.2.2.1	Der Begriff der Datenbank	218
3.2.2.2	Der Begriff des Datenbankherstellers	218
3.2.2.3	Die Rechte des Datenbankherstellers	219
3.2.2.4	Schranken des Rechts des Datenbankherstellers	220
3.2.3	Unlautere gezielte Mitbewerberbehinderung	222
3.2.4	Sonstige Leistungsschutzrechte	224
3.2.4.1	Schutz des Presseverlegers	224

3.3	Integritätsschutz	225
3.3.1	Strafrechtlicher Schutz der Datenintegrität	225
3.3.1.1	Sachbeschädigung (§ 303 StGB)	226
3.3.1.2	§ 303 a Datenveränderung	226
3.3.1.3	Computersabotage (§ 303 b)	228
3.3.1.4	§ 202 a Ausspähen von Daten	228
3.3.1.5	§ 202 b Abfangen von Daten	229
3.3.1.6	§ 202 c Vorbereiten des Ausspähens und Abfangens von Daten	230
3.3.1.7	Ausblick	231
3.3.2	Zivilrechtlicher Schutz: Daten als absolut geschützte Rechtsgüter	231
3.3.2.1	Daten auf eigenen Datenspeichern	232
3.3.2.2	Daten als absolut geschützte Rechtsgüter	233
3.3.2.3	Ansprüche aus Schutzgesetzen	235
3.3.2.4	Rechtsfolgen	235
3.4	Reglementierung der Erhebung von Big Data	237
3.4.1	Rechtliche Bewertung des Screen-Scraping	237
3.4.2	Technische Schutzmaßnahmen	238
3.4.2.1	IP-Sperren	238
3.4.2.2	Captcha	239
3.4.3	Zusammenfassung	239
3.5	Anwendungsszenarien	240
3.5.1	Auswertung des Nutzungsverhaltens im Internet	240
3.5.2	Social Media Analysen	241
3.5.3	Big Data in der Industrie (Industrie 4.0)	242
3.5.4	Zusammenfassung	244
3.6	Verträge über Daten und Datenanalysen	245
3.6.1	Wichtige Vertragstypen	246
3.6.1.1	Kaufverträge über Daten	246
3.6.1.2	Zeitlich begrenzte Datennutzung	246
3.6.1.3	Aufträge zur Datenanalyse	247
3.6.1.4	Datenerhebung im Auftrag	247
3.6.1.5	Datenspeicherung im Auftrag	248
3.6.2	Leistungsstörungen	249
3.6.3	Auftragsdatenverarbeitung	249
	Literatur	251
4	Technik	255
	Gernot Fels, Carsten Lanquillon, Hauke Mallow, Fritz Schinkel und Christ- ian Schulmeyer	
4.1	Grenzen konventioneller Business-Intelligence-Lösungen	255
4.1.1	Business Intelligence: Ein Überblick	255

4.1.1.1	Verwendung und Definitionen des Begriffs	255
4.1.1.2	Evolution entscheidungsunterstützender Systeme	256
4.1.1.3	Diskussion um das Analysespektrum	257
4.1.1.4	BI-Referenzarchitektur	258
4.1.2	Grenzen von BI-Lösungen im Kontext von Big Data	260
4.1.2.1	Volume	260
4.1.2.2	Velocity	261
4.1.2.3	Variety	262
4.1.2.4	Veracity	262
4.1.3	Zusammenfassung	263
4.2	Big Data-Lösungen	263
4.2.1	Anforderungen an Big Data-Lösungen	263
4.2.2	Big Data-Referenzarchitekturen	263
4.2.2.1	Funktionale Big Data-Referenzarchitektur	263
4.2.2.2	Erweiterung einer Data-Warehouse-Architektur mit Big Data-Technologien	275
4.2.3	Zusammenfassung und Ausblick	277
4.3	IT-Infrastrukturen für Big Data	278
4.3.1	Herausforderungen an die Infrastruktur	278
4.3.2	Verteilte Parallelverarbeitung großer Datenbestände	279
4.3.2.1	Apache Hadoop	279
4.3.2.2	Reale oder virtuelle Server?	288
4.3.3	NoSQL-Datenbanken	288
4.3.3.1	Key-Value Stores	290
4.3.3.2	Beispiel: Key-Value Store mit Produktinformationen	290
4.3.3.3	Dokument-orientierte Datenbanken (Document Stores)	291
4.3.3.4	Spaltenorientierte Datenbanken (Columnar Stores)	291
4.3.3.5	Graph-Datenbanken (Graph Databases)	292
4.3.4	In-Memory-Technologien	294
4.3.4.1	In-Memory-Datenbanken (IMDB)	295
4.3.4.2	In-Memory Data Grids (IMDG)	296
4.3.5	Verarbeitung großer Ereignisströme	297
4.3.6	Referenzarchitektur für Big Data-Infrastrukturen	300
4.3.7	Lambda-Architektur	301
4.3.7.1	Impala	303
4.3.7.2	Storm	304
4.3.8	Betrieb von Big Data-Infrastrukturen	305
4.3.8.1	IaaS, PaaS, SaaS oder sogar Data Science als Service?	306
4.4	Big Data-Analyse auf Basis technischer Methoden und Systeme	307
4.4.1	Herausforderungen an Big Data-Analyse	307
4.4.1.1	Was sind Big Data aus technischer Sicht?	307
4.4.1.2	Abgrenzung zu BI	307

4.4.1.3	Datenmengen	308
4.4.1.4	Heterogenität der Datenquellen und der Datenformate sowie fehlende Beschreibung	309
4.4.2	Daten	309
4.4.2.1	Unstrukturierte und semistrukturierte Daten	309
4.4.2.2	Text und nicht-Text-Formate (Audio, Video, Grafik, Bilder)	310
4.4.2.3	Multilinguale Daten	311
4.4.2.4	Datenzugriff	311
4.4.3	Systemische Grundlagen	312
4.4.3.1	Indexerstellung	312
4.4.3.2	In Memory Computing	312
4.4.3.3	MapReduce	313
4.4.3.4	Skalierbarkeit	313
4.4.4	Methoden	314
4.4.4.1	Suche ist nicht gleich Suche	314
4.4.4.2	Keywordbasierte Suche	315
4.4.4.3	Linguistik und Semantik	316
4.4.4.4	Wissensmodelle, Taxonomien und Ontologien	317
4.4.4.5	Assoziative Methoden der Suche	319
4.4.4.6	Case Based Reasoning (CBR)	320
4.4.4.7	Mischformen/Kombinationen	321
4.4.5	Zeitlicher Aspekt	322
4.4.5.1	Retrospektive Analysen	322
4.4.5.2	Echtzeitanalysen	322
4.4.6	Erkenntnisziele der Big Data-Analyse	323
4.4.6.1	Datengold	323
4.4.6.2	Vorhersagen	324
4.4.6.3	Schwache Signale	325
4.4.6.4	Neue Erkenntnisse (knowing the unknown unknown)	327
4.4.6.5	Relationen/Verknüpfung von Daten	328
4.4.7	Zusammenfassung	328
Literatur		329
Sachverzeichnis		331