

Knut Hildebrand • Marcus Gebauer
Holger Hinrichs • Michael Mielke
Hrsg.

Daten- und Informationsqualität

Auf dem Weg zur Information Excellence

4., überarbeitete und erweiterte Auflage

Springer Vieweg

Inhaltsverzeichnis

Teil I Informationsqualität - Grundlagen

1 Was wissen wir über Information?	3
Florian Engelmann und Christoph Großmann	
1.1 Einleitung	3
1.2 Grundlegung	5
1.3 Information im Wissens-und Informationsmanagement	7
1.4 SHANNONsche Informationstheorie	11
1.5 STEINMÜLLERs Informationsmodell	13
1.5.1 STEINMÜLLERs System-und Prozessverständnis	14
1.5.2 Information als allgemeines Modell	14
1.5.3 Modell eines Informationssystems unter Einbezug der Semiotik ...	15
1.5.4 Fazit	16
1.6 Information als Produktionsfaktor	17
1.6.1 Perspektive der Produktionstheorie	17
1.6.2 Produktionsfaktor Information	20
1.6.3 Fazit	20
1.7 Zusammenfassung des Beitrages	21
Literatur	21
2 Informationsqualität – Definitionen, Dimensionen und Begriffe	23
Jan P. Rohweder, Gerhard Kasten, Dirk Malzahn, Andrea Piro und Joachim Schmid	
2.1 Einleitung	23
2.2 IQ-Dimensionen und Definitionen	25
2.2.1 Die 15 IQ-Dimensionen im Überblick:	26
2.2.2 Graphische Darstellung der 15 IQ-Dimensionen und 4 IQ-Kategorien	27
2.2.3 Die 15 IQ-Dimensionen: Definitionen und Beispiele	29
2.3 Zusammenfassung und Ausblick	41
Literatur	42

Teil II Methoden – Techniken – Tools – Regelwerke/Standards

3 Datenqualitätsmetriken für ein ökonomisch orientiertes Qualitätsmanagement	47
Bernd Heinrich und Mathias Klier	
3.1 Einleitung	47
3.2 Anforderungen an Datenqualitätsmetriken	49
3.3 Bisherige Beiträge zur Messung von Datenqualität	50
3.4 Metriken und Messverfahren für DQ	52
3.4.1 Metrik für die DQ-Dimension Vollständigkeit	52
3.4.2 Metrik für die DQ-Dimension Fehlerfreiheit	55
3.4.3 Metrik für die DQ-Dimension Konsistenz	58
3.4.4 Metrik für die DQ-Dimension Aktualität	59
3.5 Praktische Anwendung der Metrik für Aktualität	61
3.6 Zusammenfassung und Ausblick	62
Literatur	63
4 Datenqualitätsmanagement –Steigerung der Datenqualität mit Methode ...	67
Niels Weigel	
4.1 Die Bedeutung des Total Data Quality Management	68
4.1.1 Vorgehensmodelle	68
4.1.2 Datenqualitätsmanagement sichert Ihren Unternehmenserfolg ...	69
4.2 Phasen eines ganzheitlichen Datenqualitätsmanagements	69
4.2.1 Initiierung des Datenqualitätsprojekts	70
4.2.2 Definition der Datenqualitätsanforderungen	73
4.2.3 Messung der vorhandenen Datenqualität	76
4.2.4 Analyse der Fehlerursachen	79
4.2.5 Verbesserung der Datenqualität	81
4.2.6 Permanente Überwachung der Datenqualität	82
4.3 Anreize für ein Datenqualitätsmanagement	83
Literatur	84
5 Strukturierte Datenanalyse, Profiling und Geschäftsregeln	87
Marcus Gebauer und Ulrich Windheuser	
5.1 Datenqualität	87
5.2 Merkmale der Datenqualität	89
5.3 Geschäftsregeln	92
5.4 Methoden der Datenanalyse	93
5.5 Metriken im Detail	95
5.6 Datenqualität in der Anwendung	96
Literatur	100

6 Datenbereinigung zielgerichtet eingesetzt zur permanenten Datenqualitätssteigerung	101
Marcus Zwirner	
6.1 Definition „Datenbereinigung“	102
6.2 Ursachenanalyse	103
6.3 Bewertungskriterien für Datenfehler und Korrekturmaßnahmen	103
6.4 Methoden des Datenqualitätsmanagements	107
6.5 Datenqualitätsmaßnahmen im Detail	109
6.6 Zusammenfassung	120
7 Datenintegration und Deduplizierung	121
Jens Bleiholder und Joachim Schmid	
7.1 Schritt 1: Schema Matching	124
7.2 Schritt 2: Dublettenerkennung	127
7.2.1 Auswirkungen von Dubletten	128
7.2.2 Entstehung von Dubletten	129
7.2.3 Erkennen von Dubletten	130
7.2.4 Durchführung der Dublettenerkennung	131
7.3 Schritt 3: Datenfusion	133
7.3.1 Konflikte ignorieren	134
7.3.2 Konflikte vermeiden	134
7.3.3 Konflikte auflösen	135
7.4 Erweiterungen	137
7.4.1 Strukturierung	137
7.4.2 Standardisierung	137
7.5 Zusammenfassung	138
Literatur	139
8 Definition von Datenarten zur konsistenten Kommunikation im Unternehmen	141
Andrea Piro und Marcus Gebauer	
8.1 Einleitung und Zielsetzung	141
8.1.1 Informationsqualität und Datenarten	142
8.2 Datenarten in der Informationslandschaft	143
8.3 Beschreibungskriterien	143
8.3.1 Beschreibung der Eigenschaften	144
8.3.2 Beschreibung des Kontextes	146
8.4 Beispiele für den Praxiseinsatz	148
8.4.1 Analyseebenen der Informationsqualität	148
8.4.2 Visualisierung des IQ-Status	149
8.5 Zusammenfassung	152
Literatur	153

9 Suchmaschinen und Informationsqualität: Status quo, Problemfelder, Entwicklungstendenzen	155
Christian Maaß und Gernot Gräfe	
9.1 Ausgangssituation	155
9.2 Charakterisierung algorithmenbasierter Suchmaschinen	157
9.2.1 Funktionsweise algorithmenbasierter Suchmaschinen	157
9.2.2 Anfrageabhängige Ranking-Faktoren	157
9.2.3 Anfrageunabhängige Ranking-Kriterien	159
9.3 Semantisches Web und semantische Suchmaschinen	161
9.3.1 Vision und Grundlagen des semantischen Webs	161
9.3.2 Technische Grundlagen des semantischen Web	164
9.3.3 Problemfelder und Herausforderungen im Bereich der semantischen Suche	164
9.4 Fazit und Ausblick	166
Literatur	167
10 Bedeutung der Informationsqualität bei Kaufentscheidungen im Internet	169
Gernot Gräfe und Christian Maaß	
10.1 Einleitung	169
10.2 Informationsqualität in Entscheidungsprozessen	170
10.2.1 Informationen und Kaufentscheidungen	170
10.2.2 Informationsqualitätskriterien	172
10.3 Ursachen mangelnder Informationsqualität im Internet	174
10.3.1 Opportunistische Verhaltensspielräume der Anbieter	174
10.3.2 Informationsqualität aus der Nachfragerperspektive	177
10.3.3 Gründe für Opportunismus im Internet	180
10.4 Fazit und Handlungsempfehlungen	184
Literatur	187
11 Datenqualitäts-Audits in Projekten	193
Marcus Gebauer und Michael Mielke	
11.1 Einleitung	193
11.2 Abstimmung mit anderen Regelwerken	195
11.3 Glossar	195
11.4 Gebrauch der Generischen Checkliste	196
11.5 Datenqualitätsbewertung einer Datensammlung	198
11.5.1 Anforderungen an das Management	198
11.5.2 Service Level Agreements	199
11.5.3 Organisatorische Spezifikationen	199
11.5.4 Prozess-Definitionen	200

11.5.5	Datensammlung, Datenverarbeitung und Datennutzung	202
11.5.6	Messung, Maßnahmen und Überwachung	203
11.5.7	Technische Anforderungen	204
11.5.8	Dokumentation	206
11.6	Zusammenfassung	206
12	Bewertung der Informationsqualität im Enterprise 2.0	209
	Sven Ahlheid, Gernot Gräfe, Alexander Krebs und Dirk Schuster	
12.1	Einführung	209
12.2	Beurteilung der Informationsqualität einer Enterprise	
	2.0 Wissensplattform mittels eines hybriden Ansatzes	210
12.2.1	Automatische Beurteilung der Informationsqualität	211
12.2.2	Implizites Nutzer-Feedback	213
12.2.3	Explizites Nutzer-Feedback	214
12.2.4	Zusammenwirken der drei Ansätze und Fazit	215
	Literatur	216
 Teil III Organisation		
13	Organisatorische Ansiedlung eines Datenqualitätsmanagements	223
	Jens Lüssem	
13.1	Einführung	223
13.1.1	Motivation	223
13.1.2	Gliederung des Kapitels	224
13.2	Datenqualitätsmanagement – Entwicklungsstufen und Aufgaben	225
13.2.1	Sicherung der Datenqualität	226
13.2.2	Management der Datenqualität	226
13.3	Datenqualitätsmanagement – Ansiedlung im Unternehmen	227
13.3.1	Kopplung von Datenqualitätsmanagement mit anderen Unternehmensbereichen	227
13.3.2	Folgerungen für die Ansiedlung eines Datenqualitätsmanagements	229
13.4	Datenqualitätsmanagement in Projekten	230
13.4.1	Aufgaben des Datenqualitätsmanagements in Projekten	230
13.4.2	Organisatorische Verankerung des Datenqualitätsmanagements in Projekten	231
13.5	Zusammenfassung und Ausblick	232
13.5.1	Zusammenfassung	232
13.5.2	Ausblick	233
	Literatur	233

14 Organisatorische Maßnahmen für gute Datenqualität	235
Jürg Wolf	
14.1 Messungen, Ursachen und generische Ansätze	235
14.1.1 Möglichen Arten von Datenqualitätsmängeln	235
14.1.2 Datenqualitätsmängel – Entstehung und Bekämpfung	236
14.1.3 Vier Generische Ansätze	237
14.1.4 Aus den generischen Ansätzen abgeleitete Strategien	238
14.2 Strategie A: Transparenz schafft Vertrauen	239
14.2.1 Ansatzpunkt dieser Strategie	239
14.2.2 Nutzen dieser Strategie	240
14.2.3 Nachteile und Risiken dieser Strategie	240
14.3 Strategie B: Definition von Verantwortlichkeiten	240
14.3.1 Ansatzpunkt dieser Strategie	240
14.3.2 Positionierung dieser Businessrollen im Modell	240
14.3.3 Nutzen dieser Strategie	242
14.3.4 Nachteile und Risiken dieser Strategie	242
14.4 Strategie C: gezielt Abhängigkeiten suchen	242
14.4.1 Ansatzpunkt dieser Strategie	242
14.4.2 Gezielte Definition von Master und Slave	243
14.4.3 Nutzen dieser Strategie	244
14.4.4 Nachteile und Risiken dieser Strategie	244
14.5 Strategie D: Daten-Lifecycle auf Basis des Prozesses	245
14.5.1 Ansatzpunkt dieser Strategie	245
14.5.2 Der Prozess und Lebenszyklus	245
14.5.3 Nutzen dieser Strategie	246
14.5.4 Nachteile und Risiken dieser Strategie	246
14.6 Strategie E: Niederschwellige Verbesserungs-Werkzeuge	246
14.6.1 Ansatzpunkt dieser Strategie	246
14.6.2 Beispiel eines niederschweligen Verbesserungs-Werkzeuges ...	247
14.6.3 Die Infrastruktur dieses Werkzeuges	248
14.6.4 Nutzen dieser Strategie	248
14.6.5 Nachteile und Risiken dieser Strategie	249
14.7 Vor- und Nachteile aller erwähneter Strategien	249
14.7.1 Der Prozess ist die Vorgabe	249
14.7.2 Das Saatkorn ist der Beginn	249
14.7.3 Komplexität des Systems und Datenvolumen	249
14.8 Vorgehen bei der Umsetzung dieser Strategien	250
14.8.1 Kontakt zwischen den Parteien	250
14.8.2 Management-Unterstützung	250
14.9 Schlussfolgerungen und Ausblick	250
Literatur	251

15 Informationsmanagementprozesse im Unternehmen	253
Klaus Schwinn	
15.1 Motivation	253
15.2 Ausgangslage	254
15.3 Bewertung	255
15.4 Informationsmanagementprozess	257
15.5 Schema einer Informationsplanung	259
15.6 Datenlandkarte und Datenarchitektur	260
15.7 Geschäftsprozesse und Informationsmanagementprozess	262
15.8 Qualitätsaspekte	263
15.9 Ökonomische Aspekte	266
15.10 Zusammenfassung	268
Literatur	268
16 Data Governance	269
Boris Otto und Kristin Weber	
16.1 Einführung	269
16.2 Stand der Wissenschaft und Praxis	271
16.2.1 Abgrenzung des DQM	271
16.2.2 Bedeutung des Governance-Begriffs	271
16.2.3 Data Governance	272
16.2.4 Grundmuster für IT Governance	273
16.2.5 Einflussfaktoren auf IT Governance	274
16.3 Ein Modell für Data Governance	276
16.3.1 Rollen	276
16.3.2 Aufgaben	277
16.3.3 Zuständigkeiten	280
16.3.4 Einflussfaktoren und Gestaltungsparameter	281
16.4 Zusammenfassung	284
Literatur	285
17 IQM-Reifegradmodell für die Bewertung und Verbesserung des Information Lifecycle Management Prozesses	287
Saäa Baäkarada, Marcus Gebauer, Andy Koronios und Jing Gao	
17.1 Einleitung	287
17.2 Hintergrund	288
17.2.1 Total Quality Management	288
17.2.2 QM-Reifegrad	289
17.2.3 Information Quality Management	290
17.2.4 Existierende IQM-Reifegradmodelle	291
17.3 Methodologie	291
17.3.1 Die Delphi-Methode	291

17.4	IQM-Reifegradmodell	292
17.4.1	Chaotisch	292
17.4.2	Reaktiv	293
17.4.3	Messend	294
17.4.4	Steuernd	295
17.4.5	Optimierend	295
17.5	Zusammenfassung und Ausblick	296
	Literatur	296
18	Master Data Life Cycle – Management der Materialstammdaten in SAP®	299
	Knut Hildebrand	
18.1	Stammdaten – die Basis der Informationssysteme	299
18.2	Stammdatenqualität führt zu Prozessqualität	301
18.2.1	Qualitätseigenschaften	301
18.2.2	Probleme der Datenqualität und ihre Auswirkungen	302
18.3	Master Data Life Cycle (MDLC) – der Stammdatenprozess	302
18.3.1	Statuskonzept	302
18.3.2	Hindernisse und Problemfälle	306
18.3.3	Tools der SAP®	307
18.4	Implementierung des MDLC	308
18.5	Resümee	309
	Literatur	309
19	Prinzipien erfolgreichen Informationsqualitätsmanagements im Lichte von Industrie 4.0	311
	Michael Mielke	
19.1	Big Data = Big Data Quality?	311
19.2	Datenqualität und Industrie 4.0	313
19.3	Übergeordnete Grundsätze und Einordnung der IQM-Grundsätze	315
19.4	Verantwortung für die Daten übernehmen	317
19.5	Agile DQ-Entwicklung	319
Teil IV Praxisbeispiele		
20	Ein Entscheidungsmodell zur Weitergabe persönlicher Daten im Internet	325
	Horst Treiblmaier	
20.1	Einleitung	325
20.2	Entscheidungsmodell	327
20.2.1	Intention	329
20.2.2	Nutzen	329
20.2.3	Vertrauen	331

20.2.4	Datenarten	334
20.2.5	Eingabefehler	336
20.3	Ausblick	336
	Literatur	337
21	Einführung eines proaktiven DQ-Managements	339
	Volker Landt	
21.1	Die Bremer Landesbank	339
21.1.1	Der Auftrag	340
21.2	Proaktives Datenqualitätsmanagement	340
21.3	Datenqualitätsorganisation	342
21.4	Eskalationsinstanz	343
21.5	Reporting	343
21.6	Messung von Datenqualität	344
21.6.1	Die Themen	344
21.6.2	Messsysteme	348
21.6.3	Messung bankfachlicher Datenzusammenhänge	349
21.7	Visualisierung der Messergebnisse	351
21.8	Messergebnisse und Fazit	354
22	Informationsqualität für das Management mit TOPAS®	357
	Otto-Ernst Brust, Frank Möller und Werner Skrablies	
22.1	Informationsmanagement für Führungskräfte	357
22.1.1	Qualitätskriterien von Informationen für das Management....	358
22.1.2	Absicherung der Informationsqualitätskriterien: TOPAS®-Methodik	359
22.2	TOPAS®: Methode und Modell	359
22.2.1	Geschäftsprozessmanagement (GPM) mit der TOPAS®-Methode	359
22.2.2	Regelkreis für das Geschäftsprozessmanagement	360
22.2.3	4-Ebenen-Modell (Business Excellence Kriterien)	360
22.3	Anwendung: Management von Informationen und Daten	364
22.3.1	Informationen über Strukturen und Abläufe	364
22.3.2	Informationen und Daten für die strategische und operative Planung	365
22.4	Informationserfassung	367
22.4.1	Quellen zur Informationserfassung	368
22.4.2	Anforderungen zur Sicherung der Informations- und Datenqualität	368
22.5	Informationsverarbeitung	368
22.5.1	Operative Planung, Ziele	369
22.5.2	Festlegung von Kennzahlenstrukturen (KPI) in der Balanced Scorecard	369

22.5.3	Sicherung der Informations- und Datenqualität: Kennzahlensteckbrief	370
22.5.4	Erfolgskontrolle via KPIs : Reviews und Audits	371
22.6	Informationsdarstellung	371
22.6.1	Standard-Reporting mit dem TOPAS®-Prozessmodell	371
22.6.2	IT-Portal für das Management: Business Cockpit	372
22.6.3	Effizienzsicherung der Prozess-Architektur	372
22.6.4	Identifizierung von Korrekturmaßnahmen und Kontinuierlicher Verbesserungsprozess (KVP)	373
22.7	Wirkungsspektrum von TOPAS®	374
	Literatur	375
23	Datenqualitäts-Modell der Volkswagen Financial Services AG	377
	Helena Moser	
23.1	Einleitung	377
23.2	Das Projekt „Datenqualität Strukturen/Standards und Drittmarktfähigkeit“	379
23.2.1	Warum ist Datenqualität nötig?	379
23.2.2	Projektauftrag	379
23.2.3	Projektziel	380
23.2.4	Ist-Analyse	380
23.2.5	Prozessanalyse	381
23.2.6	Sollkonzept	382
23.2.7	Das Datenqualitäts-Modell und deren Zuständigkeiten (Abb. 23.1)	384
23.2.8	Das Datenqualitäts-Modell und sein Regelwerk	388
23.2.9	Monitoring/Reports	389
23.2.10	Realisierungs- und Einführungsphase	390
23.3	Fazit	390
24	Verknüpfung von DQ-Indikatoren mit KPIs und Auswirkungen auf das Return on Investment	393
	Frank Block	
24.1	Beispiele zur Illustration von DQ-Problemen	394
24.2	Wie wirken sich DQ-Probleme auf Unternehmen aus – Der Zusammenhang zwischen Daten- und Prozessqualität	395
24.2.1	Beispiel – Call Center	396
24.2.2	Beispiel – Kundenbeziehungsmanagement (CRM)	396
24.2.3	Beispiel – Data Mining Prozess im Marketingumfeld	397
24.2.4	Beispiel – Direktmailprozess	398
24.3	Wie viel kosten schlechte Daten den Unternehmer?	400
24.4	Der Einfluss von DQ-Indikatoren auf KP-Indikatoren – wie beeinflusst Datenqualität den Unternehmenserfolg?	401

24.5	Beschreibung eines KPI orientierten DQ-Managementprozesses	403
24.5.1	Phase 1 – Selektiere zu untersuchende Komponenten	404
24.5.2	Phase 2 – Mitarbeiterbefragung	405
24.5.3	Phase 3 – DQ-Assessment	407
24.5.4	Phase 4 – Validieren und Quantifizieren	408
24.5.5	Phase 5 – DQ-Projekte definieren, Korrekturmaßnahmen durchführen	409
24.6	Fallstudie – Banque Cantonale Vaudoise (BCV)	411
	Literatur	414
25	Gewährleistung einer hohen Artikelstammdatenqualität im Global Data Synchronisation Network (GDSN)	415
	Sascha Kasper	
25.1	Global Data Synchronization Network (GDSN)	416
25.2	Bausteine des GDSN zur Optimierung der Artikelstammdatenqualität...	421
25.2.1	Data Quality Framework (DQF)	421
25.2.2	GDSN Package Measurement Rules und Implementation Guide	426
25.2.3	Data Quality Discussion Group	427
25.2.4	GDSN-Umsetzungsleitfaden zur technischen Anwendung im Rahmen der Lebensmittelinformations-Verordnung	429
25.3	Data Quality Gate	432
25.4	Zusammenfassung und Ausblick	432
	Literatur	433
	Stichwortverzeichnis	437