

Führungssysteme und ausgewählte Massnahmen zur Steuerung von Konzerndatenqualität

DISSERTATION
der Universität St. Gallen,
Hochschule für Wirtschafts-,
Rechts- und Sozialwissenschaften
sowie Internationale Beziehungen (HSG)
zur Erlangung der Würde eines
Doktors der Wirtschaftswissenschaften

vorgelegt von

Kai M. Hüner

aus

Deutschland

Genehmigt auf Antrag der Herren

Prof. Dr. Hubert Österle

und

Prof. Dr. Walter Brenner

Dissertation Nr. 3903

Logos Verlag Berlin GmbH, 2011

Die Universität St. Gallen, Hochschule für Wirtschafts-, Rechts- und Sozialwissenschaften sowie Internationale Beziehungen (HSG), gestattet hiermit die Drucklegung der vorliegenden Dissertation, ohne damit zu den darin ausgesprochenen Anschauungen Stellung zu nehmen.

St. Gallen, den 13. Mai 2011

Der Rektor:

Prof. Dr. Thomas Bieger

Zusammenfassung

Datendefekte gefährden den Geschäftserfolg von Unternehmen. So können veraltete, ungenaue oder unvollständige Daten bei der Ausführung vermeintlich einfacher Prozessaktivitäten wie dem Transport von Shampoo-Flaschen, der Verfügbarkeitsprüfung von Telekommunikationsprodukten oder dem Zählen von Gleisen falsche Ergebnisse, Verzögerungen oder andere Minderungen der Prozessleistung verursachen. Um kritische Datendefekte frühzeitig erkennen und beheben zu können, sollten Unternehmen die Qualität insbesondere ihrer konzernweit genutzten Daten dauerhaft und systematisch überwachen. Unternehmen können ausserdem versuchen, mit präventiven Massnahmen das Auftreten von Datendefekten zu verhindern oder zumindest ihre Wirkung auf die Leistung betroffener Geschäftsprozesse zu reduzieren.

Die Dissertation stellt zur Unterstützung eines qualitätsorientierten Managements konzernweit genutzter Daten vier Artefakte vor. Eine Methode zur Spezifikation geschäftsorientierter Datenqualitätskennzahlen unterstützt die Identifikation geschäftskritischer Datendefekte und die Spezifikation von Datenqualitätskennzahlen für die dauerhafte Überwachung dieser Defekte. Ein Reifegradmodell ermöglicht die Bewertung und den Vergleich von Strukturen für qualitätsorientiertes Konzerndatenmanagement. Ein Wiki-basierter fachlicher Metadatenkatalog unterstützt die kollaborative Definition und Pflege eines unternehmensweit einheitlichen Verständnisses von Geschäftsobjekten und Konzerndaten. Und ein Funktionsreferenzmodell unterstützt die bedarfsorientierte Analyse und Weiterentwicklung der Funktionalität von Informationssystemen für qualitätsorientiertes Management von Konzerndaten.

Summary

Data defects compromise the business success of companies. For example, outdated, inaccurate or incomplete data may cause wrong results, delays or decreased performance even of supposedly simple processes such as transportation of shampoo bottles, availability checks for telecommunication products or counting of railway components. For being able to identify and correct critical data defects in time, companies should regularly and systematically monitor data quality, in particular quality of corporate data. By using preventive measures companies also could try to prevent the occurrence of data defects or at least to decrease the defects' impact on business process performance.

For supporting corporate data quality management, the thesis at hand proposes four artifacts. A method for the identification of business oriented data quality metrics supports the identification of business critical data defects and the specification of data quality metrics for regularly monitoring those defects. A maturity model allows for assessing and comparing structures for corporate data quality management. A wiki-based business data dictionary supports collaborative creation and maintenance of comprehensive and unambiguous business object and corporate data definitions. And a functional reference model supports functionality analysis and enhancement of corporate data quality management applications.

Inhaltsübersicht

Teil A	1
1 Einleitung	1
1.1 Ausgangslage und Handlungsbedarf	1
1.2 Ziele, Adressaten und Nutzen der Arbeit.....	5
1.3 Forschungsmethodik und Gestaltungsprozesse	7
1.4 Aufbau der Arbeit	11
2 Grundlagen	13
2.1 Business Engineering als Bezugsrahmen	13
2.2 Datenarten und Datenqualität	14
2.3 Qualitätsorientiertes Management von Konzerndaten.....	19
3 Ergebnisse und wirtschaftliche Bewertung	23
3.1 Zusammenhang der Ergebnisse	23
3.2 Aufwands- und Nutzenbetrachtung	26
4 Zusammenfassung und Ausblick	35
4.1 Erkenntnisse.....	35
4.2 Einschränkungen.....	36
4.3 Weiterer Forschungsbedarf.....	38
Teil B	41
1 Identification of business oriented data quality metrics	41
1.1 Introduction.....	41
1.2 Background	43
1.3 Method design process.....	45
1.4 Method Composition	48
1.5 Conclusion and Further Research	54
2 Specifying data quality metrics that matter to business	57
2.1 Introduction.....	57
2.2 Definition of Fundamental Concepts	59
2.3 Literature review	61

2.4	Research methodology.....	62
2.5	Method design.....	66
2.6	Evaluation	76
2.7	Discussion and outlook.....	81
3	Methode zur Spezifikation geschäftsorientierter Datenqualitätskennzahlen..	83
3.1	Einleitung.....	84
3.2	Grundlagen.....	86
3.3	Forschungsmethodik.....	88
3.4	Stand der Forschung	91
3.5	Methodenentwurf und Anwendung der Methode.....	101
3.6	Evaluation der Methode.....	130
3.7	Schlussfolgerungen und weiterer Forschungsbedarf	133
4	Product data quality in supply chains: the case of Beiersdorf.....	135
4.1	Introduction.....	135
4.2	Data quality in cross-company supply chains.....	136
4.3	Research methodology.....	138
4.4	The Case of Beiersdorf	140
4.5	Implications for practice and research	151
5	A cybernetic view on data quality management.....	153
5.1	Introduction.....	153
5.2	State of the art	154
5.3	Research methodology.....	156
5.4	Simulation approach	156
5.5	Procedure Model for Application	164
5.6	Example of Application	165
5.7	Discussion and Future Work.....	167
6	Towards a maturity model for corporate data quality management.....	169
6.1	Introduction.....	169
6.2	Background.....	170
6.3	Research Methodology	175

6.4	CDQ maturity model.....	179
6.5	Conclusion and future work.....	185
7	The effect of using a semantic wiki for metadata management: A controlled experiment.....	187
7.1	Introduction.....	187
7.2	Background.....	189
7.3	Research methodology.....	192
7.4	Results and analysis.....	195
7.5	Conclusion and future work.....	201
8	Collaborative management of business metadata.....	203
8.1	Introduction.....	204
8.2	State of the art.....	206
8.3	Research approach.....	208
8.4	Business metadata repository design.....	209
8.5	Artifact evaluation.....	214
8.6	Conclusion and future research.....	216
9	Toward a functional reference model for master data quality management	217
9.1	Introduction.....	217
9.2	Related work.....	220
9.3	Research approach.....	221
9.4	Reference model design.....	226
9.5	Reference model application.....	229
9.6	Evaluation.....	234
9.7	Conclusion and future research.....	236
	Literaturverzeichnis.....	239
Anhang A	Komplette Publikationsliste des Autors.....	267
Anhang B	Gestaltungsprozesse.....	269
Anhang C	Erweiterung des BE-Methodenkerns.....	274
Anhang D	Anhänge zu Publikation B-2.....	276

Anhang E	Anhänge zu Publikation B-3	284
E.1	Entitätstypen des Metamodells der Methode.....	284
E.2	Datenqualitätsdimensionen	286
E.3	Kostenarten defekter Daten.....	292
E.4	Identifikation von Kausalketten.....	295
E.5	Anforderungen an Datenqualitätskennzahlen.....	301
E.6	Messverfahren.....	303
E.7	Dokumentationsvorlagen	312
E.8	Fokusgruppe und Interview-Leitfaden.....	322
Anhang F	Anhänge zu Publikation B-4	324
Anhang G	Anhänge zu Publikation B-8	327
G.1	Focus group participants	327
G.2	Interview guideline	328
Anhang H	Anhänge zu Publikation B-9	329

Inhaltsverzeichnis

Teil A	1
1 Einleitung	1
1.1 Ausgangslage und Handlungsbedarf	1
1.2 Ziele, Adressaten und Nutzen der Arbeit.....	5
1.3 Forschungsmethodik und Gestaltungsprozesse	7
1.4 Aufbau der Arbeit	11
2 Grundlagen	13
2.1 Business Engineering als Bezugsrahmen	13
2.2 Datenarten und Datenqualität	14
2.2.1 Daten als Repräsentation von Geschäftsobjekten.....	14
2.2.2 Konzerndaten.....	15
2.2.3 Datenqualität.....	17
2.3 Qualitätsorientiertes Management von Konzerndaten.....	19
3 Ergebnisse und wirtschaftliche Bewertung	23
3.1 Zusammenhang der Ergebnisse	23
3.2 Aufwands- und Nutzenbetrachtung	26
3.2.1 Kennzahlenmethode	27
3.2.2 Reifegradmodell	29
3.2.3 BDD-Wiki.....	30
3.2.4 Funktionsreferenzmodell.....	32
4 Zusammenfassung und Ausblick	35
4.1 Erkenntnisse.....	35
4.2 Einschränkungen.....	36
4.3 Weiterer Forschungsbedarf.....	38
Teil B	41
1 Identification of business oriented data quality metrics	41
1.1 Introduction.....	41
1.2 Background	43

1.2.1	Basics.....	43
1.2.2	Related work.....	44
1.2.3	Research methodology	44
1.3	Method design process.....	45
1.3.1	Overview.....	45
1.3.2	Case A: Customer data used in customer service processes	46
1.3.3	Case B: Design data used in manufacturing processes	47
1.3.4	Case C: Material data used in maintenance processes	47
1.4	Method Composition	48
1.4.1	Meta-model.....	49
1.4.2	Procedure Model.....	51
1.4.3	Roles	53
1.5	Conclusion and Further Research.....	54
2	Specifying data quality metrics that matter to business	57
2.1	Introduction.....	57
2.2	Definition of Fundamental Concepts.....	59
2.2.1	Data and information	59
2.2.2	Data quality.....	60
2.2.3	Data quality management	60
2.2.4	Data quality measurement	60
2.3	Literature review	61
2.4	Research methodology.....	62
2.4.1	Method design process	62
2.4.2	Method Engineering	64
2.4.3	Application of the method	65
2.5	Method design.....	66
2.5.1	Overview and structure of the method	66
2.5.2	Meta-model.....	68
2.5.3	Roles	68
2.5.4	Procedure model and techniques	70
2.6	Evaluation	76

2.6.1	Overall evaluation approach.....	76
2.6.2	Evaluation results.....	78
2.7	Discussion and outlook.....	81
2.7.1	Research results and limitations	81
2.7.2	Outlook for future research.....	82
3	Methode zur Spezifikation geschäftsorientierter Datenqualitätskennzahlen..	83
3.1	Einleitung.....	84
3.2	Grundlagen.....	86
3.2.1	Business Engineering als Bezugsrahmen	86
3.2.2	Datenqualität und Datenqualitätsmanagement	88
3.3	Forschungsmethodik.....	88
3.3.1	Forschungsprozess.....	89
3.3.2	Methoden-Engineering	89
3.4	Stand der Forschung	91
3.4.1	Batini et al.....	91
3.4.2	Caballero et al.....	92
3.4.3	DAMA	92
3.4.4	Larry P. English.....	93
3.4.5	Heinrich et al.....	94
3.4.6	IBM.....	94
3.4.7	ISO / IEEE	95
3.4.8	David Loshin	95
3.4.9	Jack E. Olson	96
3.4.10	Price et al.	97
3.4.11	Thomas C. Redman	97
3.4.12	Wang et al.....	98
3.4.13	Vergleichende Bewertung vorgestellter Ansätze	100
3.5	Methodenentwurf und Anwendung der Methode.....	101
3.5.1	Metamodell.....	102
3.5.2	Methodenfragmente.....	104
3.6	Evaluation der Methode.....	130

3.7	Schlussfolgerungen und weiterer Forschungsbedarf	133
4	Product data quality in supply chains: the case of Beiersdorf	135
4.1	Introduction	135
4.2	Data quality in cross-company supply chains	136
4.2.1	Data quality management	136
4.2.2	Data quality measurement	137
4.2.3	Data sharing in supply chains	138
4.3	Research methodology	138
4.4	The Case of Beiersdorf	140
4.4.1	Current situation	140
4.4.2	Data quality measurement	144
4.4.3	Findings and discussion	150
4.5	Implications for practice and research	151
5	A cybernetic view on data quality management	153
5.1	Introduction	153
5.2	State of the art	154
5.2.1	Measuring data quality	154
5.2.2	Costs and benefits of data quality (management)	155
5.3	Research methodology	156
5.4	Simulation approach	156
5.4.1	Data Quality Management as a closed-loop control system	156
5.4.2	Simulation objectives	158
5.4.3	Impact Patterns	161
5.4.4	Simulation procedure	163
5.5	Procedure Model for Application	164
5.6	Example of Application	165
5.6.1	Business Scenario	165
5.6.2	Simulation Model and Application	166
5.7	Discussion and Future Work	167
6	Towards a maturity model for corporate data quality management	169

6.1	Introduction.....	169
6.2	Background.....	170
6.2.1	Corporate data quality management.....	170
6.2.2	Maturity models.....	172
6.2.3	Analysis of existing maturity models.....	172
6.3	Research Methodology.....	175
6.3.1	Research context.....	175
6.3.2	Design research.....	176
6.3.3	Reference modeling.....	177
6.4	CDQ maturity model.....	179
6.4.1	Model requirements.....	179
6.4.2	Model construction.....	181
6.4.3	Model discussion and evaluation.....	184
6.5	Conclusion and future work.....	185
7	The effect of using a semantic wiki for metadata management: A controlled experiment.....	187
7.1	Introduction.....	187
7.1.1	Corporate metadata management.....	187
7.1.2	Business scenario.....	188
7.2	Background.....	189
7.2.1	Definition of metadata.....	189
7.2.2	Wikis and semantic wikis.....	190
7.3	Research methodology.....	192
7.3.1	Research objective.....	192
7.3.2	Scenario.....	192
7.3.3	Experiment design.....	193
7.3.4	Experiment implementation.....	194
7.4	Results and analysis.....	195
7.4.1	Statistical background.....	195
7.4.2	Cross-over analysis.....	196
7.4.3	Analysis of T_{Find} and T_{Link}	197

7.4.4	Analysis of T _{Collect}	199
7.4.5	Threats to validity	200
7.4.6	Discussion	201
7.5	Conclusion and future work	201
8	Collaborative management of business metadata	203
8.1	Introduction	204
8.1.1	Motivation and problem statement	204
8.1.2	Research objective and paper structure	205
8.2	State of the art	206
8.2.1	Business metadata management	206
8.2.2	Systems for collaborative metadata management	207
8.3	Research approach	208
8.4	Business metadata repository design	209
8.4.1	Requirements for effective business metadata management	209
8.4.2	Processes for collaborative management of business metadata	211
8.4.3	A wiki-based prototype	213
8.5	Artifact evaluation	214
8.5.1	Focus group evaluation results	214
8.5.2	Results of prototype testing at Bayer CropScience	215
8.6	Conclusion and future research	216
9	Toward a functional reference model for master data quality management	217
9.1	Introduction	217
9.1.1	Motivation and problem statement	217
9.1.2	Research question and contribution	219
9.2	Related work	220
9.2.1	Data quality management	220
9.2.2	Master data management	220
9.2.3	MDM and DQM	221
9.3	Research approach	221
9.3.1	Overview	221

9.3.2	Research process.....	222
9.4	Reference model design.....	226
9.4.1	Design foundations.....	226
9.4.2	Model overview.....	227
9.4.3	Functional details.....	228
9.5	Reference model application.....	229
9.5.1	Case study overview.....	229
9.5.2	Case Study Approach and Results.....	231
9.6	Evaluation.....	234
9.6.1	Multi-perspective artifact evaluation.....	234
9.6.2	Compliance with GOM.....	235
9.7	Conclusion and future research.....	236
	Literaturverzeichnis.....	239
Anhang A	Komplette Publikationsliste des Autors.....	267
Anhang B	Gestaltungsprozesse.....	269
Anhang C	Erweiterung des BE-Methodenkerns.....	274
Anhang D	Anhänge zu Publikation B-2.....	276
Anhang E	Anhänge zu Publikation B-3.....	284
E.1	Entitätstypen des Metamodells der Methode.....	284
E.2	Datenqualitätsdimensionen.....	286
E.3	Kostenarten defekter Daten.....	292
E.4	Identifikation von Kausalketten.....	295
E.5	Anforderungen an Datenqualitätskennzahlen.....	301
E.6	Messverfahren.....	303
E.7	Dokumentationsvorlagen.....	312
E.8	Fokusgruppe und Interview-Leitfaden.....	322
Anhang F	Anhänge zu Publikation B-4.....	324
Anhang G	Anhänge zu Publikation B-8.....	327
G.1	Focus group participants.....	327

G.2 Interview guideline	328
Anhang H Anhänge zu Publikation B-9	329

Abkürzungsverzeichnis

1SYNC	Produktdaten-Pool des Unternehmens Beiersdorf AG
AFNOR	Association française de normalisation
AG	Aktiengesellschaft
APO	Advanced Planner and Optimizer (Applikation für die Produktionsplanung)
ARIS	Architektur integrierter Informationssysteme
AS	Präfix zur Kennzeichnungen von Erweiterungen des BE-Methodenkerns durch Alexander Schmidt [Schmidt 2010]
BDD	Fachlicher Metadatenkatalog (engl. business data dictionary)
BE	Business Engineering
BW	Business Warehouse
ca.	ungefähr (auf Lateinisch: circa)
CC	Kompetenzzentrum (engl. competence center)
CDQ	Konzerndatenqualität (engl. corporate data quality)
CDQM	Qualitätsorientiertes Management von Konzerndaten (engl. corporate data quality management)
CHF	Schweizer Franken (ISO-Code)
CMM	Capability Maturity Model (Reifegradmodell)
CRM	Customer Relationship Management
d. h.	das heisst
DAMA	Data Management Association
DD	Datendefekt (engl. data defect)
DOI	Digital Object Identifier (Bezeichner für digitale Objekte)
DP	Gestaltungsprinzip (engl. design principle)
DQ	Datenqualität (engl. data quality)
DQ-Dimension	Datenqualitätsdimension
DQI	Data Quality Index
DQ-Kennzahl	Datenqualitätskennzahl
DQM	Datenqualitätsmanagement
DQ-Messung	Datenqualitätsmessung
DR	Begründung für Gestaltungsentscheidung (engl. design rationale)
Dr.	Doktor
DSR	Design Science Research

DSS	Decision Support System
DTAG	Deutsche Telekom AG
EBP	Enterprise Buyer Professional (Applikation für den Einkauf)
EFQM	European Foundation for Quality Management
engl.	auf Englisch
ERP	Enterprise Resource Planning
et al.	und andere (auf Lateinisch: et alii)
EU	Europäische Union
EUR	Euro (ISO-Code)
GI-FB WI	Fachbereich Wirtschaftsinformatik der Gesellschaft für Informatik
GS1	Global Standards One
GTIN	Global Trade Item Identification Number
HSG	Universität St. Gallen (früher: Hochschule St. Gallen)
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IS	Informationssystem
ISO	Internationale Organisation für Normung (engl. International Organization for Standardization)
IT	Informationstechnologie
LuFV	Leistungs- und Finanzierungsvereinbarung
MDM	Stammdatenmanagement (engl. master data management)
MDQM	Qualitätsorientiertes Stammdatenmanagement (engl. master data quality management)
MDW	Master Data Workbench (Applikation für die Erzeugung von Konzerndaten)
ME	Methoden-Engineering (engl. Method Engineering)
Mio.	Million
MM	Reifegradmodell (engl. maturity model)
MOF	Object Management Facility
Nr.	Nummer
OMG	Object Management Group
org.	im Original
PhD	Doktor (auf Lateinisch: philosophiae doctor)
PID	Reglertyp aus der Regelungstechnik (engl. proportional integral derivative)

PIM	Product Information Management
PLM	Product Lifecycle Management
Prof.	Professor
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
s.	siehe
S.	auf Seite
s. u.	siehe unten
S.A.	Aktiengesellschaften (auf Französisch: société anonyme)
SEI	Software Engineering Institute der Carnegie Mellon University
u. a.	unter anderem
vgl.	vergleiche
VR	Validierungsregel (engl. validation rule)
WKWI	Wissenschaftlichen Kommission Wirtschaftsinformatik im Verband der Hochschullehrer für Betriebswirtschaft e. V.
z. B.	zum Beispiel

Teil A

1 Einleitung

Mit der Entwicklung von Führungssystemen und Massnahmen zur Steuerung der Qualität von Konzerndaten (engl. corporate data) adressiert die Dissertation einen in den letzten Jahren gewachsenen Bedarf von Unternehmen [Friedman et al. 2007, S. 13f, Thoo et al. 2010, S. 11]. Ursachen für die zunehmende Bedeutung von Datenqualität (insb. der Qualität konzernweit genutzter Daten) sind u. a. die steigende Zahl gesetzlicher Anforderungen an die Datenverwaltung in Unternehmen, die Standardisierung zwischenbetrieblicher Prozesse sowie innerbetriebliche Initiativen zur Senkung von Kosten (vgl. Beispiele in Abschnitt 1.1).

Die vorliegende kumulative Dissertation gliedert sich in zwei Teile. Teil A motiviert die Forschungsziele, erläutert die Gestaltungsprozesse der Ergebnisse und ordnet die entwickelten Konzepte in einen Bezugsrahmen ein. Teil B besteht aus wissenschaftlich publizierten Aufsätzen, die die Forschungsergebnisse vorstellen.

1.1 Ausgangslage und Handlungsbedarf

Daten repräsentieren Geschäftsobjekte und somit die Gegenstände der Leistungserstellung in Unternehmen [Österle 1995, S. 87, 185f]. Mehrdeutige, ungenaue oder unvollständige Repräsentationen von Geschäftsobjekten können u. a. zu falschen Entscheidungen, fehlerhaften Produkten oder unzureichender Bereitstellung gesetzlich vorgeschriebener Informationen führen. Die folgenden zwei Beispiele erläutern den Zusammenhang zwischen Datendefekten und Geschäftsproblemen in Unternehmen.

Das Unternehmen Deutsche Telekom AG bietet eine Verfügbarkeitsprüfung für ressourcenintensive Telekommunikationsprodukte an. Über diesen Dienst können (potentielle) Kunden prüfen, ob z. B. IP-TV (über das Internet übertragene Fernsehprogramme) an einer bestimmten Adresse verfügbar ist. Bei einer Anfrage ist zu prüfen, ob die entsprechende Leitung grundsätzlich und unter Berücksichtigung bereits verkaufter Produkte in der Lage ist, die erforderliche Leistung bereitzustellen. Diese Entscheidung erfordert die Analyse von Netzinfrastruktur-, Kunden- und Produktdaten. Ungenau spezifizierte Produktanforderungen, Lücken in der Dokumentation der Netzinfrastruktur oder veraltete Kundendaten können falsche Verfügbarkeitsantworten verursachen.

Der beschriebene Fall zeigt, dass Datendefekte möglichen Umsatz der Deutschen Telekom verhindern können. Denn erhält ein Kunde trotz der tatsächlichen Verfügbarkeit eines angefragten Produkts die Antwort, dass das Produkt nicht verfügbar ist, wird er

es mit hoher Wahrscheinlichkeit nicht bestellen und nach alternativen Anbietern suchen.

Das Unternehmen Beiersdorf AG verwaltet logistische Produktdaten (z. B. Abmessungen, Gewichte) in einem Informationssystem (IS). Dort ist u. a. das erwartete Gewicht einer Shampoo-Flasche (berechnet aus vorgesehener Füllmenge und Dichte der Füllung) hinterlegt. Logistische Planungsprozesse nutzen dieses Datenelement für die Erstellung von Packlisten. Dabei wird u. a. die Anzahl der Shampoo-Flaschen einer Palette festgelegt. Nach der Produktion kann das Produktgewicht von dem im IS gespeicherten Datenelement abweichen¹. Ein erhöhtes Gewicht der einzelnen Flaschen kann dazu führen, dass das Gewicht einer Palette mit Shampoo-Flaschen über dem für Lagerplätze oder Transportmittel zulässigen Maximalgewicht liegt.

Im diesem Fall führen Datendefekte zwar nicht unmittelbar zu geringerem Umsatz von Beiersdorf. Sie verursachen jedoch Mehrarbeit und Verzögerungen durch das Umpacken von Paletten und verringern so die Leistung der Logistikprozesse. Vergessen Mitarbeiter, die vorgenommenen Änderungen (hier: Anzahl Shampoo-Flaschen pro Palette) im IS zu dokumentieren, entsteht ein weiterer Datendefekt, der fehlerhafte Lieferungen an Kunden verursachen kann.

Neben der Auswirkung von Datendefekten auf die Leistung von Geschäftsprozessen zeigt der Fall, dass verbesserte Datenmanagementstrukturen Datendefekte verhindern können. Denn mit einer Prüfung des Flaschengewichts nach der Produktion der ersten Shampoo-Flasche und der Korrektur des theoretisch berechneten Gewichts würden die Packlisten folgender Produktionsläufe automatisch die korrekte Stückzahl enthalten. Und auch das beschriebene Risiko von Lieferungen mit zu wenigen Shampoo-Flaschen könnte eine Datenpflegeaktivität reduzieren, die nach dem Umpacken einer Palette die Korrektur der Stückzahlen im IS fordert.

Datendefekte verursachen jedoch nicht nur Probleme in Primäraktivitäten (z. B. Produktion, Vertrieb), sondern können auch die Leistung von Unterstützungsprozessen mindern. Verwenden Entscheidungsunterstützungssysteme (engl. decision support systems, DSS) defekte Daten, kann dies zu ungenauen Analysen und falschen Entscheidungen führen. Im folgenden Beispiel verhindern Datendefekte die Realisierung von Einsparpotenzialen und schmälern den Erfolg einer ressourcenintensiven Prozessharmonisierung und Applikationskonsolidierung des Unternehmens Bayer CropScience AG. Das Unternehmen nutzt aufgrund der beschriebenen Problematik eine Kennzahl zur Überwachung von Datendefekten [Ebner et al. 2011, S. 10-15].

¹ Das Überfüllen von Shampoo-Flaschen kann betriebswirtschaftlich sogar sinnvoll sein, falls die eigentlich vorgesehene Füllmenge in einer transparenten Flasche den Eindruck erweckt, dass die Flasche nicht voll ist.

Bayer CropScience hat weltweit Geschäftsprozesse vereinheitlicht und fachliche Applikationen zusammengelegt. Ziel dieses Projekts ist u. a. die bessere Unterstützung von Entscheidungsprozessen. Fehlende und unvollständige Datenelemente in Produkthierarchien führen aber zu Problemen bei der Bedarfsplanung, da Wirkstoffbedarfe auf Konzernebene nicht korrekt konsolidiert werden können.

Um Geschäftsprobleme als Auswirkung von Datendefekten zu verhindern, gibt es grundsätzlich zwei Möglichkeiten: Ein bereits aufgetretener Datendefekt kann repariert werden (reaktives Datenqualitätsmanagement, DQM). Oder es wird versucht, das Auftreten von Datendefekten zu verhindern (präventives DQM, vgl. Abschnitt 2.3). Ein Erfolgsfaktor für die Vermeidung von Datendefekten ist die unternehmensweit einheitliche Beschreibung von Geschäftsobjekten. Denn Mehrdeutigkeiten können zu Missverständnissen bei der Ausführung von Geschäftsprozessen (z. B. bei der Eingabe logistischer Daten für eine Shampoo-Flasche) und letztlich zu Datendefekten (z. B. falsches Produktgewicht) führen. Und auch Bayer CropScience wird die konzernweite Bedarfskonsolidierung nur dann realisieren können, wenn Materialien konzernweit einheitlich verstanden und die Produkthierarchien entsprechend diesem Verständnis gepflegt werden. Das folgende Beispiel unterstreicht die Notwendigkeit unternehmensweit einheitlich beschriebener Geschäftsobjekte.

Das Unternehmen DB Netz AG ist die für die Schieneninfrastruktur verantwortliche Tochtergesellschaft der Deutschen Bahn AG. Die DB Netz ist gegenüber der Bundesrepublik Deutschland (Bund) durch die Leistungs- und Finanzierungsvereinbarung (LuFV) verpflichtet, jährlich ein Infrastrukturkataster vorzulegen, das die von ihr betriebenen Schienenwege mit allen wesentlichen Merkmalen der Betriebsanlagen aufführt [Bund 2004, S. 10f]. Auf Basis der darin enthaltenen Stückliste der Gleise, Tunnel, Weichen und anderen Anlagenarten bestimmt der Bund die Höhe der Finanzmittel, die er der DB Netz zur Instandhaltung der Anlagen zur Verfügung stellt.

Als Voraussetzung für hohe Qualität des Infrastrukturkatasters hat die DB Netz die Verwendung gleicher Definitionen für die Infrastrukturanlagen in allen Regionen und Unternehmensfunktionen identifiziert. Denn ohne eine unternehmensweit einheitliche Beschreibungen der Geschäftsobjekttypen ist unklar, welche Anlagen bei der Inventarisierung überhaupt zu zählen sind. Das Verständnis der Geschäftsobjekte ist jedoch regional unterschiedlich: Ein S-Bahn-Tunnel ist in einem Regionalbereich ein unterirdisches Streckenstück, wobei die Distanz zwischen Tunneleingang und Tunnelausgang die Tunnellänge repräsentiert. In einem anderen Regionalbereich ist die Tunnellänge die Distanz zwischen zwei aufeinander folgenden unterirdischen Bahnhöfen.

Datendefekte im Infrastrukturkataster der DB Netz können zu geringeren Einnahmen des Unternehmens führen oder Rückzahlungsforderungen [Bund 2004, S. 16f] verursachen. Die folgenden Beispiele beschreiben vier weitere Gesetze und Standards, die wie die Leistungs- und Finanzierungsvereinbarung zwischen Bund und DB Netz Anforderungen an die Qualität von Daten und die Qualität der Datenverwaltung stellen. Verstöße gegen diese Vorgaben können zu gesetzlichen Strafen, negativer öffentlicher Aufmerksamkeit oder Problemen bei der Ausführung zwischenbetrieblicher Prozesse führen.

Das Standardisierungsgremium Global Standards One (GSI) spezifiziert u. a. die Global Trade Item Identification Number (GTIN) zur weltweit eindeutigen Identifikation von Produkten [GSI 2010, S. 168-175]. Externe Dienstleister bieten basierend auf diesem von GSI bereitgestellten Nummernsystem sogenannte Datenpools zur multilateralen Synchronisation von Produktdaten an. Stellen Konsumgüterhersteller Produktdaten in einem solchen Pool zur Verfügung, müssen sie die von GSI definierten Regeln für die Vergabe von GTINs beachten. So ist beispielsweise für eine Shampoo-Flasche eine neue GTIN zu vergeben, wenn sich das Gewicht der Flasche um mehr als 20 Prozent ändert.

Ein französischer Gesetzentwurf sieht vor, schrittweise die Auszeichnung aller auf dem französischen Markt angebotenen Produkte mit einem sogenannten Carbon Footprint vorzuschreiben [Frankreich 2010, S. Artikel 228]. Konsumgüterhersteller sind dann verpflichtet, die Menge des bei Herstellung und Distribution von Produkten freigesetzten Kohlenstoffdioxids korrekt angeben zu können.

Die Chemikalienverordnung REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) der Europäischen Union (EU) verpflichtet Hersteller und Lieferanten chemischer Erzeugnisse zur Bereitstellung von Informationen, falls der Anteil bestimmter Stoffe in Erzeugnissen über vorgeschriebenen Schwellwerten liegt [EU 2006, S. Artikel 33 Absatz 1 und 2].

Das deutsche Telekommunikationsgesetz schreibt vor, dass ein Telekommunikationsunternehmen Kundendaten nach spätestens einem Jahr löschen muss, wenn ein Kunde seinen Vertrag mit dem Unternehmen kündigt [Bund 2004, S. Paragraph 95 Absatz 3].

Die Beispiele zeigen, dass Datendefekte den Geschäftserfolg von Unternehmen gefährden und Datenqualität daher ein Erfolgsfaktor für Unternehmen ist. Um geschäftskritische Datendefekte frühzeitig erkennen und verhindern zu können, sollten Unternehmen die Qualität ihrer Konzerndaten und ihrer Strukturen für qualitätsorientiertes Management von Konzerndaten (engl. corporate data quality management, CDQM)

dauerhaft überwachen. Für diese Überwachung sind Kennzahlen erforderlich, die die Erfüllung betrieblicher und gesetzlicher Anforderungen an Datenqualität und Datenverwaltung operationalisieren, d. h. technisch messbar und vergleichbar machen. Dabei repräsentieren die Kennzahlen die gestellten Anforderungen z. B. durch Regeln, deren Einhaltung ein IS überprüfen kann.

Neben Artefakten zur Überwachung von Datenqualität und CDQM stellt die Dissertation zwei Artefakte zur Unterstützung eines präventiven CDQM vor. Die beschriebenen Beispiele motivieren insgesamt den Handlungsbedarf für die Arbeit mit den folgenden Fragestellungen:

- Welches methodische Vorgehen ermöglicht Unternehmen die Identifikation geschäftskritischer Datendefekte und die Spezifikation von Kennzahlen für die Überwachung dieser Defekte? Welche Zusammenhänge zwischen Datendefektkosten und Kosten von CDQM-Massnahmen sind bei der Entwicklung von Datenqualitätskennzahlen (DQ-Kennzahlen) zu berücksichtigen?
- Wie kann ein Unternehmen die Qualität seiner CDQM-Funktion bewerten, den Aufbau oder die Umstrukturierung einer solchen Unternehmensfunktion überwachen und Erfahrungen anderer Unternehmen nutzen, um eigene Strukturen für CDQM weiterzuentwickeln?
- Wie kann ein Unternehmen Datenobjekt- und Geschäftsobjekttypen einheitlich und umfassend fachlich beschreiben und den Aufwand der dauerhaften Pflege dieser Beschreibungen auf viele Mitarbeiter verteilen?
- Welche Funktionalität sollte ein IS zur Unterstützung von CDQM in Unternehmen bereitstellen?

1.2 Ziele, Adressaten und Nutzen der Arbeit

Ziel der Dissertation ist eine von der Praxis nutzbare und wissenschaftlich fundierte Konzeptualisierung von Datenqualität, um die Qualität von Konzerndaten systematisch bewerten und verbessern zu können. Die Arbeit stellt dazu die folgenden vier Artefakte zur Unterstützung von CDQM-Funktionen in Unternehmen vor.

- Eine *Methode zur Spezifikation geschäftsorientierter Datenqualitätskennzahlen* (kurz: Kennzahlenmethode) unterstützt die Identifikation geschäftskritischer Datendefekte und die Spezifikation von DQ-Kennzahlen für die dauerhafte Überwachung dieser Defekte.
- Ein *Referenzmodell zur Bewertung des Reifegrads von CDQM* (kurz: Reifegradmodell) ermöglicht die Bewertung und den Vergleich von CDQM-Funktionen. Das Reifegradmodell unterstützt ausserdem die Identifikation von Handlungsbedarfen beim Aufbau und der Umstrukturierung von CDQM.

- Ein *Wiki-basierter fachlicher Metadatenkatalog* (engl. wiki-based business data dictionary, kurz: Wiki-BDD) unterstützt die kollaborative Definition und Pflege eines unternehmensweit einheitlichen Verständnisses von Geschäftsobjekten und Konzerndaten.
- Ein *Funktionsreferenzmodell für IS zur Unterstützung von CDQM* (kurz: Funktionsreferenzmodell) unterstützt die bedarfsorientierte Analyse und Weiterentwicklung der Funktionalität von IS für CDQM.

Gemäss den Prinzipien der Konsortialforschung [Österle/Otto 2010, S. 4] erhebt die Dissertation den Anspruch, nützliche Artefakte für Praxis und Wissenschaft vorzustellen. Den konkreten Nutzen der einzelnen Artefakte diskutieren Abschnitt 3.2 und die einzelnen Publikationen in Teil B. Dieser einleitende Abschnitt beschreibt daher nur die Adressaten der Arbeit und allgemeine Nutzenpotenziale.

In der Praxis adressiert die Dissertation Mitarbeiter, die das Ziel haben, die Qualität von Konzerndaten ihrer Unternehmen zu überwachen, zu verbessern oder zu sichern. Die Erfahrung aus Forschungsprojekten mit Unternehmen zeigt, dass diese Mitarbeiter überwiegend Funktionsbereichen der Informationstechnologie (IT), des Prozessmanagements oder des Supply Chain Managements angehören oder explizit für CDQM geschaffenen Organisationseinheiten zugeordnet sind. Die Artefakte sollen diese Mitarbeiter bei der Ausführung von CDQM-Aufgaben unterstützen. Insbesondere die Spezifikation von DQ-Kennzahlen ist ein Ziel von Unternehmen, die bereits CDQM-Strukturen etabliert haben und nun den Nutzen verbesserter Datenpflegeprozesse, erweiterter Datenarchitekturen oder neuer CDQM-IS überprüfen wollen. Das Reifegradmodell adressiert insbesondere Unternehmen, die CDQM neu aufbauen und dafür eine Vorlage mit Handlungsempfehlungen suchen. Das Modell kann aber auch Unternehmen unterstützen, die sich in einem bestimmten Aufgabenbereich (vgl. Abschnitt 2.3) verbessern wollen. Eine Diskussion von Aufwand und Nutzen (vgl. Kapitel 3) unterstützt die Adressaten in der Praxis ausserdem bei der Entscheidung, ob die Anwendung eines der Artefakte im Einzelfall nützlich ist.

Für die wissenschaftliche Diskussion beschreibt die Dissertation nicht nur die Forschungsergebnisse, sondern ausserdem die Gestaltungsprozesse der einzelnen Artefakte. Die Explikation von Gestaltungsentscheidungen und Evaluationsergebnissen ermöglicht die wissenschaftliche Überprüfung und Weiterentwicklung der Artefakte und bietet zukünftigen Forschungsarbeiten die Möglichkeit, Erkenntnisse zur Gestaltung von Artefakten zu entwickeln. Das Metamodell der Kennzahlenmethode, das Reifegradmodell und das Funktionsreferenzmodell sind ausserdem nach wissenschaftlichen Prinzipien entwickelte Realitätsrepräsentationen und adressieren das grundsätzliche Erkenntnisziel der Wirtschaftsinformatik (vgl. Abschnitt 1.3). Das Verständnis betrieblicher Realität ist dabei eine Voraussetzung für die Gestaltung von Artefakten, die in Unternehmen Nutzen stiften. Das Metamodell der Kennzahlenmethode beschreibt

beispielsweise Zusammenhänge zwischen Datenqualität und der Prozessleistung von Unternehmen und liefert so – neben der Grundlage für die Gestaltung eines nützlichen Artefakts – einen Beitrag zu der wissenschaftlichen Diskussion des Nutzens von IS [DeLone/McLean 1992, DeLone/McLean 2003, Urbach et al. 2009, Buhl et al. 2010]. Als weiteres Beispiel beschreibt das Funktionsreferenzmodell fachliche Anforderungen an ein CDQM-IS und liefert so einen Beitrag für die anforderungsgerechte IS-Entwicklung.

1.3 Forschungsmethodik und Gestaltungsprozesse

Als angewandte Sozialwissenschaft hat die Wirtschaftsinformatik, im Gegensatz zu Grundlagenwissenschaften, einen konstitutiven Praxisbezug und strebt nach Regeln und Modellen zur Schaffung neuer Realitäten [Ulrich 1984, S. 179]. Dieses Gestaltungsziel der Wirtschaftsinformatik [Rautenstrauch 1993, S. 3, Österle et al. 2011, S. 2]) orientiert sich am Selbstverständnis technischer Wissenschaften: Die Realität ist dabei nicht eigentliches Untersuchungsobjekt, sondern Ausgangspunkt für die Untersuchung möglicher zukünftiger Realitäten mit dem Ziel, zweckorientierte Artefakte unter Berücksichtigung gegebener Rahmenbedingungen (z. B. betriebliche Anforderungen, Naturgesetze) zu schaffen [Lenk 1979, S. 189f, Ulrich 1984, S. 174]. Verständnis der Realität und die Fähigkeit, Entwicklungen der Realität prognostizieren zu können (Erkenntnisziele), unterstützen die Gestaltung und Nutzung von Artefakten [Becker et al. 2003, S. 12, vom Brocke 2003, S. 2], sind aber nicht alleiniges Ziel angewandter Forschung [Ulrich 1984, S. 192f].

Die in der vorliegenden Dissertation vorgestellten Artefakte sind Ergebnis gestaltungsorientierter Forschung, die Design Science Research (DSR) als forschungsmethodischem Paradigma folgt. DSR fokussiert das beschriebene Gestaltungsziel der Wirtschaftsinformatik auf die Gestaltung von Artefakten [Simon 1996, S. 1-8] als nützliche Lösung praktischer Probleme [Hevner et al. 2004, S. 76, Wilde/Hess 2007, S. 181] und definiert die Artefakttypen Konstrukt, Modell, Methode und Instanz [March/Smith 1995, S. 256-258]. Gemäss definierten Anforderungen an die Gestaltung von DSR-Artefakten [Hevner et al. 2004, S. 82-90] umfasst der Gestaltungsprozess die Problemidentifikation, die Definition von Gestaltungszielen, die Gestaltung von Artefakten als Problemlösung, die Evaluation der Artefakte sowie nötige Anpassungen mit erneuter Evaluation [Gregor 2006, S. 628f, Peffers et al. 2008, S. 52-56]. Die Evaluation eines Artefakts prüft dabei die Erreichung von Gestaltungs- und Erkenntnisziel: Erstens sollte ein Artefakt einen Beitrag zur Lösung des zuvor definierten Problems leisten und zweitens die Wissensbasis der Praxis und Wissenschaft erweitern [March/Storey 2008, S. 726].

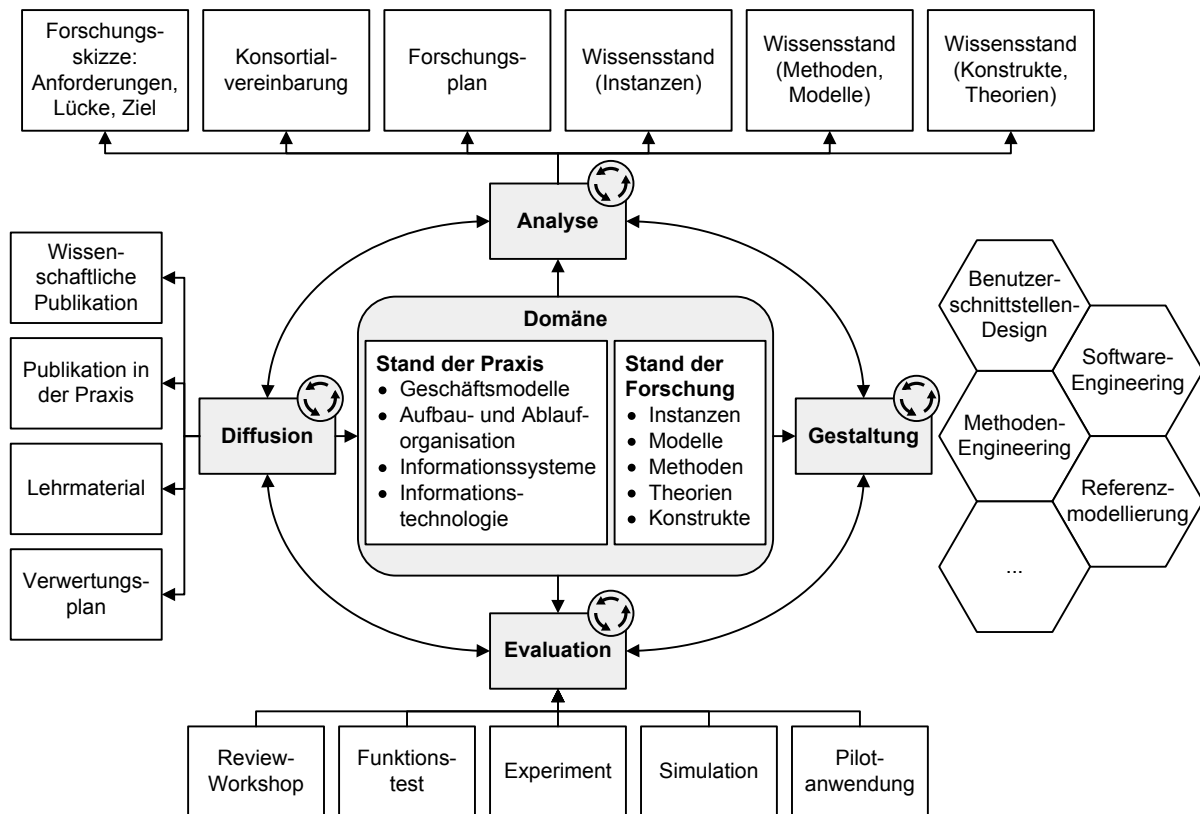


Abbildung A1-1: Konsortialforschung im Überblick [Österle/Otto 2010, S. 6]

Die Gestaltungsprozesse der vorgestellten Artefakte folgen einer Methode für Konsortialforschung [Österle/Otto 2010, S. 4-9]. Konsortialforschung unterstützt die Erreichung von Gestaltungs- und Erkenntniszielen und hat sich für die Gestaltung von Artefakten zur Unterstützung von CDQM als geeigneter Forschungsansatz erwiesen [Otto/Österle 2010]. Abbildung A1-1 zeigt die vier von der Methode vorgeschlagenen Phasen eines Konsortialforschungsprozesses mit jeweils anwendbaren Techniken und erzeugten Ergebnissen. Das Vorgehensmodell der Methode orientiert sich an einem allgemein für DSR vorgeschlagenen Forschungsprozessmodell [Peppers et al. 2008, S. 52-56], setzt aber einen stärkeren Fokus auf die Identifikation praxisrelevanter Forschungsziele (Analysephase) und die Nutzenbewertung der Artefakte (Evaluationsphase).

Die vorgestellten Artefakte sind Ergebnisse der Konsortialforschungsprojekte *Kompetenzzentrum Corporate Data Quality* (CC CDQ, November 2006 bis Oktober 2008) und CC CDQ2 (Dezember 2008 bis September 2010). Die Forschungsdomäne, also der in dem Forschungsprojekt untersuchte Gegenstand [Nunamaker et al. 1991, S. 91-93], ist CDQM. Sie umfasst u. a. Fragestellungen der Forschungsthemen *Datenqualitätsmanagement* (DQM), *Metadatenmanagement* und *Stammdatenmanagement* (MDM). Beide Projekte sind Teil des Forschungsprogramms *Business Engineering* (BE), das an der Universität St. Gallen bereits seit über 20 Jahren Konsortialforschung betreibt [Österle/Otto 2010, S. 2].

Im Verlauf der beiden Konsortialforschungsprojekte sind insgesamt 20 zweitägige Workshops durchgeführt worden, in denen Fachexperten aus Unternehmen, Forscher und weitere Praxisvertreter (z. B. Hersteller von Anwendungs-Software) Artefakte zur Unterstützung von CDQM sowie Fragestellungen und Handlungsbedarfe der Domäne präsentiert und diskutiert haben. Die Fachexperten der Unternehmen haben ausserdem gemeinsam mit Forschern in Projekten neu gestaltete Artefakte eingesetzt und so die Evaluation und Verbesserung der Artefakte unterstützt. Die folgenden zwei Forschungstechniken haben die Gestaltung der vorgestellten Artefakte dominiert:

- *Fokusgruppeninterview.* In einem Fokusgruppeninterview befragt ein Forscher (der Moderator, möglich sind auch mehrere Moderatoren) eine Gruppe von Fachexperten (die Fokusgruppe) und dokumentiert die Ergebnisse (z. B. Antworten auf konkrete Fragen, Bewertungen, Diskussionen) für spätere Analysen [Morgan/Krueger 1993, Stewart et al. 2007, Rosemann/Vessey 2008]. Im Verlauf der Gestaltungsprozesse der vorgestellten Artefakte sind Fokusgruppeninterviews im Rahmen von Konsortialworkshops durchgeführt worden, um Forschungsziele zu identifizieren (Analysephase), Teile von Artefakten zu gestalten (Gestaltungsphase) oder den Nutzen von Artefakten zu bewerten (Evaluationsphase). Die zur Evaluation durchgeführten Fokusgruppeninterviews haben für die Bewertung der Artefakte eine Struktur verwendet, die auf Evaluationsperspektiven für Referenzmodelle (Wirtschaftlichkeit, Anwendung, Konstruktion, Epistemologie) aufbaut [Frank 2007, S. 123-136].
- *Partizipative Fallstudie.* In einem Aktionsforschungsprojekt analysieren Forscher gemeinsam mit Fachexperten eines Unternehmens ein Problem des Unternehmens und erarbeiten gemeinsam eine Lösung [Baskerville/Wood-Harper 1996]. Die Forscher sind dabei nicht nur Beobachter, sondern haben die Möglichkeit, durch Anpassung der Lösung deren Wirkung auf das Unternehmen direkt zu beeinflussen [Susman/Evered 1978, S. 580]. Kanonische Aktionsforschung folgt einem Vorgehensmodell mit fünf Phasen zur Problemanalyse, Massnahmenplanung, Anwendung der Massnahme, Evaluation der Wirkung und Generalisierung der Ergebnisse [Susman/Evered 1978, S. 588f, Baskerville/Wood-Harper 1998, S. 96f]. Eine partizipative Fallstudie ist ein Aktionsforschungsprojekt, in dem dieses Paradigma aus pragmatischen Gründen (z. B. Problemanalyse bereits vor Projektbeginn durchgeführt) nicht eingehalten wird, Forscher und Fachexperten aber dennoch gemeinsam analysieren und gestalten [Baskerville 1997, S. 40f]. Im Verlauf der Gestaltungsprozesse der vorgestellten Artefakte sind partizipative Fallstudien durchgeführt worden, um Artefakte zu gestalten (Gestaltungsphase) oder ihre Anwendbarkeit zu demonstrieren und dabei ihren Nutzen zu bewerten (Evaluationsphase). Das übergeordnete Konsortialforschungsprojekt gleicht die im Vergleich zu einem Aktionsforschungsprojekt geringere Stringenz einer partizipativen Fallstudie

aus, da es z. B. den Rahmen zur Artefaktevaluation durch Fokusgruppen bietet und eine wiederholte Anwendung von Artefakten in ähnlichen partizipativen Fallstudien ermöglicht.

Weitere angewendete Forschungstechniken wie Literaturrecherche oder Gestaltungstechniken (z. B. Methoden-Engineering, Referenzmodellierung) werden in den einzelnen Publikationen dieser Dissertation erläutert (vgl. Teil B). Tabelle A1-1 gibt einen Überblick zu den Gestaltungsprozessen der vier vorgestellten Artefakte. Anhang B enthält detailliertere Beschreibungen der einzelnen Gestaltungsprozesse. Ein typischer Gestaltungsprozess der vorgestellten Artefakte umfasst

- die Diskussion einer praxisrelevanten Problemstellung und Definition eines zugehörigen Forschungsziels in Fokusgruppeninterviews,
- die Analyse der Domäne (Stand der Forschung und Praxis) durch Fallstudien, Fokusgruppeninterviews und Literaturrecherche,
- die Gestaltung eines Artefakts als Problemlösung unter Anwendung von Gestaltungstechniken (z. B. Methoden-Engineering, Referenzmodellierung),
- die Demonstration, Evaluation und Anpassung des Artefakts in Fokusgruppeninterviews und partizipativen Fallstudien und
- die Dokumentation und Publikation des Artefakts für Praxis und Wissenschaft.

Artefakt	Analyse	Gestaltung	Evaluation	Diffusion
Kennzahlenmethode	2 Fokusgruppeninterviews 1 Literaturrecherche	Methoden-Engineering 1 Fokusgruppeninterview 3 partizipative Fallstudien 4 Gestaltungszyklen	5 Fokusgruppeninterviews 3 partizipative Fallstudien	5 Publikationen (s. Kapitel B-1, B-5, B-3, B-4 und B-2)
Reifegradmodell	1 Fokusgruppeninterview 1 Literaturrecherche	Referenzmodellierung 1 Fokusgruppeninterview 2 Gestaltungszyklen	1 Fokusgruppeninterview 2 partizipative Fallstudien	3 Publikationen (s. Kapitel B-1)
BDD-Wiki	1 Fokusgruppeninterview 1 Literaturrecherche	Software-Engineering 1 Fokusgruppeninterview 1 kontrolliertes Experiment 1 Gestaltungszyklus	1 Fokusgruppeninterview 1 Funktionstest	4 Publikationen (s. Kapitel B-7 und B-8)
Funktionsreferenzmodell	1 partizipative Fallstudie 1 Fokusgruppeninterview 1 Produktevaluation	Referenzmodellierung 1 Literaturrecherche 3 Gestaltungszyklen	3 Fokusgruppeninterviews 1 Produktanalyse 1 partizipative Fallstudie	3 Publikationen (s. Kapitel B-9)

Tabelle A1-1: Gestaltungsprozesse der vorgestellten Artefakte im Überblick

Der Überblick zu den Gestaltungsprozessen dieser Dissertation zeigt eine wesentliche Charakteristik von Konsortialforschung: In den einzelnen Phasen sind Fachexperten

unterschiedlicher Unternehmen an dem Gestaltungsprozess eines Artefakts beteiligt. Diese multilaterale Zusammenarbeit zwischen Forschung und Praxis [Österle/Otto 2010, S. 3] sichert den konstitutiven Praxisbezug der Forschung [Ulrich 1984, S. 179] und gewährleistet wissenschaftliche Stringenz durch die Möglichkeit, Artefakte in mehreren Zyklen zu gestalten und zu evaluieren [Simon 1996, S. 28f, Hevner et al. 2004, S. 88-90].

1.4 Aufbau der Arbeit

Die vorliegende kumulative Dissertation gliedert sich in zwei Teile. Teil A motiviert CDQM als grundlegende Forschungsdomäne, erläutert Konsortialforschung als angewendete Forschungsmethode und ordnet Grundlagen und Forschungsergebnisse in das Metamodell des BE-Methodenkerns ein. Teil B umfasst neun Publikationen, die insgesamt vier Artefakte als Forschungsergebnisse der Dissertation vorstellen. Abbildung A1-2 zeigt den Aufbau der Arbeit im Überblick.

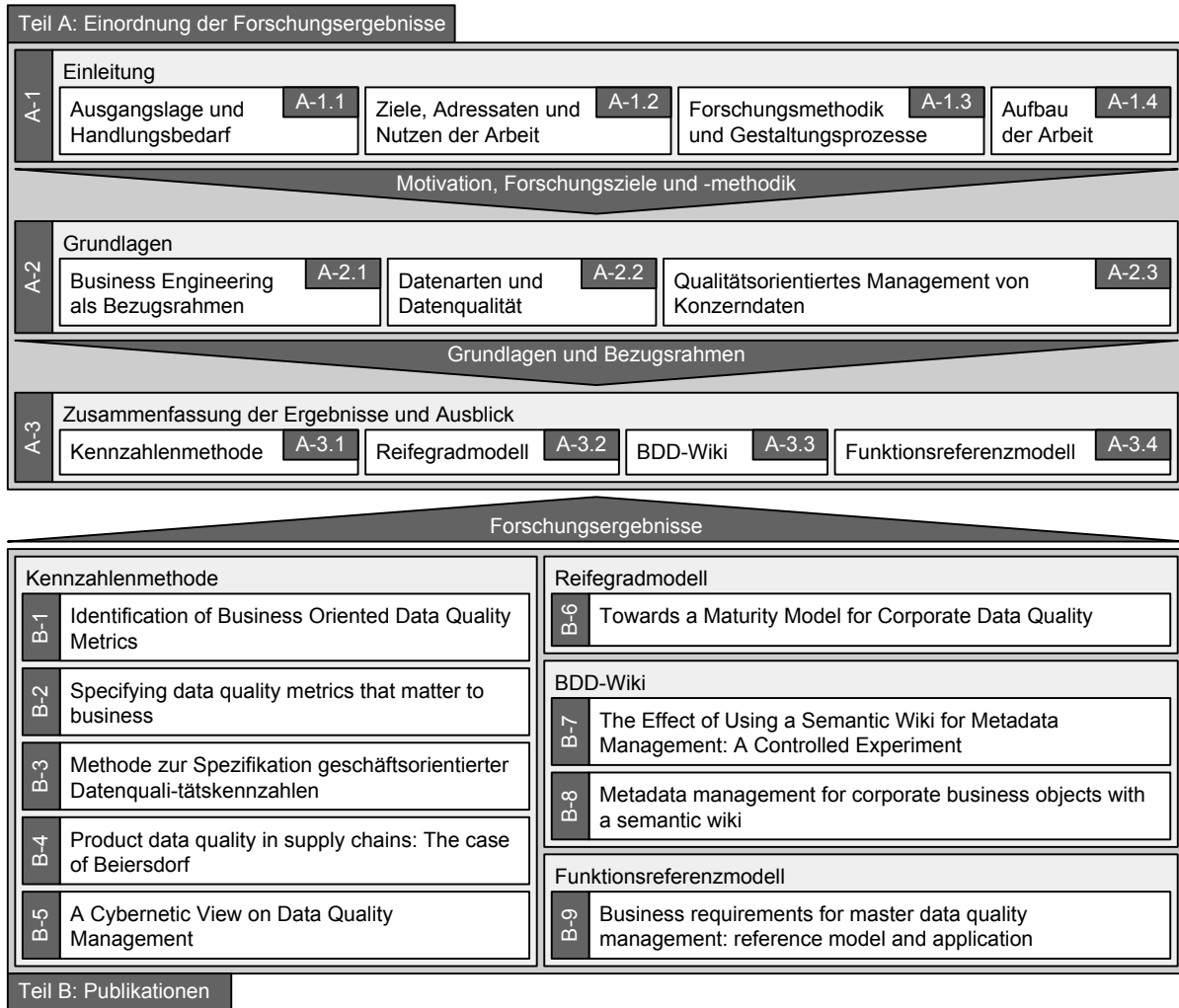


Abbildung A1-2: Aufbau der Dissertation

Den inhaltlichen Schwerpunkt der Dissertation bilden die Publikationen in Teil B. Sie stellen die einzelnen Artefakte und ihre Gestaltungsprozesse vor und diskutieren ihren

praktischen und wissenschaftlichen Nutzen. Da jede Publikation als eigenständiger wissenschaftlicher Beitrag das jeweils vorgestellte Forschungsergebnis motiviert, erklärt und diskutiert, enthält Teil B insbesondere an Grundlagen redundante Inhalte. Zudem verwenden die einzelnen Publikationen aufgrund ihrer über drei Jahre verteilten Entstehung keine einheitliche Terminologie. Teil A ordnet daher die Forschungsergebnisse in das Metamodell des BE-Methodenkerns ein und versucht durch diesen Bezugsrahmen, die Zusammenhänge zwischen den einzelnen Artefakten zu verdeutlichen und ein einheitliches Verständnis der Konzepte zu gewährleisten. Für ein homogenes Erscheinungsbild der Dissertation gibt Teil B die Publikationen nicht in der Originalformatierung, sondern in einem einheitlichen Format mit einem einheitlichen Zitationsstil und fortlaufend nummerierten Abbildungen und Tabellen wieder. Jeder Publikation ist ausserdem eine Tabelle mit bibliografischen Angaben vorangestellt. Die Verzeichnisse enthalten Angaben für die gesamte Dissertation, d. h. gemeinsam für Teil A und B.

2 Grundlagen

Die vorgestellten Artefakte nutzen das Metamodell des Methodenkerns des BE [Österle et al. 2007, S. 191-194] als begrifflichen Bezugsrahmen. Abschnitt 2.1 gibt daher einen kurzen Überblick zu BE und dem Metamodell des BE-Methodenkerns. Abschnitt 2.2 definiert die von der Dissertation fokussierten Datenobjekte als Konzerndaten und erläutert das der Arbeit zugrunde liegende Datenqualitätsverständnis. Abschnitt 2.3 definiert CDQM als Forschungsdomäne der Dissertation. Für den Stand der Forschung zu den einzelnen Artefakten sei auf die Publikationen in Teil B verwiesen.

2.1 Business Engineering als Bezugsrahmen

Der St. Galler Ansatz des BE [Österle 1995, S. 1-12, Österle/Blessing 2003, S. 80-84, Österle/Blessing 2005, S. 12-15] konzeptualisiert Transformationsprojekte von Unternehmen auf den Modellierungsebenen Strategie, Prozess und System durch Methoden [Heym/Österle 1993, Gutzwiller 1994, Nuseibeh et al. 1996]. Der BE-Methodenkern konsolidiert ähnliche Konzepte verschiedener BE-Ansätze und spezifiziert Gestaltungsobjekte der Ergebnisse einer Methodenanwendung in einem Metamodell [Österle et al. 2007, S. 192-194, Höning 2009, S. 96-116, 252-257]. Abbildung A2-1 zeigt die bei der Einordnung der vorgestellten Artefakte betrachteten Konzepte des BE-Methodenkerns.

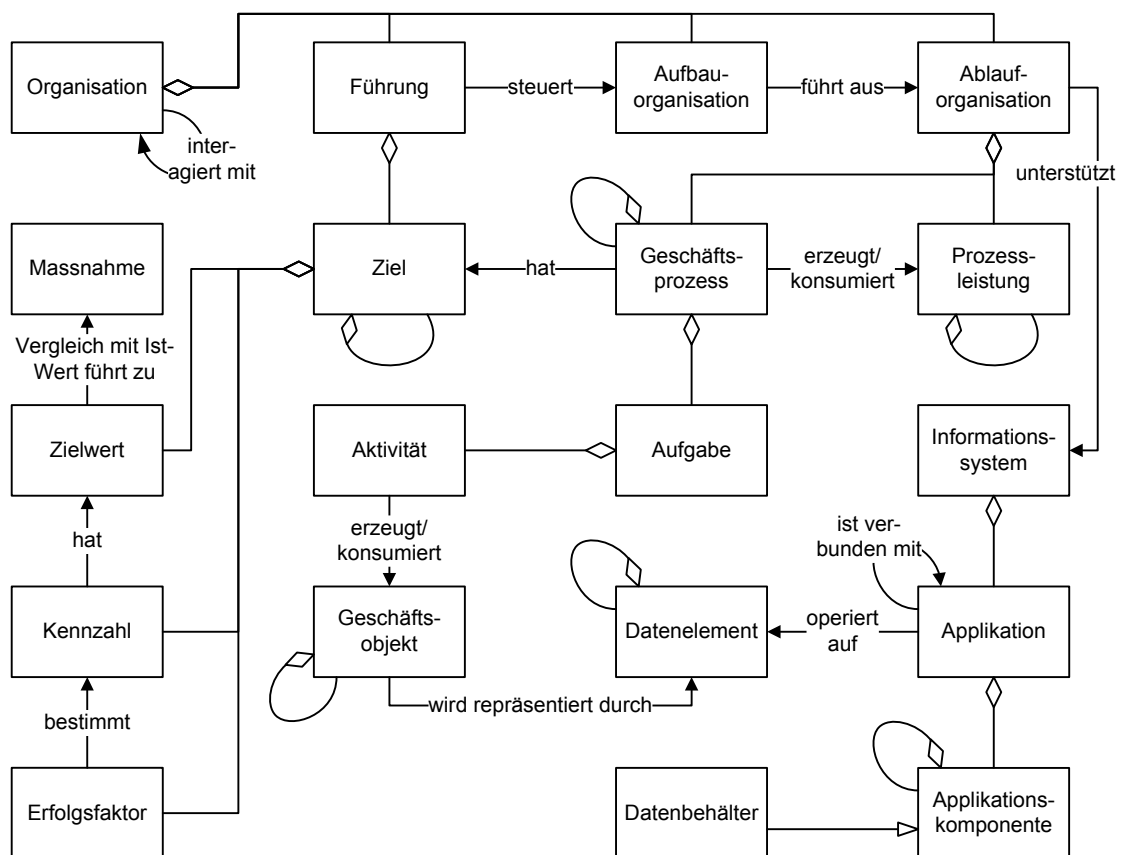


Abbildung A2-1: Konzepte des BE-Methodenkerns [Höning 2009, S. 252-257]

Die Metamodellierung verwendet die Metadatenarchitektur Object Management Facility (MOF) [OMG 2006] der Object Management Group (OMG) als Bezugsrahmen und ordnet das Metamodell des BE-Methodenkerns auf Ebene 2 (M2, zwei Ebenen über M0, der Ebene von Realweltobjekten [OMG 2009, S. 14-19]) ein. Die Darstellungen der Metamodellentitäten und -relationen verwenden Klassendiagramme der Unified Modeling Language [OMG 2006, S. 23-142] als Modellnotation, da sich diese als Notationsstandard für Metamodelle in Praxis und Wissenschaft etabliert haben.

Das Metamodell einer Methode zur Stammdatenintegration [Schmidt 2010, S. 101, 221-225] differenziert Typen und Instanzen von Datenelement und Geschäftsobjekten und ordnet ausserdem Metadaten als Konzept in den BE-Methodenkern ein. Abbildung A2-2 zeigt einen Ausschnitt des Metamodells der Methode. Das Präfix *BE::* kennzeichnet Metaentitätstypen und Relationen des BE-Methodenkerns, ergänzte Metaentitätstypen sind grau eingefärbt. Die Erweiterung des BE-Methodenkerns durch die Methode zur Stammdatenintegration ermöglicht eine differenziertere Einordnung der Konzepte der vorgestellten Artefakte (vgl. Kapitel 3).

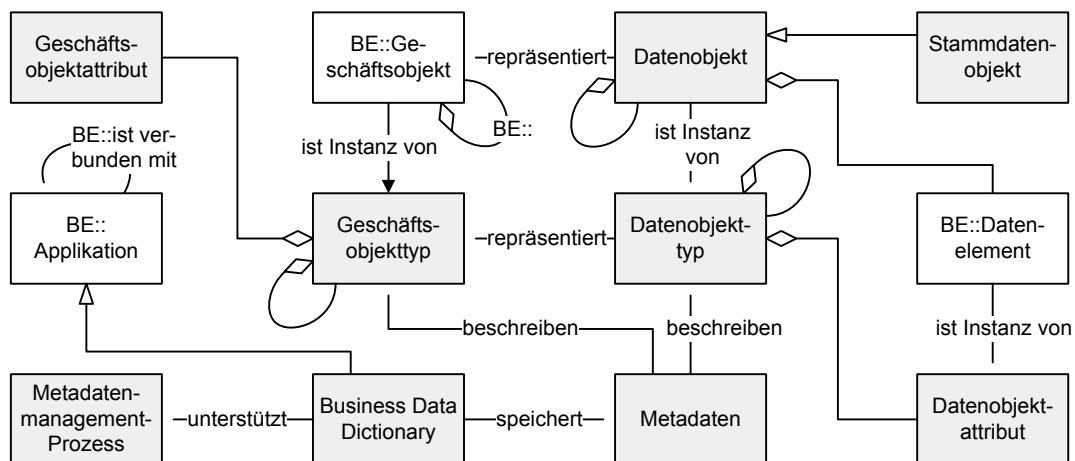


Abbildung A2-2: Erweiterung des BE-Methodenkerns [Schmidt 2010, S. 101]

2.2 Datenarten und Datenqualität

Dieser Abschnitt definiert das der Dissertation zugrundeliegende Verständnis von Daten als Repräsentation von Geschäftsobjekten und von Datenqualität als Erfüllungsgrad kontextspezifischer und subjektiver Anforderungen an Daten. Abschnitt 2.2.2 nimmt ausserdem eine Definition der von der Dissertation fokussierten Datenobjekte als Konzerndaten vor.

2.2.1 Daten als Repräsentation von Geschäftsobjekten

Praxis und Wissenschaft unterscheiden die Konzepte *Daten* und *Information* [Bourdreau/Couillard 1999, S. 25f, English 1999, S. 52, Tuomi 1999, S. 45-48, Boissot/Canals 2004, S. 45-62]. Eine übliche Differenzierung ist, dass Information durch die subjektive und kontextspezifische Interpretation von Daten (Repräsentation objek-

tiver Fakten) entsteht [Price/Shanks 2005, S. 89], die Grundlage unternehmerischer Entscheidungen ist und in Geschäftsprozessen verarbeitet wird [Mertens et al. 2005, S. 53, Stahlknecht/Hasenkamp 2005, S. 9f]. Diese Arbeit betrachtet jedoch den Zweck von Daten- oder Informationsverarbeitung, die Ablauforganisation eines Unternehmens bei der Leistungserstellung zu unterstützen (vgl. Abbildung A2-1). Ob die für die technische Verarbeitung erforderliche Repräsentation von Geschäftsobjekten als Daten oder als Information bezeichnet wird, erscheint dabei unerheblich. Ebenso wie andere Arbeiten mit ähnlichem Fokus [Wang 1998, S. 59, Pipino et al. 2002, S. 212, McKinney Jr./Yoos 2010, S. 331] unterscheidet die Dissertation daher die beiden Konzepte nicht.

Die Arbeit folgt der Modellierung des BE-Methodenkerns, der in seinem Metamodell die Repräsentation von Geschäftsobjekten durch das Konzept *Datenelement* beschreibt (vgl. Abbildung A2-1). Das Metamodell der Methode zur Stammdatenintegration differenziert dieses Verständnis durch weitere Konzepte: Datenobjektattribute repräsentieren als Typ von Datenelementen zweckorientiert gemeinsame Eigenschaften von Geschäftsobjekten (vgl. Abbildung A2-2). Für den Zweck einer telefonischen Kundenakquise ist beispielsweise die Repräsentation der Namen und Telefonnummern (Geschäftsobjektattribute) potentieller Kunden (Geschäftsobjekte) durch zwei Zeichenfolgen (Datenobjektattribute) ausreichend. Für den Versand personalisierter Anschreiben per Post sollte jedoch der Datenobjekttyp, der die potentiellen Kunden repräsentiert, um Datenobjektattribute zur Repräsentation des Geschlechts (z. B. durch eine Zeichenfolge als Anrede oder ein einzelnes Zeichen) und der Postanschrift ergänzt werden.

2.2.2 Konzerndaten

Konzerndatenobjekte (kurz: Konzerndaten²) einer Konzerndatenklasse (z. B. Kontodaten, Kundendaten, Produktdaten, Rechnungsdaten, Vertragsdaten) sind Datenobjekte,

- die Geschäftsobjekte (z. B. Konten, Kunden, Produkte, Rechnungen) und betriebliche Vorgänge (z. B. Adressänderung, Kontoüberweisung, Lieferung, Reservierung) eines Unternehmens repräsentieren, und
- die durch Datenobjekttypen und Metadaten für mehrere Applikationen, Geschäftsprozesse und Organisationseinheiten (meist unternehmens- oder konzernweit) einheitlich und unmissverständlich spezifiziert sind (oder sein sollten).

Abbildung A2-2 ordnet die Konzepte *Konzerndatenobjekt* und *Konzerndatenklasse* unter Berücksichtigung der vorgegebenen Erweiterungsregeln [Höning 2009, S. 219-227] in das erweiterte Metamodell des BE-Methodenkerns ein (vgl. Anhang C für detailliertere Beschreibung der Metaentitätstypen). Erweiterungen der Methode zur

² Der Bezeichner *Konzerndaten* soll eine Verwechslung mit dem Bezeichner *Unternehmensdaten* vermeiden, den andere Arbeiten [Mure/Zwick 2009, Plötner et al. 2010, pp. 221-230] für quantitative Eigenschaften von Unternehmen (z. B. Anzahl Mitarbeiter, Vorstandsgehälter, Umsatz) verwenden.

Stammdatenintegration [Schmidt 2010, S. 101, 221-225] kennzeichnet dabei das Präfix *AS::*:

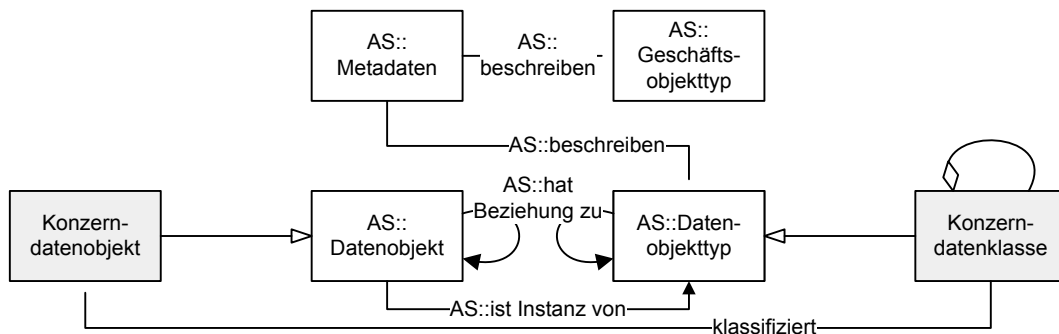


Abbildung A2-3: Konzerndaten im Metamodell des BE-Methodenkerns

Die Definition von Konzerndaten basiert auf den Erfahrungen der in Anhang B beschriebenen partizipativen Fallstudien und Fokusgruppeninterviews. Diese haben gezeigt, dass eine Differenzierung von Konzerndaten durch weitere Datenarten (z. B. Stammdaten, Bewegungsdaten) zwar eine Orientierungshilfe bei der Auswahl zu überwachender oder zu beschreibender Datenobjekte sein kann. Entscheidender ist aber die Wahrscheinlichkeit, dass bestimmte Datenobjekte geschäftskritische Probleme verursachen. Und die genannten Fallstudien und Interviews zeigen, dass Fachexperten bei dieser Bewertung insbesondere die konzernweite Nutzung von Datenobjekten berücksichtigen, und weniger die Zugehörigkeit der Datenobjekte zu einer bestimmten Datenart (vgl. Kapitel B-4).

Dennoch ist eine Differenzierung von Datenarten grundsätzlich möglich und kann z. B. die Definition eines Betrachtungsbereichs für die Spezifikation von DQ-Kennzahlen oder für die Nutzung eines BDDs unterstützen. Aus betriebswirtschaftlicher Sicht lassen sich die von einem Unternehmen bei der Leistungserstellung genutzten Datenobjekte anhand der Eigenschaften *Änderungshäufigkeit*, *Grad der existenziellen Unabhängigkeit* und *Volumenkonstanz* in die vier Kategorien *Änderungsdaten*, *Bestandsdaten*, *Bewegungsdaten* und *Stammdaten* unterteilen [Wedekind 2001, S. 72, Hansen/Neumann 2005, S. 8f, Schemm 2008, S. 21f, Schmidt 2010, S. 17-19].

- *Stammdaten*. Zustandsorientierte Datenobjekte mit relativ konstantem Volumen und geringer Änderungshäufigkeit [Loshin 2008, S. 8, Mertens 2009, S. 21]. Sie repräsentieren die Kerngeschäftsobjekte eines Unternehmens und haben einen eindeutigen Bezeichner [White et al. 2006, S. 63, Smith/Potter 2009, S. 2]. Datenelemente, die typischerweise als Stammdaten bezeichnet werden, sind Adressen, Alter und Namen von Kunden oder Abmessungen, Gewicht und Stücklistenpositionen von Produkten.
- *Bewegungsdaten*. Abwicklungsorientierte Datenobjekte, die betriebswirtschaftliche Vorgänge repräsentieren und sich auf Stammdaten beziehen [Dreibelbis et al. 2008, S. 35]. Die Positionen einer Rechnung, die verkaufte Produkte

(Stammdaten) auflisten, werden beispielsweise als Bewegungsdaten bezeichnet. Sie haben keinen konkreten Zeitbezug und eine begrenzte Lebensdauer [Loos 1999, S. 229]. Da jeder IS-unterstützte Geschäftsvorgang neue Bewegungsdaten erzeugt, nimmt ihr Volumen im Zeitverlauf zu.

- *Bestandsdaten*. Zustandsorientierte Datenobjekte, die die betriebliche Werte- und Mengenstruktur repräsentieren (z. B. Kontostände, Lagerbestände) [Hansen/Neumann 2005]. Ihr Volumen ist ähnlich konstant wie das Volumen von Stammdaten.
- *Änderungsdaten*. Abwicklungsorientierte Datenobjekte, die ausschliesslich Stammdaten verändern (z. B. eine Adressänderung) [Stahlknecht/Hasenkamp 2005, S. 138]. Ihr Volumen nimmt im Zeitverlauf zu, aber weniger als das Volumen von Bewegungsdaten.

2.2.3 Datenqualität

Datenqualität ist Gegenstand von Forschungsarbeiten, die das Phänomen durch Datenqualitätsdimensionen (DQ-Dimensionen) konzeptualisieren (vgl. Kapitel B-3 für einen Überblick). Ein Beispiel ist die Definition von 15 Dimensionen (u. a. Aktualität, Objektivität, Vollständigkeit) zur Beschreibung der von Datennutzern in einem bestimmten Nutzungskontext subjektiv wahrgenommenen Qualität genutzter Daten [Wang/Strong 1996, S. 13-20]. Diese Liste ist das Ergebnis einer zweistufigen Befragung von Datennutzern (1. Identifikation möglicher Dimensionen, 2. Bewertung der gesammelten Dimensionen). Die Autoren fassen die Aspekte Kontextabhängigkeit (Nutzungskontext) und Subjektivität (Datennutzer) in der Definition von Datenqualität als „the data’s fitness for use“ (org. „data that are fit for use by data consumers“) zusammen [Wang/Strong 1996, S. 6] und prägen damit das auch dieser Dissertation zugrunde liegende Datenqualitätsverständnis. Erfahrungen aus Berufs- und Beratungspraxis führen zu ähnlichen Dimensionslisten [Redman 1996, S. 245-266, English 1999, S. 87-118, Loshin 2001, S. 102-123].

Andere Arbeiten stellen DQ-Dimensionen als Ergebnis ontologischer [Wang/Wang 1996] und semiotischer [Price/Shanks 2005] Schlussfolgerungen vor. Ein Datendefekt (org. data deficiency) ist dabei eine Abweichung zwischen der unmittelbaren Wahrnehmung der Realität durch einen Datennutzer und dessen indirekter Wahrnehmung der Realität durch die Betrachtung eines IS (hier synonym zu Daten), das die Realität repräsentiert [Wang/Wang 1996, S. 88]. Die Autoren differenzieren vier Defekttypen (u. a. mehrdeutige Repräsentation, unvollständige Repräsentation) und definieren für jeden Typ eine DQ-Dimension (u. a. eindeutig, vollständig). Die semiotische Konzeptualisierung von Datenqualität [Price/Shanks 2005, S. 91-93] unterscheidet die folgenden drei Kategorien von DQ-Dimensionen basierend auf der Differenzierung von Syntaktik, Semantik und Pragmatik [Morris 1938].

- *Syntaktische Datenqualität.* Beschreibt die Konformität gespeicherter Datenelemente und deren gespeicherter Spezifikation durch Metadaten (z. B. zulässiger Wertebereich für ein Gewicht).
- *Semantische Datenqualität.* Beschreibt die Qualität der Abbildung von Realweltobjekten durch gespeicherte Datenelemente unter Berücksichtigung des Ziels der Objektabbildung (z. B. Unterstützung der Versandlogistik) und der Datenspeicherung (z. B. Genauigkeit eines gespeicherten Gewichts im Vergleich zum tatsächlichen Gewicht einer Shampoo-Flasche).
- *Pragmatische Datenqualität.* Beschreibt die Nutzbarkeit gespeicherter Datenelemente für einen bestimmten Anwendungsfall, wobei ein Anwendungsfall durch eine Aktivität (z. B. Rechnungsstellung), einen Kontext (z. B. Nutzung einer bestimmten Enterprise Resource Planning (ERP)-Applikation) und Eigenschaften des Datennutzers (z. B. neuer Mitarbeiter) konzeptualisiert wird.

Im Beispiel der Produktion von Shampoo-Flaschen bei Beiersdorf (vgl. Abschnitt 1.1) sind die Shampoo-Flaschen Realweltobjekte. Die Abweichung des im IS gespeicherten Flaschengewichts von dem tatsächlichen Gewicht einer Shampoo-Flasche ist ein semantischer Datendefekt [Price/Shanks 2005, S. 92f]. Denn die Betrachtung des IS (theoretisch berechnetes Flaschengewicht) führt dazu, dass ein falscher Zustand des Realwertobjekts (ein zu geringes Gewicht) angenommen wird [Wand/Wang 1996, S. 91f]. Im Raster der 15 DQ-Dimensionen ist der Defekt auf der Genauigkeitsdimension einzuordnen [Wand/Wang 1996, S. 14].

Die theoretische Betrachtung des Beispiels bestätigt die bereits einleitend beschriebene Feststellung, dass Datendefekte die Leistung von Geschäftsprozessen beeinträchtigen können: Im Beispiel müssen Paletten, deren Gewicht ein von Lager- oder Transportdienstleistern vorgegebenes Maximalgewicht überschreiten, umgepackt werden. Und bereits zugesicherte Lieferungen an Kunden verzögern sich. Datenqualität (im Beispiel die präzise Repräsentation des Produktgewichts durch das IS, konzeptualisiert durch die DQ-Dimension Genauigkeit) ist daher ein Erfolgsfaktor für die Leistungserstellung und Teil der Prozessleistung (vgl. Abbildung A2-4). Da Metadatendefekte (z. B. fehlende oder mehrdeutige fachliche Spezifikation des Produktgewichts) ebenfalls die Prozessleistung beeinträchtigen können, ist auch die Qualität von Metadaten zu bewerten und zu überwachen.

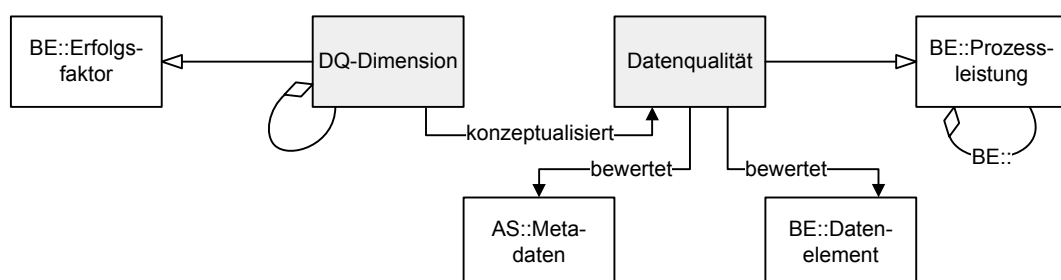


Abbildung A2-4: Datenqualität im Metamodell des BE-Methodenkerns

2.3 Qualitätsorientiertes Management von Konzerndaten

Der Ursprung eines unternehmensweit koordinierten Managements von Konzerndaten ist das Datenmanagement als administrative Aufgabe eines unternehmensweiten Informationsmanagements [Heinrich 2002, S. 21, Krcmar 2005, S. 35]. Datenmanagement umfasst dabei alle organisatorischen und technischen Aufgaben, die der Planung, Beschaffung, Organisation, Nutzung und Entsorgung von Daten im Unternehmen dienen [Dippold et al. 2005, S. 21f]. Zentrale Aufgaben sind die Definition einer Datenmanagementstrategie, die Festlegung der organisatorischen Verantwortung für Pflege und Erfassung der Daten sowie die Bereitstellung der Daten in exakt definierter und untereinander abgestimmter Form [Krcmar 2005, S. 111]. Weitere Arbeiten ergänzen u. a. die Führung und Überwachung des Datenmanagements sowie Metadatenmanagement [DAMA 2009, S. 6f] und definieren ähnliche Aufgaben für MDM als Ausprägung des Datenmanagements für Stammdaten [White et al. 2006, S. 2f, Schemm 2008, S. 20-25, Smith/Potter 2009, S. 64-68, Schmidt 2010, S. 19-21].

DQM umfasst Massnahmen zur Verbesserung der Qualität von Daten [Batini/Scannapieco 2006, S. 69-71], geht aber über die rein reaktive Verbesserung der Datenqualität (d. h. Identifikation und Bereinigung von Datendefekten) hinaus [Shankaranarayanan/Cai 2006, S. 303f]. Vielmehr beinhaltet DQM neben Massnahmen zur reaktiven Verbesserung von Datenqualität (d. h. der Korrektur von Datendefekten, z. B. dem Ergänzen einer fehlenden Kundentelefonnummer) präventive Massnahmen mit dem Ziel, Datendefekte zu verhindern und dadurch Datenqualität zu sichern [English 1999, S. 69-81, Eppler/Helfert 2004, S. 317]. Eine Übersicht zu DQM-Ansätzen geben Kapitel B-3 und weitere Arbeiten [Batini/Scannapieco 2006, S. 161-181, Batini et al. 2009, S. 16:3-16:31, Weber 2009, S. 24-36]. Die Unterscheidung präventiver und reaktiver Massnahmen bezieht sich auf den Zweck der jeweiligen Massnahme: Eine präventive Massnahme hat das Ziel, zukünftige Datendefekte zu verhindern oder die Wahrscheinlichkeit ihres Auftretens zu reduzieren. Reaktive Massnahmen beheben die Auswirkung bereits bestehender Datendefekte. Eine präventive Massnahme kann daher auch die Reaktion auf einen Datendefekt sein und das Ziel haben, ähnliche Defekte zukünftig zu verhindern.

Die Beziehung zwischen Datenmanagement und DQM beschreiben verschiedene Arbeiten unterschiedlich: Einige Autoren ordnen DQM als Aufgabenbereich des Datenmanagements ein [Dreibelbis et al. 2008, S. 123-125, DAMA 2009, S. 6]. Und andere Autoren beschreiben dem Datenmanagement zugeschriebene Aufgaben, wie die Festlegung der organisatorischen Verantwortung für Pflege und Erfassung von Daten [Krcmar 2005, S. 112-114], als präventives DQM [Redman 1996, S. 273-288, Wang et al. 1998, S. 102-104, English 1999, Weber et al. 2009b, S. 4:6f]. Die Dissertation versucht daher nicht, die eine oder andere Zuordnung zu motivieren, sondern fokussiert die Sicherung und Verbesserung der Qualität von Konzerndaten als gemeinsames Ziel beider Bereiche [Wang 1998, S. 60f, Krcmar 2005, S. 111, Loshin 2008, S. 17f]. Ab-

Abbildung A2-5 zeigt die Einordnung des CDQM-Prozesses als Unterstützungsprozess [Porter 1985, S. 37-48] in das Metamodell des BE-Methodenkerns. Datenqualität ist dabei die durch den CDQM-Prozess erzeugte Prozessleistung. Präventive und reaktive CDQM-Massnahmen, die je nach Zielerreichung des CDQM-Prozesses bei der Ausführung von CDQM-Aktivitäten angewendet werden (vgl. Abbildung A2-1), schützen oder verbessern die Qualität der Konzerndaten.

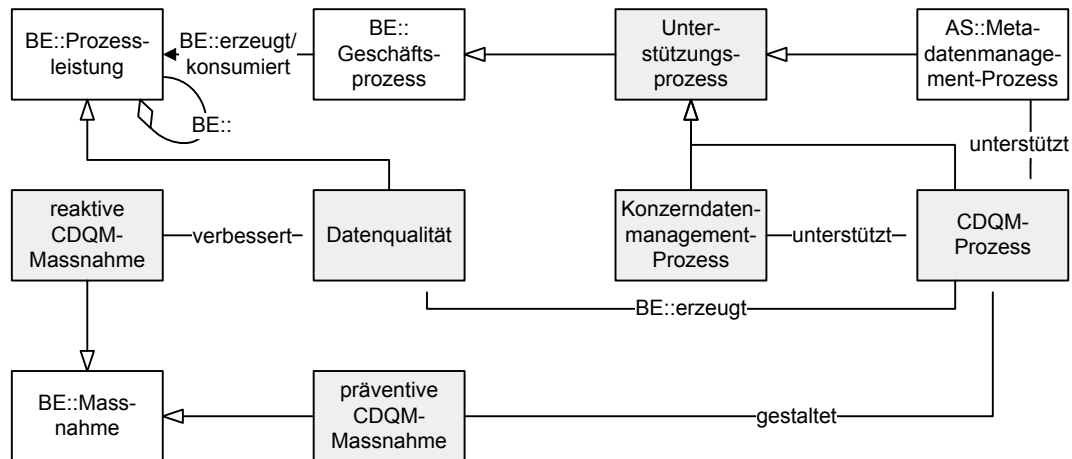


Abbildung A2-5: CDQM-Prozess im Metamodell des BE-Methodenkerns

Im Verlauf der Konsortialforschungsprojekte CC CDQ und CC CDQ2 (vgl. Abschnitt 1.3) ist ein Ordnungsrahmen gestaltet worden, der Aufgaben von CDQM durch 15 Erfolgsfaktoren und sechs Aufgabenbereiche strukturiert (vgl. Abbildung A2-6) [Otto et al. 2007, S. 919-923, Otto et al. 2008, S. 215-228, Hüner et al. 2009, S. 232]. Die CDQM-Erfolgsfaktoren (Spezialisierung des Metaentitätstyps *BE::Erfolgsfaktor*, vgl. Abschnitt 3.1) beschreiben dabei Strukturen zur Unterstützung von CDQM (z. B. DQ-Kennzahlen zur Überwachung geschäftskritischer Datendefekte). Für jeden Erfolgsfaktor sind ausserdem Aufgaben definiert, die den Aufbau der jeweils beschriebenen Strukturen unterstützen.

Abbildung A2-6 zeigt die sechs Aufgabenbereiche des CDQM-Ordnungsrahmens auf den drei Ebenen des St. Galler BE-Ansatzes (vgl. Abschnitt 2.1). Der gemeinsam mit der European Foundation for Quality Management (EFQM) standardisierte CDQM-Ordnungsrahmen [EFQM 2011] beschreibt detailliert die einzelnen Aufgaben und Erfolgsfaktoren. Der Ordnungsrahmen ist u. a. in Fokusgruppeninterviews und partizipativen Fallstudien im Rahmen des Gestaltungsprozesses des vorgestellten Reifegradmodells (vgl. Anhang B). Die Dissertation betrachtet den CDQM-Ordnungsrahmen daher als praktisch anwendbare und wissenschaftlich fundierte Strukturierung von CDQM.

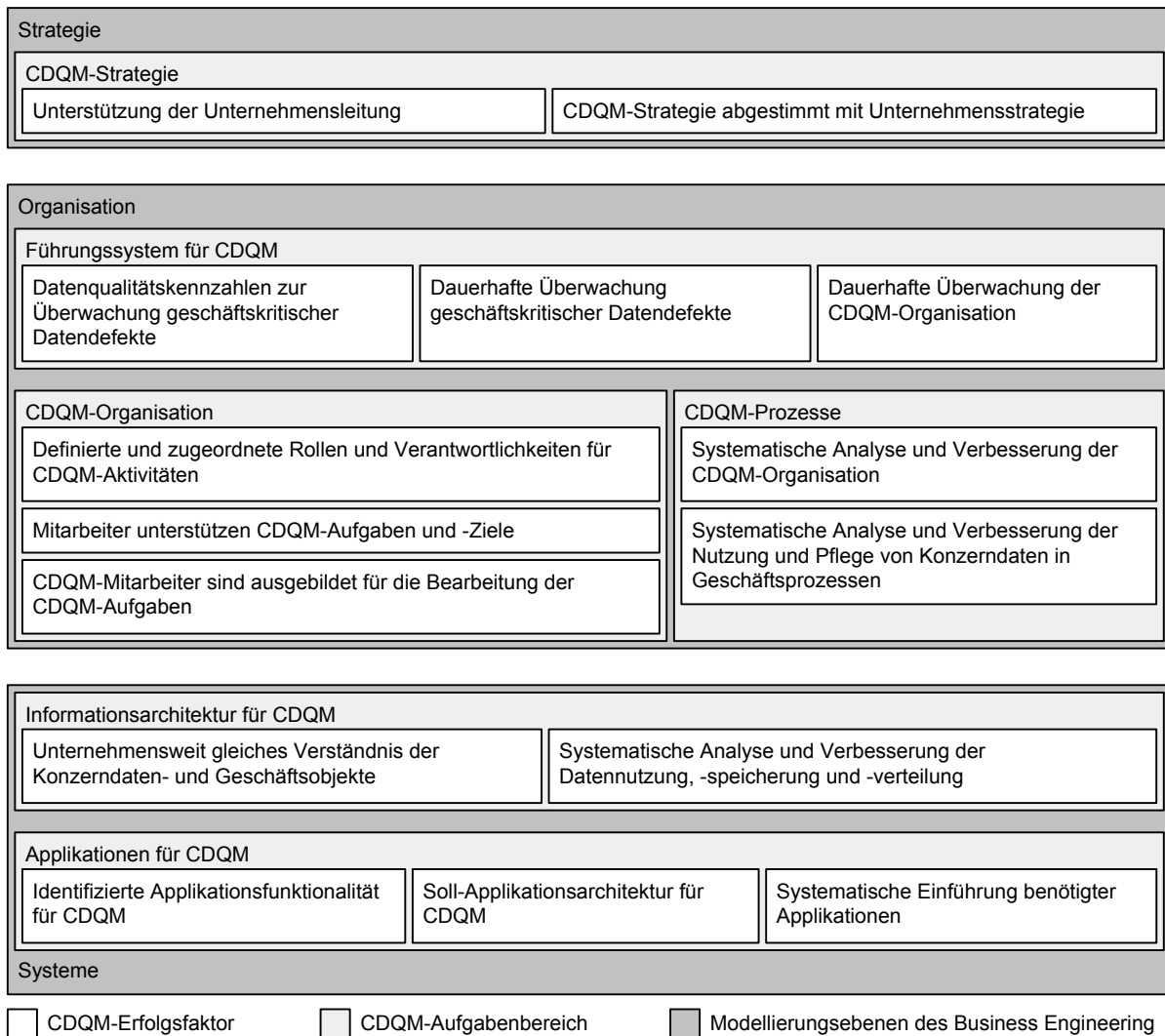


Abbildung A2-6: CDQM-Ordnungsrahmen mit Zielen nach EFQM [2011]

3 Ergebnisse und wirtschaftliche Bewertung

Dieses Kapitel erläutert den Zusammenhang der einzelnen Ergebnisse der Dissertation und ordnet die im Teil B der Arbeit vorgestellten Artefakte in den BE-Methodenkern ein. Das Metamodell des BE-Methodenkerns dient dabei als Bezugsrahmen, der ein einheitliches und zusammenhängendes Verständnis der Konzepte gewährleistet. Abschnitt 3.2 diskutiert ausserdem Aufwand und Nutzen der Artefaktanwendung.

3.1 Zusammenhang der Ergebnisse

Abbildung A3-1 zeigt die Einordnung der vorgestellten Artefakte in den BE-Methodenkern. Die Artefakte werden dabei mit Konzepten des Metamodells (vgl. Abschnitt 2.1), den in den Abschnitten 2.2 und 2.3 erläuterten Erweiterungen und weiteren, artefaktspezifischen Ergänzungen, beschrieben (vgl. Anhang C für eine Definition der erweiterten Metaentitätstypen).

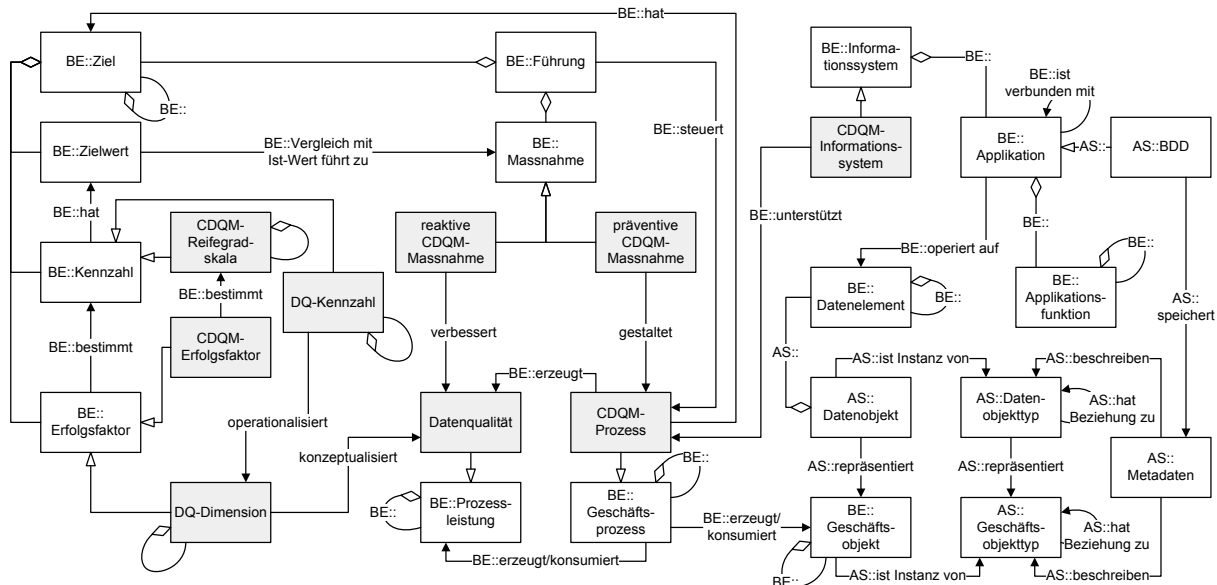


Abbildung A3-1: Konzepte der vorgestellten Artefakte im Überblick

Ein Führungssystem zur Steuerung der Qualität von Konzerndaten eines Unternehmens umfasst DQ-Kennzahlen und eine CDQM-Reifegradskala.

- Die Kennzahlen dienen der Überwachung der Konzerndatenqualität.
- Die Skala des Reifegradmodells ermöglicht eine Fortschrittskontrolle beim Aufbau von Strukturen (z. B. DQ-Kennzahlen) für CDQM sowie einen Vergleich mit CDQM-Funktionen anderer Unternehmen.

Neben Artefakten, die den Aufbau eines solchen Führungssystems in einem Unternehmen unterstützen, stellt die Dissertation zwei Artefakte für präventive CDQM-Massnahmen vor. Ziel der Anwendung dieser Massnahmen ist die Vermeidung von Datendefekten.

- Das BDD-Wiki unterstützt die unternehmensweit einheitliche Beschreibung von Geschäftsobjekten und versucht Datendefekte zu vermeiden, deren Ursache fehlendes oder unterschiedliches Verständnis von Geschäftsobjekten ist (vgl. Abschnitt 1.1, Beispiel der DB Netz).
- Die Funktionsarchitektur unterstützt die bedarfsorientierte Bereitstellung von IS für CDQM. Sie beschreibt Funktionen, die u. a. durch ein BDD bereitgestellt werden können.

Das zentrale Ergebnis der Dissertation ist die Kennzahlenmethode (vgl. Kapitel B-3). Sie unterstützt Unternehmen bei der Spezifikation von DQ-Kennzahlen zur Überwachung geschäftskritischer Datendefekte. Der Spezifikationsprozess umfasst (neben der Überprüfung und Dokumentation der Ergebnisse) zwei wesentliche Phasen: In der ersten Phase identifiziert der Konzern-Datensteward³ Datendefekte, die bei der Ausführung von Prozessaktivitäten Probleme verursachen, und bewertet die Auswirkung der Geschäftsprobleme auf die Ziele des Unternehmens. In der zweiten Phase spezifiziert er Validierungsregeln zur Überwachung der identifizierten Datendefekte. Die Validierungsregeln bilden das Messverfahren für die spezifizierten DQ-Kennzahlen.

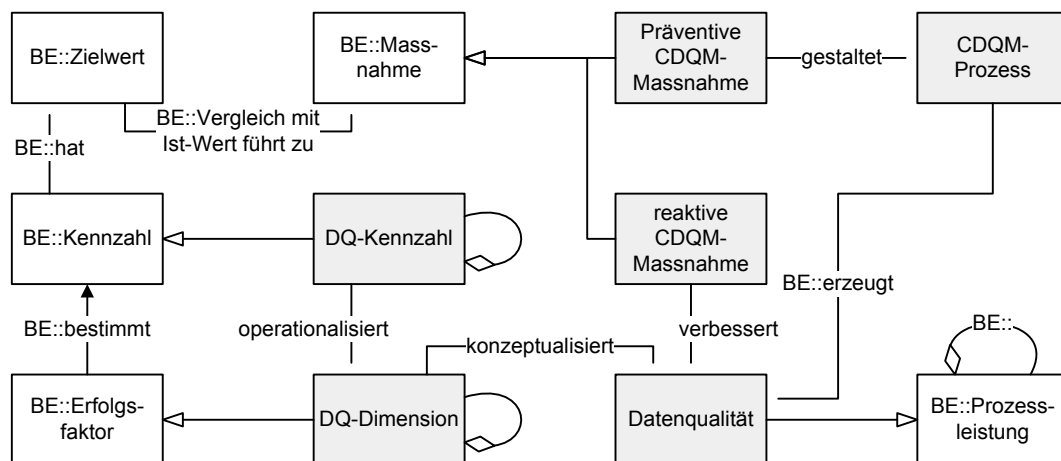


Abbildung A3-2: Konzepte der Kennzahlenmethode

Abbildung A3-2 zeigt den Metaentitätstyp für DQ-Kennzahlen im Kontext des BE-Methodenkerns. Die Unterscheidung von Konzeptualisierung und Operationalisierung entspricht dabei dem zuvor beschriebenen Vorgehen: Zuerst werden Datendefekte (beobachtetes Phänomen, z. B. fehlerhafte Adresdaten) mit DQ-Dimensionen (standardisierte Konzepte, z. B. Aktualität, Vollständigkeit) beschrieben. Das Ergebnis ist eine standardisierte Problembeschreibung (z. B. „die Adresse ist veraltet“, oder „bei der Adresse fehlt die Hausnummer“). Um die Defekte technisch messen zu können, operationalisieren DQ-Kennzahlen diese Beschreibung z. B. durch Validierungsregeln, deren Einhaltung ein IS überprüfen kann.

³ Konzern-Datensteward ist eine von der Methode definierte Rolle. Für die Beschreibung der Methode (vgl. Kapitel B3) ist der Konzern-Datensteward die Person, die die Anwendung der Methode leitet und die Zusammenarbeit der übrigen Rollen koordiniert.

Das Reifegradmodell nutzt den CDQM-Ordnungsrahmen (vgl. Abschnitt 2.3) zur Modellierung der CDQM-Funktion eines Unternehmens. Die im Ordnungsrahmen beschriebenen Erfolgsfaktoren für CDQM bestimmten dabei die Bewertungsskala des Reifegradmodells (vgl. Abbildung A3-3). Die Erreichung von Zielen für die Neugestaltung oder die Umstrukturierung einer CDQM-Funktion (z. B. die Spezifikation von DQ-Kennzahlen) kann auf dieser Skala bewertet werden.

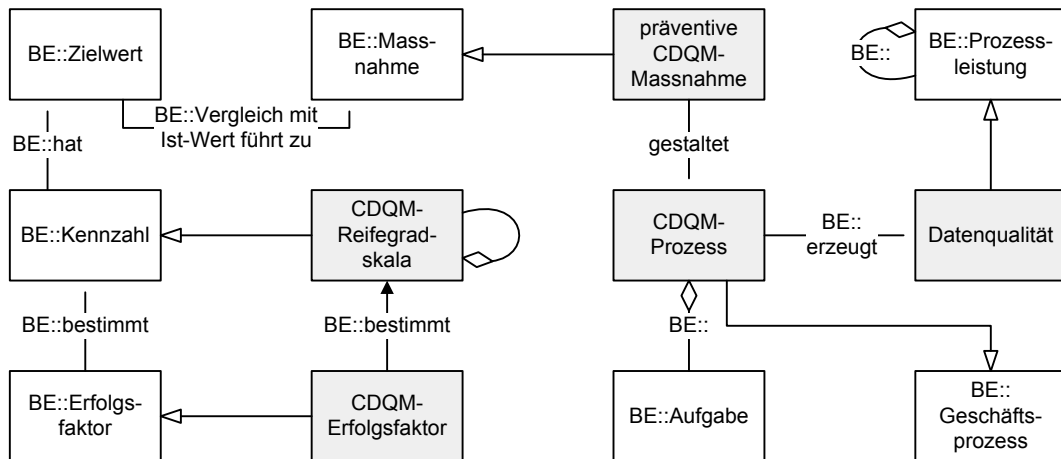


Abbildung A3-3: Konzepte des Reifegradmodells

Der CDQM-Ordnungsrahmen definiert neben der Spezifikation und Nutzung von DQ-Kennzahlen weitere Erfolgsfaktoren für CDQM, u. a. die unternehmensweit einheitliche Beschreibung von Geschäftsobjekten und die regelmässige Überprüfung der für CDQM angebotenen Applikationsfunktionen. Die beiden im Folgenden beschriebenen Artefakte unterstützen Massnahmen, deren regelmässige Durchführung diese Erfolgsfaktoren sichert.

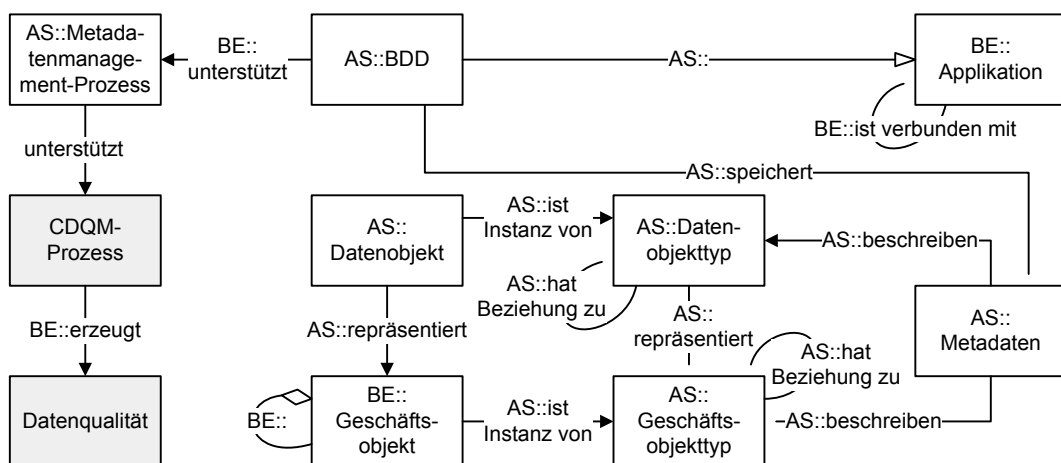


Abbildung A3-4: Konzepte des BDD-Wikis

Ein BDD ist ein fachlicher Metadatenkatalog und dient der unternehmensweit einheitlichen Beschreibung von Geschäftsobjekt- und Konzerndatenobjekttypen aus einer fachlichen Perspektive [Schmidt/Otto 2008, S. 215-217, Schmidt 2010, S. 31f]. Es unterstützt als Applikation den Metadatenmanagement-Prozess und somit den CDQM-Prozess eines Unternehmens (vgl. Abbildung A3-4). Fachliche Metadaten informieren

Datennutzer aus Fachbereichen über Bedeutung, Herkunft, Struktur und Qualität der von ihnen verwendeten Daten- und Geschäftsobjekte [Tozer 1999, S. 123-125, Foshay et al. 2007, S. 72, Inmon et al. 2008, S. 12-16]. Ziel der Nutzung eines BDD ist ein unternehmensweit einheitliches Verständnis von Geschäfts- und Konzerndatenobjekten. Die Kapitel B-7 und B-8 erläutern die Nutzung eines Wikis als BDD und diskutieren Vor- und Nachteile.

Das Funktionsreferenzmodell beschreibt 72 fachliche Applikationsfunktionen eines IS zur Unterstützung von CDQM (vgl. Abbildung A3-5). Die Funktionen bilden ein Raster für die Evaluation von Software-Produkten oder den Vergleich von CDQM-IS verschiedener Unternehmen. Das Modell unterstützt ausserdem die Bewertung der Bereitstellung benötigter Applikationsfunktionen für CDQM sowie die Einführung neuer Software.

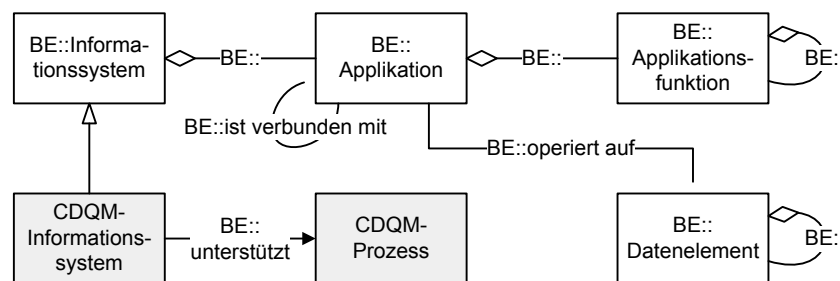


Abbildung A3-5: Konzepte des Funktionsreferenzmodells

3.2 Aufwands- und Nutzenbetrachtung

Dieser Abschnitt stellt Aufwand und Nutzen der Anwendung der vier Artefakte gegenüber. Grundlage der Bewertung ist die Nutzung der Artefakte in partizipativen Fallstudien (vgl. Teil B und Anhang B). Das Ziel dieser Gegenüberstellung ist nicht eine allgemeingültige Nutzenbewertung, sondern vielmehr die Unterstützung möglicher Nutzer der Artefakte bei der Planung von CDQM-Massnahmen. Daher wird der Gestaltungsaufwand der Artefakte nicht berücksichtigt, sondern lediglich der Aufwand für die Nutzung der Artefakte.

Der Aufwand für die Artefaktanwendung wird in Tagen und Kosten (kalkuliert über angenommene Tagessätze) angegeben. Die Aufstellung berücksichtigt keine monetären Aufwände wie z. B. Kosten für Software-Lizenzen oder -Betrieb, da diese stark von den Gegebenheiten der jeweiligen Unternehmen abhängen. Zudem sind für keines der Artefakte besondere Ressourcen bereitzustellen. Nutzenpotenziale der jeweiligen Artefaktanwendung werden qualitativ beschrieben und in die Nutzenkategorien Kostensenkung, Qualitätsverbesserung und Zeitersparnis eingeordnet. Beispiele erläutern ausserdem, wie Unternehmen den Kosten der Artefaktanwendung einen monetären Nutzen gegenübergestellt haben.

3.2.1 Kennzahlenmethode

Tabelle A3-1 zeigt exemplarisch den Aufwand der Methodenanwendung in einer partizipativen Fallstudie (vgl. Kapitel B-4). Das Ergebnis der Methodenanwendung umfasst sieben DQ-Kennzahlen mit insgesamt 32 Validierungsregeln zur Überwachung von Produktdaten. Die technische Implementierung dieser Regeln und einer geeigneten Darstellung für Messergebnisse ist noch nicht abgeschlossen, der Aufwand wird auf ca. 50 Tage geschätzt. Das Projektteam bestand aus drei internen Mitarbeitern der CDQM-Funktion und einem externen Mitarbeiter. An sieben Workshops zur Identifikation geschäftskritischer Datendefekte haben zusätzlich insgesamt 16 interne Mitarbeiter verschiedener Fachbereiche teilgenommen.

Aktivität	interne Mitarbeiter	externe Mitarbeiter
Vorbereitung (Aktivitäten I.1 und I.2)	2 Workshops: 3 x 2 Tage	Vorbereitung : 1 Tag 2 Workshops: 2 Tage
Workshops (I.3)	7 Workshops: 37 x ½ Tag	Vorbereitung: 1 Tag 7 Workshops: 7 x ½ Tag
Spezifikation (II.1 und II.2)	2 Workshops: 3 x 2 Tage	Analyse: 2 Tage 2 Workshops: 2 Tage
Dokumentation (III.1 und III.2)	Prüfung: 1 x ½ Tag	Dokumentation: 2 Tage
Summe Tage	31 Tage	13 ½ Tage
Summe Kosten⁴	31'000 CHF	27'000 CHF

Tabelle A3-1: Anwendungsaufwand der Kennzahlenmethode

Workshops mit Mitarbeitern unterschiedlicher Fachbereiche verursachen zwar hohen Koordinationsaufwand und insgesamt hohen zeitlichen Aufwand durch die Teilnahme mehrerer Personen, sie führen aber auch zu einer im Vergleich zu Einzelinterviews umfassenderen Identifikation von Datendefekten. Daher sollten mindestens ein fachbereichsübergreifender Workshop durchgeführt oder mehrere Personen eines Fachbereichs gemeinsam befragt werden.

Ein Konsumgüterhersteller hat im Winter 2009 Produkte aus einem deutschen Werk per Lastwagen nach Russland geliefert. Die im Verkauf flüssigen Produkte kamen dort in gefrorenem Zustand an, da für den Transport bei dem beauftragten Logistikdienstleister kein Thermotransport beantragt wurde. Die Ware wurde reklamiert und musste neu ausgeliefert werden. Der Schaden wird vom Unternehmen auf ca. 100'000 EUR geschätzt. Bei der Analyse dieses Geschäftsproblems im Verlauf der Anwendung der Kennzahlenmethode hat das Unternehmen festgestellt, dass Toleranzen für Transporttemperaturen vieler Produkte nicht gepflegt sind. Eine Kennzahl zur Überwachung der Vollständigkeit des zugehörigen Datenelements soll helfen, zukünftig ähnliche Geschäftsprobleme zu vermeiden.

⁴ Annahmen für die Kostenkalkulation: 2'000 CHF pro Tag für einen externen Mitarbeiter, 1'000 CHF pro Tag für einen internen Mitarbeiter.

Das Unternehmen hat das geschilderte Beispiel erfolgreich verwendet, um den Nutzen von DQ-Messungen zu erläutern. Allerdings ist bei diesem und ähnlichen Beispielen stets zu überlegen, welcher Anteil der zukünftig potentiell verhinderten Geschäftsprobleme tatsächlich als Nutzen der Kennzahlenmethode veranschlagt werden kann. Denn nicht jede fehlende Toleranz für Transporttemperaturen verursacht automatisch einen Schaden von 100'000 EUR. Trotzdem kann die Erläuterung eines Geschäftsproblems, dessen Schaden in etwa den Kosten der Anwendung der Kennzahlenmethode entspricht, ein Argument für die Spezifikation geschäftsorientierter DQ-Kennzahlen sein.

Grundsätzlich ist bei der Nutzenbewertung der Kennzahlenmethode der Nutzen der spezifizierten Kennzahlen und der Nutzen der Methode zu unterscheiden. Fachexperten haben DQ-Kennzahlen in einem Fokusgruppeninterview (vgl. Kapitel B-2) als Voraussetzung für CDQM beschrieben, da ohne eine dauerhafte Überwachung von Datenqualität effektives CDQM nicht möglich sei. DQ-Kennzahlen haben nach dieser Einschätzung die folgenden allgemeinen Nutzenpotenziale:

- *Kostensenkung.* Die Anwendung von DQ-Kennzahlen führt zur Identifikation defekter Datenobjekte. So ist es möglich, reaktive CDQM-Massnahmen nur für tatsächlich defekte Datenobjekte anzuwenden.
- *Qualitätsverbesserung.* Messwerte einheitlich spezifizierter DQ-Kennzahlen, deren Messverfahren ein IS wiederholt ausführen kann, sind vergleichbar und ermöglichen aussagekräftigere Analysen als Ergebnisse von Befragungen oder Stichproben.
- *Zeitersparnis.* Das technisch ausführbare Messverfahren der DQ-Kennzahlen ermöglicht die Identifikation von Datendefekten durch ein IS und ersetzt manuelle Analysen.

Die Anwendung der Kennzahlenmethode sichert insbesondere den Geschäftsbezug der spezifizierten DQ-Kennzahlen und die Möglichkeit, das spezifizierte Messverfahren technisch zu implementieren und von einem IS ausführen zu lassen (vgl. *Qualitätsverbesserung*). Die Kennzahlenmode hat ausserdem die folgenden Nutzenpotentiale:

- *Kostensenkung.* Die Analyse der durch Datendefekte verursachten Geschäftsprobleme ermöglicht eine quantitative Schätzung des monetären Schadens von Datendefekten (vgl. Kapitel B-5). Entlang dieser Kausalketten ist die Argumentation eines monetären Nutzens von CDQM-Massnahmen möglich. Allerdings ist zu beachten, dass die durch eine Korrektur oder Vermeidung von Datendefekten erzielte Kostensenkung keine direkte Folge der Methodenanwendung ist.
- *Qualitätsverbesserung.* Das Vorgehensmodell der Kennzahlenmethode sichert die Identifikation kritischer Geschäftsprobleme vor der Spezifikation von Validierungsregeln zur Überwachung der zugehörigen Datendefekte. Diese Top-

Down-Analyse (vgl. Kapitel B3) reduziert das Risiko, Datendefekte nur aufgrund ihrer Messbarkeit auszuwählen. Eine mögliche Fehlsteuerung des CDQM-Prozesses durch die Konzentration von CDQM-Massnahmen nur auf überwachte Datendefekte ist jedoch zu beachten. Eine regelmässige Wiederholung der Identifikation kritischer Geschäftsprobleme (Aktivitäten I bis III der Kennzahlenmethode, vgl. Kapitel B3) kann dieses Risiko reduzieren.

- *Zeitersparnis.* Der zeitliche Aufwand der Methodenanwendung ist nicht unerheblich (vgl. Tabelle A3-1). Die methodische Dokumentation aller Teilergebnisse (z. B. Anforderungen, Kausalketten) reduziert jedoch den Aufwand bei Anpassungen oder Erweiterungen des Kennzahlensystems.

3.2.2 Reifegradmodell

Tabelle A3-2 zeigt exemplarisch den Aufwand einer Reifegradbewertung in einem Workshop. Um Bewertungen verschiedener Unternehmensfunktionen vergleichen zu können, ist die Aktivität *Workshop* mehrfach durchzuführen mit entsprechend erhöhtem Aufwand. Im Beispiel führt ein externer Mitarbeiter die Bewertung durch und stimmt den Betrachtungsbereich und die Ziele der Bewertung mit einem internen Mitarbeiter ab. Am Workshop für die Reifegradbewertung nehmen weitere vier interne Mitarbeiter teil.

Aktivität	interne Mitarbeiter	externe Mitarbeiter
Vorbereitung	1 Workshop: 1 x 1 Tag	1 Workshop: 1 Tag
Workshops	1 Workshop: 5 x ½ Tag	Vorbereitung: 1 Tag 1 Workshop: ½ Tag
Analyse		Analyse: 2 Tag
Dokumentation	Prüfung: 1 x ½ Tag	Dokumentation: 2 Tag
Summe	4 Tage	6 ½ Tage
Summe Kosten	4'000 CHF	13'000 CHF

Tabelle A3-2: Anwendungsaufwand des Reifegradmodells

Verschiedene Energieversorger haben im Anschluss an Reifegradbewertungen in den einzelnen Unternehmen auf einem gemeinsamen Workshop ihre Ansätze für DQ-Messungen diskutiert. Für den zugehörigen Aufgabebereich (Führungssystem für CDQM) ist in allen Unternehmen ein niedriger Reifegrad identifiziert worden. Ein Unternehmen hat DQ-Kennzahlen für Kundendaten vorgestellt und die Fokusgruppe zur Prüfung und Ergänzung des Ansatzes genutzt. Die übrigen Unternehmen können die vorgestellten Kennzahlen als Basis für eigene DQ-Messungen nutzen (vgl. 58'000 CHF für die Anwendung der Kennzahlenmethode, Abschnitt 3.2.1).

Das Beispiel zeigt, wie der durch eine standardisierte Reifegradbewertung mögliche Erfahrungsaustausch genutzt werden kann, um CDQM-Strukturen anderer Unterneh-

men zu nutzen und dadurch Kosten zu sparen. Grundsätzlich lassen sich folgende Nutzenpotenziale einer Reifegradbewertung unterscheiden:

- *Qualitätsverbesserung.* Die Bewertung der Strukturen einer CDQM-Funktion auf Basis eines bereits mehrfach erfolgreich angewendeten Domänenmodells (CDQM-Ordnungsrahmen) erhöht im Vergleich zu selbstdefinierten Kriterien die Wahrscheinlichkeit, kritische Handlungsbedarfe (z. B. das Fehlen von DQ-Kennzahlen) zu identifizieren. Der Vergleich mit anderen Unternehmen ermöglicht ausserdem die Identifikation von Zielwerten für einzelne Erfolgsfaktoren. Unternehmen können diese als Richtwerte nutzen, um den Aufwand für nötige Verbesserungen eigener CDQM-Strukturen besser abschätzen zu können.
- *Zeitersparnis.* Das Reifegradmodell umfasst neben der Reifegradskala auch Empfehlungen für mögliche CDQM-Massnahmen (vgl. Aufgaben des CDQM-Ordnungsrahmens). Eine Reifegradbewertung vereinfacht daher die Auswahl geeigneter Massnahmen.

3.2.3 BDD-Wiki

Tabelle A3-3 zeigt exemplarisch den Aufwand zur Bereitstellung eines BDD-Wikis in einer partizipativen Fallstudie (vgl. Kapitel B-8). Das Projektteam bestand aus zwei internen Mitarbeitern der CDQM-Funktion und zwei externen Mitarbeitern (Konzeption, Implementierung). Fünf weitere interne Mitarbeiter haben die Modellierung von Nutzungs- und Pflegeprozessen in Interviews unterstützt. Das Unternehmen nutzte bereits vor Projektbeginn ein BDD, aus dem Metadaten übernommen werden konnten.

Aktivität	interne Mitarbeiter	externe Mitarbeiter
Planung	1 Workshop: 2 x 1 Tag	1 Workshop: 1 x 1 Tag
Modellierung Nutzungs- und Pflegeprozesse	5 Interviews: 5 x ½ Tag Prüfung: 1 x ½ Tag	Vorbereitung: 1 x 1 Tag 5 Interviews: 5 x ½ Tag Modellierung: 1 x 2 Tage
Modellierung Metadatentypen	Prüfung: 1 x ½ Tag	Modellierung: 1 x 1 Tag
Anpassung und Erweiterung benötigter Funktionen		Spezifikation: 1 x 2 Tage Implementierung: 1 x 6 Tage
Migration vorhandener Metadaten		Migration: 1 x 2 Tage
Funktionstest	5 Tests: 5 x ½ Tag	Vorbereitung: 1 x 1 Tag Analyse: 1 x 1 Tag
Summe	8 Tage	19 ½ Tage
Summe Kosten	8'000 CHF	39'000 CHF

Tabelle A3-3: Aufwand zur Bereitstellung eines BDD-Wikis

Der Aufwand für die technische Einführung und den Betrieb eines BDD-Wikis ist nicht höher als für vergleichbare Applikation zur Definition und Pflege fachlicher Me-

tadaten. Die Applikation wurde auf Basis einer Standard-Software (MediaWiki⁵, Wiki-Plattform von Wikipedia⁶) entwickelt, für deren Betrieb standardisierte Applikationen sowie Einführungs- und Betriebsunterstützung externer Dienstleister verfügbar sind. Die für die weltweite Zusammenarbeit von Menschen auch mit geringer technischer Ausbildung entwickelte Wiki-Plattform vereinfacht zudem die fachliche Einführung.

Der Pflegeaufwand fachlicher Metadaten ist allerdings ein kritischer Faktor. Beschreibungen der korrekten Verwendung von Datenobjekten oder Handlungsanweisungen für bekannte Fehlerfälle sind nur zwei Beispiele fachlicher Metadaten, die manuell zu pflegen und zwischen mehreren Fachbereichen abzustimmen sind. Diese regelmässig durchzuführenden Tätigkeiten kosten zwar Zeit, sind aber auch kritisch für den Erfolg eines BDD. Die Nutzung eines Wikis kann den hohen Pflegeaufwand zwar nicht reduzieren, jedoch für den einzelnen Mitarbeiter akzeptabel erscheinen lassen. Denn der zu erwartende Kollaborationseffekt, d. h. die Mitarbeit von vielen Mitarbeitern des Unternehmens bei der Pflege fachlicher Metadaten, verteilt den Aufwand über das gesamte Unternehmen und belastet nicht nur einzelne Mitarbeiter. Durch die verstärkte Zusammenarbeit sind ausserdem Qualitätsverbesserungen (Nutzenpotenziale, s. u.) zu erwarten. Für Unternehmen, in denen fachliche Metadaten nicht bereits vorhanden sind, ist zusätzlich der Aufwand zur initialen Identifikation der einheitlich zu beschreibenden Daten- und Geschäftsobjekte zu berücksichtigen [Schmidt 2010, S. 203-205].

Das Telekommunikationsunternehmen Deutsche Telekom AG nutzt ein Wiki zur Pflege fachlicher Metadaten (fachlogisches Datenmodell, Business Object Model [Schmidt et al. 2010, S. 22f]). Fachexperten des Unternehmens schätzen die durch die Verwendung eines unternehmensweit einheitlichen Datenmodells möglichen Einsparungen auf ca. 0.8% des IT-Budgets [Schmidt 2010, S. 207-09, Schmidt et al. 2010, S. 30f]. Bei einem Anteil des IT-Budgets am Unternehmensumsatz von 4.6% [Smith/Potter 2009, S. 31] entspricht dies ca. 24 Mio. EUR.

Die Nutzenpotenziale des BDD-Wikis sind grundsätzlich vergleichbar mit den Nutzenpotenzialen der bereits genannten Methode zur Stammdatenintegration [Schmidt 2010, S. 205-210]. Hinzu kommen mögliche Qualitätsverbesserungen durch kollaborative Pflege.

- *Qualitätsverbesserung.* Die Nutzung einer Kollaborationsplattform für die Pflege fachlicher Metadaten ermöglicht die Beteiligung vieler Mitarbeiter am Pflegeprozess. Dies ermöglicht im Vergleich zur Pflege durch einzelne Mitarbeiter umfassendere Beschreibungen und frühzeitige Korrekturen falscher Metadaten. Die verstärkte Kommunikation zwischen Mitarbeitern unterschiedlicher Abteilungen, die im Verlauf der gemeinschaftlichen Pflege zu erwarten

⁵ Website zu MediaWiki: <http://www.mediawiki.org>

⁶ Website der deutschen Wikipedia: <http://de.wikipedia.org>

ist, verbessert zudem das Verständnis von Geschäftsobjekten. Denn Diskussionen über die Bedeutung einzelner Geschäftsobjekte oder Änderungen einer Beschreibung aufgrund neuer, bisher nicht berücksichtigter Anforderungen anderer Abteilungen bleiben stärker in Erinnerung als eine zentral bereitgestellte Definition. Dieser Effekt führt zwar nicht unmittelbar zu einer Verbesserung der Metadatenqualität, er verstärkt jedoch den grundsätzlichen Nutzen eines BDD.

- *Zeitersparnis.* Die zentrale Bereitstellung fachlicher Metadaten reduziert den Rechercheaufwand, falls (neue) Mitarbeiter Informationen zu Geschäftsobjekten (z. B. bei der Pflege logistischer Daten) benötigen. Die Nutzung einer Kollaborationsplattform für die Pflege von Metadaten verteilt zudem den Pflegeaufwand auf mehrere Personen. Dabei ist jedoch auch der Aufwand zu berücksichtigen, den die Nutzung einer neuen Applikation verursacht. Alle Nutzer müssen sich mit der Handhabung des Werkzeugs befassen und Zeit für die Pflege von Metadaten einplanen.

Zwar konnten bisher Einsparpotenziale wie die Reduzierung von Mehrarbeit oder von Fehlern nicht quantifiziert werden. In Fokusgruppeninterviews zur Bewertung der Methode und des BDD-Wikis (vgl. Anhang B) haben Fachexperten aber bestätigt, dass ein einheitliches Verständnis von Daten- und Geschäftsobjekten hilft, diese Potenziale zu realisieren.

3.2.4 Funktionsreferenzmodell

Tabelle A3-4 zeigt exemplarisch den Aufwand zur Nutzung des Funktionsreferenzmodells in einer partizipativen Fallstudie (vgl. Kapitel B-9). Das Ergebnis der Anwendung ist eine Bewertung angebotener Applikationsfunktionen und die Identifikation zusätzlich benötigter Funktionalität. Das Projektteam bestand aus drei internen Mitarbeitern der CDQM-Funktion, zwei internen Mitarbeitern der IT-Organisation und einem externen Mitarbeiter.

Aktivität	interne Mitarbeiter	externe Mitarbeiter
Vorbereitung	1 Workshop: 1 x ½ Tag	Vorbereitung: 1 Tag 1 Workshop: ½ Tag
Workshops	2 Workshops: 10 x ½ Tag	Vorbereitung: 1 Tag 2 Workshops: 2 x ½ Tag
Analyse		Analyse: 2 Tage
Dokumentation	Prüfung: 1 x ½ Tag	Dokumentation: 2 Tage
Summe	6 Tage	7 ½ Tage
Summe Kosten	6'000 CHF	15'000 CHF

Tabelle A3-4: Anwendungsaufwand des Funktionsreferenzmodells

Bei der bisher einzigen Anwendung des Funktionsreferenzmodells (vgl. Kapitel B-9) ist keine monetäre Nutzenbewertung durchgeführt worden. Primäres Ziel der Anwendung war der Erfahrungsaustausch mit anderen Unternehmen. Ob die Ergebnisse dieses Austauschs zur Reduzierung von Software-Lizenzen oder zur Verbesserung bestimmter Funktionalität (Fokus des Austauschs sind Archivierung und Vergabe von GTINs) führen, ist abzuwarten. Grundsätzlich lassen sich aber folgende Nutzenpotenziale des Funktionsreferenzmodells unterscheiden:

- *Kostensenkung*. Die Analyse der Ist-Aufnahme (Ergebnis der Bewertung angebotener Applikationsfunktionen) kann unnötige Software-Lizenzen und Konsolidierungspotenziale der Applikationslandschaft aufzeigen.
- *Qualitätsverbesserung*. Die Bewertung angebotener Applikationsfunktionen ermöglicht die gezielte Verbesserung häufig genutzter oder unzureichend bereitgestellter Funktionen. Zudem verbessert die gemeinsame Analyse von Applikationsfunktionalität die Kommunikation zwischen Fachbereichen und IT-Organisation.
- *Zeitersparnis*. Das Funktionsreferenzmodells vereinfacht den Austausch von Wissen zwischen Unternehmen und den Vergleich angebotener Software, da es Funktionen eindeutig benennt und ordnet.

4 Zusammenfassung und Ausblick

Dieses Kapitel fasst die Erkenntnisse der Beiträge aus Teil B zusammen und diskutiert Einschränkungen für die Interpretation dieser Erkenntnisse. Abschnitt 4.3 gibt ausserdem einen Ausblick auf weiteren Forschungsbedarf.

4.1 Erkenntnisse

Die Anwendung der Kennzahlenmethode zeigt, dass die Operationalisierung unternehmens- und problemspezifisch identifizierter DQ-Dimensionen durch Validierungsregeln die Spezifikation fachlich nutzbarer DQ-Kennzahlen ermöglicht. In partizipativen Fallstudien hat die Anwendung der Methode mit akzeptablem Aufwand zu nützlichen Ergebnissen geführt (vgl. Kapitel B-4 und Anhang B). Die Modellierung von CDQM-Massnahmen in einem kybernetischen System (vgl. Kapitel B-5) zeigt ausserdem, dass die in Abbildung A3-2 dargestellte Konzeptualisierung auch Kosten-Nutzen-Analysen ermöglicht. Die Simulation von CDQM-Szenarien, d. h. von Annahmen zu Datendefektkosten, zu Kosten von CDQM-Massnahmen und zur Wirkung der Massnahmen, unterstützt dabei die Investitionsplanung für CDQM.

Im Verlauf der Konsortialforschungsprojekte CC CDQ und CC CDQ2 (vgl. Abschnitt 1.3) sind Inhalt und Struktur des CDQM-Ordnungsrahmens in Fokusgruppeninterviews und partizipativen Fallstudien evaluiert und genutzt worden (vgl. Anhang B). Dabei hat sich gezeigt, dass der Ordnungsrahmen aus Sicht von Fachexperten die wesentlichen Aufgaben einer CDQM-Funktion beschreibt und für Aufbau und Umstrukturierung einer CDQM-Funktion nützliche Orientierung bietet. Fachexperten und Forscher haben bei der Anwendung und Evaluation des Ordnungsrahmens keine Änderungsbedarfe identifizieren können. Der CDQM-Ordnungsrahmen ist daher eine adäquate und ausreichend stabile Repräsentation der Domäne CDQM.

Die Nutzung des CDQM-Ordnungsrahmens als Domänenmodell für Reifegradbewertungen (vgl. Anhang B) zeigt, dass die Strukturierung der Aufgaben in sechs Bereiche und die Definition von 15 Erfolgsfaktoren (vgl. Abbildung A2-6) auch für diesen Zweck geeignet ist. Es ist möglich, unterschiedliche Unternehmensstrukturen und Aufgabenbereiche auf die Struktur des Ordnungsrahmens abzubilden und anhand eines Fragenkatalogs bewerten zu lassen.

Die Identifikation von Anforderungen an ein BDD zeigt, dass Kollaboration bei der Pflege fachlicher Metadaten und Integration der Metadaten in fachlich genutzte Applikationen Erfolgsfaktoren für fachliches Metadatenmanagement sind (vgl. Kapitel B-8). Dies legt die Nutzung eines Wikis als BDD nahe, da Wikis allgemein Kollaboration fördern und auch Unternehmen zunehmend Wikis zur Unterstützung kollaborationsintensiver Prozesse nutzen [Wagner/Majchrzak 2007, S. 22-30, Bughin et al. 2009, S. 14].

Das Ergebnis eines Experiments (vgl. Kapitel B-7) zeigt, dass die Navigations- und Suchfunktionalität semantischer Wikis die Nutzung fachlicher Metadaten im Vergleich zu einem normalen Wiki besser unterstützt. Das Experiment bestätigt ausserdem, dass ein semantisches Wiki im betrieblichen Kontext für fachliches Metadatenmanagement ohne Einführungs- oder Schulungsmassnahmen eingesetzt werden kann. Szenario-Tests in einer partizipativen Fallstudie und Fokusgruppeninterviews (vgl. Anhang B) bestätigen weiter, dass das BDD-Wiki die Anforderungen an ein BDD erfüllt und für fachliches Metadatenmanagement genutzt werden kann.

Die Anwendung des Funktionsreferenzmodells (vgl. Gestaltungsprozess in Anhang B) zeigt, dass das Modell die für CDQM benötigte Funktionalität betrieblich genutzter IS adäquat repräsentiert. So haben Software-Hersteller die Beschreibung ihrer Produkte mit den Konzepten des Modells geprüft und akzeptiert. Ausserdem hat Beiersdorf das Modell für eine Analyse seiner Applikationslandschaft genutzt und plant die Einführung neuer Applikationen zur Unterstützung von CDQM auf der Basis dieser Analyse (vgl. Kapitel B-9). Beide Anwendungen zeigen, dass in der Praxis nicht eine dedizierte Applikation die für CDQM benötigte Funktionalität bereitstellt, sondern mehrere, ursprünglich nicht für DQM oder MDM vorgesehene Applikationen. So werden seitens der Software-Hersteller u. a. Customer Relationship Management (CRM)-, DSS- oder ERP-Applikationen für einzelne Funktionen angegeben, oder auch Middleware-Komponenten oder Modellierungswerkzeuge [Otto/Hüner 2009, S. 46-57]. Diese Erkenntnis unterstützt die in Abschnitt 2.3 beschriebene Modellierung eines CDQM-Prozesses als Unterstützungsprozess, der auch Aufgaben anderer Unterstützungsprozesse (z. B. Konzerndatenmanagement, Metadatenmanagement) umfasst.

4.2 Einschränkungen

Die Evaluation aller Artefakte durch ihre Anwendung in partizipativen Fallstudien erfordert grundsätzlich die Einschränkung der Erkenntnisse auf die Unternehmen, die die Artefakte angewendet haben. Auch die positive Nutzenbewertung in Fokusgruppeninterviews ist als subjektive Bewertung einzelner Fachexperten zu betrachten. Die Anwendung der Artefakte in mehreren Unternehmen aus unterschiedlichen Industriesektoren sowie die Gestaltungsprozesse mit bis zu vier Gestaltungs- und Evaluationszyklen (vgl. Tabelle A1-1 und Anhang B) lassen allerdings vermuten, dass die Artefakte auch in weiteren Fällen effektiv anwendbar sind.

Bei der Methodenanwendung ist zu beachten, dass DQ-Kennzahlen der beschriebenen Struktur nicht ausschliesslich durch Anwendung der Kennzahlenmethode spezifiziert werden können. Die Methode ist lediglich als eine Möglichkeit für die Spezifikation geschäftsorientierter DQ-Kennzahlen zu sehen. So hat Bayer CropScience beispielsweise vergleichbare Regeln für DQ-Messungen im Verlauf eines Projekts zur Harmonisierung von Geschäftsprozessen definiert [Ebner et al. 2011, S. 10-12]. Der Auslöser waren Defekte der Produkthierarchie (fehlende und falsche Zuordnungen von Produk-

ten), die zu Problemen bei der konzernweiten Konsolidierung von Wirkstoffbedarfen, einem zentralen Ziel des gesamten Harmonisierungsprojekts, geführt haben. Bayer CropScience hat seit dem Beginn der DQ-Messungen (2008) die zur Überwachung der Produkthierarchie definierten sieben Regeln um Regeln für weitere Bereiche (z. B. Beschaffungsart, Bewertungsklasse) ergänzt und überwacht mittlerweile (2011) ca. 80 Regeln für Materialdaten. Die Überwachung von Kunden- und Lieferantendaten ist ebenfalls geplant.

Die Interpretation eines quantitativen Reifegrads, der auf Basis einer qualitativen Beantwortung eines Fragenkatalogs berechnet wird, ist grundsätzlich wegen der Verwendung unterschiedlicher Skalenniveaus einzuschränken. Um Vergleiche und arithmetische Operationen zu ermöglichen, sollte der Reifegrad als quantitatives Mass auf einer Verhältnisskala (z. B. auf einer Prozentskala) abgebildet werden. Um mit dem vorgestellten Reifegradmodell vergleichende Bewertungen zu ermöglichen, sind die Fragen und die Bewertungsskala gemeinsam mit der EFQM standardisiert worden und folgen einer auch in anderen Domänen erprobten Bewertungslogik [EFQM 2011]. Ein für alle Fragen gleiches Bewertungsraster gewährleistet dabei näherungsweise äquidistante Bewertungen. Bei der Interpretation einer Reifegradmessung ist die Subjektivität der Bewertungen aber stets zu beachten. Die Bewertung durch externe, geschulte Assessorinnen kann die Objektivität der Messung erhöhen und insbesondere die Vergleichbarkeit von Reifegraden verschiedener Unternehmen gewährleisten.

Die Eigenschaft semantischer Wikis, die Nutzung fachlicher Metadaten im Vergleich zu einem normalen Wiki besser zu unterstützen, ist durch die angewendete Forschungsmethodik eingeschränkt. Das durchgeführte Experiment belegt den Effekt zwar statistisch signifikant, allerdings nur für das im Experiment modellierte Anwendungsszenario und die genutzte Technologie. Da insbesondere die Funktionalität der Benutzerschnittstelle die subjektiv zu bewertende Benutzerfreundlichkeit beeinflusst, kann die Erkenntnis nicht ohne weitere Untersuchungen auf andere Anwendungsszenarien und Wiki-Plattformen übertragen werden.

Der in Fokusgruppeninterviews und einer partizipativen Fallstudie (vgl. Anhang B) bestätigte Nutzen des BDD-Wikis ist ebenfalls nicht ohne Einschränkungen auch in anderen Fällen zu erwarten. Die genannten allgemeinen Einsparpotenziale unternehmensweit einheitlich beschriebener Geschäftsobjekte sind zwar bei der erfolgreichen Nutzung eines BDD zu erwarten, allerdings ist ihnen der hohe Pflegeaufwand für fachliche Metadaten und das damit verbundene Risiko veralteter und unvollständiger Beschreibungen gegenüberzustellen.

Die Senkung des Pflegeaufwands durch die Nutzung des BDD-Wikis wurde in Szenariotests im Rahmen einer partizipativen Fallstudie bestätigt (vgl. Kapitel B-8). Dieser Effekt alleine wird allerdings nicht ausreichen, um im betrieblichen Alltag dauerhaft die Pflege fachlicher Metadaten zu gewährleisten. Weitere, insbesondere organisatori-

sche Massnahmen, sollten die Nutzung des BDD-Wikis unterstützen. Die Definition von Verantwortlichkeiten für einzelne Datenobjekt- und Geschäftsobjekttypen oder die Definition von Metadatenpflege als Mitarbeiterziel sind nur zwei Beispiele solcher Massnahmen.

Der Fokus des Funktionsreferenzmodells auf die fachkonzeptionelle Funktionssicht eines CDQM-IS schränkt dessen Nutzung auf die beschriebenen und ähnliche Anwendungsfälle (Analyse, Bebauungsplanung, Kommunikation) ein. Für die Gestaltung und Entwicklung neuer CDQM-IS ist das Modell nicht geeignet, es fehlen insbesondere die Ebene der Datenverarbeitung sowie die Organisations- und Steuerungssicht.

4.3 Weiterer Forschungsbedarf

Zwar sind bei der Spezifikation von DQ-Kennzahlen individuelle Charakteristika der jeweiligen Unternehmen zu berücksichtigen. Bereits spezifizierte Kennzahlen können aber als Vorlage dienen und den Aufwand des Spezifikationsprozesses reduzieren. Zudem können auch Kennzahlen zur Überwachung allgemeiner oder branchenspezifischer Datendefekte spezifiziert und von mehreren Unternehmen verwendet werden. Gesetze wie REACH oder das deutsche Telekommunikationsgesetz (vgl. Abschnitt 1.1) enthalten z. B. Vorgaben, deren Einhaltung durch Kennzahlen überwacht werden könnte. Die Ergänzung der Sammlung von Kausalketten und DQ-Kennzahlen (vgl. Kapitel B-3) ist daher ein Gestaltungsziel für weitere Forschung.

Weiteren Forschungsbedarf bietet ausserdem die Kosten-Nutzen-Analyse für CDQM-Massnahmen. Kapitel B-5 stellt hierzu zwar bereits ein kybernetisches Modell unter Verwendung des Konzepts geschäftsorientierter DQ-Kennzahlen vor. Für detailliertere Analysen ist das Modell allerdings z. B. durch eine Differenzierung von CDQM-Kostenarten zu erweitern. Neben diesem Gestaltungsziel bietet die Untersuchung der Effektivität von CDQM-Massnahmen Möglichkeiten für empirische Forschung. In Unternehmen, die DQ-Kennzahlen zur Messung von Datenqualität nutzen, kann die Wirkung von Bereinigungsmassnahmen oder verbesserter Datenpflegeprozesse auf Messwerte der DQ-Kennzahlen und auf die Leistung der jeweils betrachteten Geschäftsprozesse untersucht werden. Erkenntnisse zu Kosten und Effektivität einzelner Massnahmen wären für die Praxis interessante Forschungsergebnisse.

Ein weiteres Erkenntnisziel bietet die empirische Untersuchung geschäftsorientierter DQ-Kennzahlen. Ihre dauerhafte Nutzung sollte zu einer Qualitätsverbesserung der überwachten Daten und zu weniger Geschäftsproblemen führen, da das Ziel der Kennzahlennutzung die frühzeitige Erkennung und Vermeidung geschäftskritischer Datendefekte ist. Bei diesen Untersuchungen ist jedoch zu berücksichtigen, dass Datendefekte nicht die einzige Ursache von Geschäftsproblemen sind.

Für das Reifegradmodell besteht weiterer Forschungsbedarf insbesondere in der Sammlung und Analyse weiterer Reifegradbewertungen. Die Standardisierung des

Reifegradmodells gemeinsam mit der EFQM gewährleistet die Stabilität des Domänenmodells und ermöglicht vergleichbare Bewertungen von Unternehmen sowie mehrere Bewertungen gleicher Unternehmen über einen längeren Zeitraum. Die Analyse der Bewertungsergebnisse kann helfen, erfolgreiche Massnahmen und Vorgehensmodelle für den Aufbau von CDQM-Funktionen zu entwickeln oder industriespezifische Zielwerte für die einzelnen Aufgabenbereiche zu identifizieren.

Überwacht ein Unternehmen neben dem CDQM-Reifegrad auch die Qualität ausgewählter Datenobjekte mit DQ-Kennzahlen, sind vergleichende Analysen von Reifegrad und Datenqualität möglich. Diese Untersuchungen würden es ermöglichen, die Effektivität der durch den CDQM-Ordnungsrahmen vorgeschlagenen Massnahmen zu bewerten. Auch hier ist aber zu beachten, dass nicht nur CDQM-Massnahmen die Qualität von Konzerndaten beeinflussen.

Für die Weiterentwicklung des BDD-Wikis ist insbesondere die Integration fachlicher Metadaten in fachlich genutzte Applikationen wie CRM- oder ERP-Systeme ein Gestaltungsziel. Der bisher genutzte Prototyp bietet hierzu zwar bereits die notwendigen technischen Voraussetzungen, allerdings ist u. a. noch zu untersuchen, für welche fachlichen Aufgaben welche Metadaten bereitzustellen sind und in welchem Umfang Pflegefunktionalität des BDD-Wikis auch in fachlichen Anwendungen angeboten werden sollte. Die Gestaltung eines fachkonzeptionellen Referenzmodells für kollaboratives fachliches Metadatenmanagement könnte die technische Entwicklung des BDD-Wikis unterstützen. Ein solches Modell könnte ausserdem als eigenständiges Artefakt Handlungsempfehlungen für die Organisation von Unternehmensfunktionen geben, die für fachliches Metadatenmanagement zuständig sind.

Weiterer empirischer Forschungsbedarf besteht in der Untersuchung des Effekts offener Kollaboration (Wiki-Prinzip) auf die Qualität fachlicher Metadaten. Forscher haben bisher nur wenige Fälle dokumentiert, in denen Unternehmen Wikis für fachliche Spezifikationen nutzen. Ein positiver Effekt der Nutzung eines Wikis auf die Qualität der gemeinsam erstellten Spezifikationen (z. B. fachliche Beschreibung von Daten- und Geschäftsobjekten) ist dabei bisher nicht belegt worden. Gleichwohl lässt die teilweise hohe Qualität fachlicher Beschreibungen in öffentlichen Wikis einen solchen Effekt vermuten.

Eine weitere Möglichkeit für wissenschaftliche Untersuchungen bietet die Identifikation der Daten- und Geschäftsobjekte, für die Metadaten unternehmensweit einheitlich beschrieben werden sollten. Diesen Prozess unterstützt die Methode zur Stammdatenintegration [Schmidt/Otto 2008, S. 105-140], allerdings mit hohem Aufwand für die Identifikation (u. a. Analyse von Datenflüssen und Prozessdokumentationen) und initiale Beschreibung [Schmidt 2010, S. 203-205]. Die Unterstützung bereits dieser Aufgaben (und nicht nur der dann folgenden Pflegeaktivitäten) durch das BDD-Wiki könnte den Aufwand deutlich senken.

Zukünftige Forschungsarbeiten sollten ausserdem versuchen, das Funktionsreferenzmodell um weitere Ebenen und Sichten zu erweitern. So könnten auf Basis von Fallstudien Referenzstrukturen für CDQM-Funktionen (z. B. Führungsstrukturen, Verantwortlichkeiten) identifiziert werden und die Basis für Rollen- und Berechtigungskonzepte als Organisationssicht des Referenzmodells bilden. Ausserdem könnten Abhängigkeiten zwischen einzelnen Funktionen des Referenzmodells identifiziert und typische CDQM-Aufgaben und -Aktivitäten (z. B. Anpassung der Datenarchitektur, Identifikation und Entfernung von Dubletten, Messen von Datenqualität) als Referenzprozesse modelliert werden. Ziel dabei sollte allerdings nicht die Gestaltung und Entwicklung von CDQM-IS für die Bereitstellung aller im Referenzmodell beschriebenen Funktionen sein, denn die Anwendung des Modells zeigt, dass dies kaum möglich ist (vgl. Kapitel B-9). Die Entwicklung spezialisierter Applikationen (z. B. für DQ-Messungen, die Modellierung von Datenpflegeprozessen oder die Pflege fachlicher Metadaten) ist aber ein realistisches und von der Praxis gefordertes Ziel, soweit die Integration der Applikationen in bestehende Applikationsarchitekturen mit akzeptablem Aufwand möglich ist.

Teil B

1 Identification of business oriented data quality metrics

Titel	Identification of Business Oriented Data Quality Metrics
Autoren	Boris Otto; Kai M. Hüner; Hubert Österle Universität St. Gallen, Institut für Wirtschaftsinformatik Müller-Friedberg-Strasse 8, 9000 St. Gallen, Schweiz {boris.otto kai.huener hubert.oesterle}@unisg.ch
Publikationsorgan	Bowen, P.; Elmagarmid, A. K.; Österle, H.; Sattler, K.-U. (Hrsg.): Proceedings of the 14th International Conference on Information Quality, Potsdam, Deutschland, 2009
Status	Veröffentlicht

Tabelle B1-1: Bibliographische Angaben zum Beitrag „Identification of business oriented data quality metrics“

Abstract

Corporate data of poor quality can have a negative impact on the performance of business processes and thereby the success of companies. Similar to machine tools corporate data show signs of wear (imaging a moving customer with a new address, for example) and have to be monitored continuously for quality defects. Effective quality control of corporate data requires metrics that monitor potential data defects with the most significant impact on the performance of a company's processes. However, due to company specific success factors and IT landscapes, it is hardly possible to provide generic metrics that can be implemented without any adjustment. This paper presents a method for the identification of business oriented data quality metrics. The presented approach takes into account company specific requirements from both a business and an IT perspective. The method's design and evaluation process is discussed in the context of three real-world cases.

1.1 Introduction

Poor corporate data can have a negative impact on the performance and the success of companies [Redman 1998, Batini/Scannapieco 2006]. Corporate data are generated in business processes (when entering design parameters for manufacturing a watch movement, for example), and they are used in business processes (when manufacturing a watch movement using design drawings, for example). If critical data are missing, a business process may be severely impeded. And if critical data are flawed (e.g. inaccurate values), undesired results are very likely to occur. With regard to the exam-

ple of watch movement manufacturing, missing design drawings would increase process cycle time, and inaccurate data in the design drawings would result in the production of scrap. Apart from impairing business process performance, poor data quality (DQ) may result in inappropriate or even wrong decision-making regarding a company's overall business strategy [Orr 1998]. For example, consolidating procurement volumes for achieving optimized procurement conditions is possible only if a supplier or several subdivisions of one supplier can be uniquely identified. If the quality of supplier master data is poor, the decision in favor of or against a certain supplier will be based on poor information.

While the fact that DQ does affect business process performance becomes obvious by simple examples like the one mentioned above, the question still remains as to what DQ aspects have what impact on what aspects of business process performance. The definition of DQ as 'fitness for use' [Redman 1996, Wang/Strong 1996] leads to a set of requirements on data as determined by a specific business context. While a specific set of corporate data can be highly appropriate for one purpose, the same set of data can be totally inappropriate for another [Batini/Scannapieco 2006].

Both theoretical research and practical experiences made in business projects have shown that knowing about some generic interrelations (e.g. missing data usually contribute to decreased process cycle time) provides no sufficient basis for sustainable data quality controlling (e.g. identifying critical business problems caused by data defects as a first step). While literature offers a number of dimensions and frameworks for identifying various DQ aspects [Wang/Strong 1996, Redman 2001, English 2002], there has hardly been any presentation of concrete DQ metrics or procedures to identify such metrics. As a matter of fact, this is what has been explicitly declared further theoretical work to be done [Batini/Scannapieco 2006, Lee et al. 2006]. Also, the interrelatedness between data defects and their impact on business process performance usually is illustrated by means of specific cases [Strong et al. 1997a, Redman 1998, Fisher/Kingma 2001], what suggests that company specific aspects seem to be of crucial importance.

As taking into consideration the importance of company specific aspects for identifying the interrelations between DQ and business process performance, the development of a method supposed to collect company specific requirements, include findings known from previous research, and provide DQ metrics in relation to process metrics, seems to be appropriate. The research question to be answered is: How can companies identify those DQ metrics which are relevant to their business process metrics?

The paper discusses two central components of the method suggested, namely the procedure model and the involved roles. A part of the method's meta-model is also presented in order to clarify the semantics of the concepts proposed. The main part of the paper starts with information on the background as well as related work of the topic to

be examined, and outlines design science research and method engineering as the research approaches the work is based on. Following these approaches, the next section describes how the method was applied iteratively in three business cases and how it was adapted accordingly. Finally, the last section gives a brief conclusion of work at hand and offers an outlook on future research to be conducted.

1.2 Background

1.2.1 Basics

Information systems provide data in a certain business context (a business process in which customer master data are used, for example). When data are used by human beings, they turn into information, and information finally turns into knowledge. A detailed discussion on the differentiation of data, information, and knowledge is offered by [Davenport/Prusak 1998, Bourdreau/Couillard 1999, Spiegler 2000, Boisot/Canals 2004], for example. This paper postulates that data contain pieces of information, and information can be extracted by interpreting data, e.g. by subordinating data under existing categories or schemas. Data interpretation usually is done by computer programs, e.g. by interpreting a certain sequence of characters as a date of birth. So, while any transformation of data into information usually (though not always) is independent of the user of the data, it is by any means dependent on the context the data are used in (the computer program, in the example) [Tuomi 1999]. Knowledge, finally, is generated by linking pieces of information with one another, or by linking pieces of information with existing knowledge. This transformation does depend on the user of the information, and on their specific situation. The result of any such transformation process can be seen in the effect generated by any concrete action of the information consumer in a certain context. For example: In a DQ monitoring process, a figure (piece of data) is interpreted as a certain value of a metric (information) triggering a maintenance activity (effect) because a certain threshold value (existing knowledge) has been exceeded (linking of information and knowledge).

Regardless of such clear theoretical differentiation between data and information, practitioners use the term ‘data’ in a broader sense. Master data (e.g. customer or materials master data) are not just values (e.g. 0721) but comprise also the act of interpreting by means of certain schemas (e.g. a telephone area code) or in a certain context (e.g. area code plus a customer’s phone number). As the method to be presented in this article does not so much aim at a theoretical differentiation of certain terms but rather focuses on the practical use of data in business processes, the paper favors broader semantics of the term ‘data’ (cf. Figure B1-1).

DQ is defined by the degree of benefit (or value) perceived by a user using certain data in a certain context (‘fitness for use’) [Redman 1996, Wang/Strong 1996]. The benefit

the data bring about (i.e. their quality) is described by a set of quality dimensions from the user's point of view, such as timeliness, completeness, and believability. In order to obtain concrete values for DQ, these DQ dimensions need to be operationalized by DQ metrics [Batini/Scannapieco 2006, Lee et al. 2006, Batini et al. 2007, Caballero et al. 2007]. For any DQ metric a measuring method has to be specified that determines a point of measuring (where does the measurement take place?), the object (what is to be measured?), a measuring tool (how is the measuring done?), and a measuring scale [Batini/Scannapieco 2006].

1.2.2 Related work

There have been numerous theoretical studies on the identification of DQ dimensions [Wang et al. 1995, Redman 1996, Wang/Strong 1996, English 1999, Eppler 2006, Lee et al. 2006]. Some of these studies even include a definition of metrics for DQ measurement. However, the metrics proposed are either generic and do not include a description of possible measuring techniques [English 1999], or they refer to certain domains or even single specific cases only [Heinrich et al. 2008]. Also, the impact of DQ on companies' business process performance or on companies' capabilities in general has been examined by many experts [Redman 1998, English 2002, Sheng/Mykytyn 2002, Sheng 2003, Batini/Scannapieco 2006]. However, what has rarely been provided yet are concrete measurements of DQ or any attempts of quantification of any stated impact on business process performance.

While provision of generally applicable DQ metrics is broadly desired by practitioners and explicitly mentioned to be in the focus of further scientific research [Batini/Scannapieco 2006, Lee et al. 2006], the necessity to adapt metrics to specific contexts (e.g. a specific database) is given by definition due to context sensitivity of DQ ('fitness for use'). Therefore, some scientific studies dealt with procedures as to how measuring techniques could be specified in certain contexts [Naumann/Rolker 2000, Burgess et al. 2004, Gebauer et al. 2005, Gustavsson 2006]. An overview on various techniques is offered by Batini et al. [2007], who also describe a method by which DQ metrics can be specified. A similar approach comes from Caballero et al. [Caballero et al. 2008], who describe a method for specifying techniques for DQ measurement on the basis of a specific DQ related information model [Caballero et al. 2007].

1.2.3 Research methodology

The paper offers a description of the design process of the proposed method for the identification of business oriented DQ metrics. Thus, the paper's main topic is a method as an artifact of design science research (DSR). As a consequence, the paper does not present and discuss any concrete DQ metrics, but the method itself and the process of its design. The paper uses DSR as a methodological framework for the general de-

sign process, and method engineering (ME) as a concrete guideline for designing the method. The method is designed and evaluated together with various companies of an applied research program.

DSR is a framework for design oriented research, aiming at the design of solutions to practical problems [March/Smith 1995, Hevner et al. 2004]. Outcomes of DSR are artifacts, i.e. constructs, models, methods, or instantiations. In this context, a method is defined as a procedure applied to solve a problem (e.g. an algorithm or best practices). Regarding several requirements [Hevner et al. 2004], the design process of a DSR artifact comprises phases of constructing, evaluating, and adapting the artifact [March/Smith 1995, Gregor 2006].

A method in the context of ME represents a form of integrated, systematic procedure for developing information systems [Heym/Österle 1993, Nuseibeh et al. 1996]. ME deals with the design of methods [Brinkkemper 1996], and applies method principles (i.e. an engineering design approach of information systems) to the design of methods (i.e. the design of a method's components *design activities*, *design results*, a *meta-model*, *roles*, and *techniques*) [Gutzwiller 1994]. Design activities aim at producing one or more defined design results. Activities can be structured hierarchically and can be part of a process sequence. The total of process sequences constitute the procedure model as one central component of a method [Leist/Zellner 2006]. Being a formal language, the meta-model defines the syntax and the semantics of the results of the activities. Roles aggregate several activities, which are executed by individuals or boards assigned to a certain role. Roles are always involved in activities in one way or another (e.g. 'responsible'). Techniques describe how results or groups of logically associated results are produced 'within' activities [Brinkkemper 1996].

1.3 Method design process

1.3.1 Overview

The paper presents three cases in which, together with the companies involved, business problems and causing data defects are identified and related DQ metrics are specified. The following subsections describe the use and the design process of the proposed method in the context of the three presented cases. Table B1-4 summarizes the cases' contribution to the components of the method. Table B1-5 provides a description of the activities of the procedure model (i.e. the design state after Case C) which are already mentioned in the case descriptions. The companies mentioned in the three cases are members of the consortium research project Competence Center Corporate Data Quality (CC CDQ) which is part of the research program Business Engineering (BE HSG) of the University of St. Gallen.

1.3.2 Case A: Customer data used in customer service processes

Company A is a German telecommunications provider with 234,123 employees worldwide, about 20 million customers, and an annual revenue of 61.12 billion Euros (in 2008). Company A was the first company that evaluated the method. The search for DQ metrics in this company was basically triggered by problems occurring in business processes (invoices were sent to wrong addresses due to flawed address data, for example) and by increasing inconsistencies in customer master data (missing dates of birth, dummy entries, etc.). As these data usually were collected and used by company employees (in call centers or retail stores, for example), a questionnaire was developed to conduct interviews with employees in order to identify business problems (cf. activity I.3 of the procedure model). The questionnaire was structured according to the fifteen DQ dimensions [Wang/Strong 1996], and for each dimension the questionnaire contained a number of (partly company specific) questions. This kind of structured interviewing made it quite easy to assign business problems that could be identified to DQ dimensions (cf. activity I.3). However, no specification of concrete DQ metrics (cf. phase II of the procedure model) has been done so far. Primary objective was to identify DQ problems. Other, more complex tasks (e.g. the specification of measuring points, which must be considered quite time consuming due to large data volumes and a complex system landscape) will be done at a later stage. Therefore, upon documenting the business problems identified and classifying them according to the DQ dimensions the project was completed.

	Procedure model	Roles
Confirmation	The use of a questionnaire for identifying cause-effect chains between business problems and causing data defects has proven to be very helpful, and so has the questionnaire's structuring according to the fifteen DQ dimensions. Only few problems occurred in interviews regarding interviewees' understanding of DQ.	Involving data users and collectors has proven to be very reasonable. A lot of business problems could be identified which the chief data steward had not been aware of before.
New findings	At the beginning a concrete benefit (referring to the business process in which problems are occurring) should be specified. The specification of the measurement approach is a complex task that should be described in detail.	Persons conducting interviews should really be interested not only in identifying business problems but also in designing concrete DQ metrics including measurement devices and procedures. To be able to provide concrete specifications for DQ metrics also persons from IT departments (i.e. the persons responsible for data providing systems) should be involved.

Table B1-1: Findings from Case A

1.3.3 Case B: Design data used in manufacturing processes

Company B is a Swiss producer of watch movements. Unlike in Company A, in Company B the responsibility for designing DQ metrics was clearly assigned to a certain role, the chief data steward. Moreover, at the beginning a concrete benefit referring to a certain process (i.e. a manufacture process in a certain production line) plus all IT systems involved in this process were identified. Thus, interviewees were both process users and technical data stewards.

	Procedure model	Roles
Confirmation	Detailing the documentation of activity II.2 (i.e. defining the four given steps for specifying a DQ metric) has proven to be very reasonable, as that made it possible to determine single aspects very specifically (e.g. specification of measuring points in focus group interviews with technical data stewards). As expected, the addition of activities I.1 and I.2 helped to focus on 'relevant' cause-effect chains and so improved the result of I.3.	Involving both business and technical data stewards in interviews for identification of DQ problems has proven to be very helpful, as, for example, points of measuring could be identified already during the interviews.
New findings	Upon specification of DQ metrics it was discussed whether there are any general requirements on DQ metrics, and if so, if these requirements are met. Appropriate activities should be included.	No new findings could be identified.

Table B1-2: Findings from Case B

Like with Company A, the findings of the interviews at Company B were documented and structured. In addition, the documentation for Company B included an informal description of potential DQ metrics (cf. activity I.3) as agreed upon by the interviewees. This documentation then was iteratively developed into a specification of data to be measured (e.g. all customer data of a particular database), of a measuring device, of a specification of measuring scales, and of measurement procedures (cf. activity II.2). This project led to a requirements specification for the implementation of eight DQ metrics, which meanwhile have been implemented and are being used in the manufacturing process.

1.3.4 Case C: Material data used in maintenance processes

Company C is a German machine manufacturer and automotive industry supplier with 282,123 employees worldwide, 290 manufacturing sites worldwide, and an annual revenue of 45.12 billion Euros (in 2008). For the project with Company C, a machine maintenance process was identified for being in the focus of identifying DQ problems. The primary objective was to achieve improved conditions for the procurement of spare parts and accessories by means of cross-site procurement, with identical parts to be identified over technical characteristics (e.g. weight, dimensions, material). In the

course of the project it turned out that both general requirements on DQ metrics and company specific requirements (e.g. caused by the system landscape or internal reporting requirements) have to be regarded, so activity II.1 and activity III.1 were added to the procedure model.

	Procedure model	Roles
Confirmation	Examining requirements on DQ metrics has proven to be very helpful in order to keep in mind overall design objectives. By taking into account requirements it is possible to adjust expectations regarding the use of DQ metrics at an early stage.	No new findings from our project with Company B had to be evaluated. The role model basically has proven to be applicable for Company C too.
New findings	To evaluate the quality of the DQ metrics developed, it is necessary to examine the degree of mitigation of the DQ problems identified. Therefore it is necessary to identify not only business processes but also KPIs allowing to monitor these processes.	To evaluate the quality of DQ metrics the respective process owners should be involved (e.g. for identifying KPIs whose values are supposed to be manipulated by improved DQ metrics).

Table B1-3: Findings from Case C

1.4 Method Composition

Table B1-4 shows how each case contributed to the development of the method's components. The line labels represent a qualitative statement on the degree of contribution for the design of the documentation model, the procedure model, the roles, and the techniques. The meta-model is left out here, as it is primarily deduced from the documentation model.

Contribution	Documentation model	Procedure model	Roles	Techniques
High	<i>Case C.</i> A template for documenting and ranking requirements for DQ metrics (cf. activity II.1) that can be used as a checklist (cf. activity III.1) was designed.	<i>Case A.</i> The procedure model comprised three activities comparable to current activities I.3, II.2 and II.3. Activity II.2 was just performed on a very high level compared to cases B and C. <i>Case B.</i> Activities I.1 (yet without focusing process metrics as well) and I.2 were added. Activity II.2 was described in more detail.	<i>Case A.</i> The interviews were conducted by a chief data steward with process users. The order was placed by a sponsor. <i>Case B.</i> Both business and technical data stewards were included into the interviews. The role of a process owner was intended as well but not yet involved.	<i>Case C.</i> A list of generic requirements for DQ metrics to be used in activity II.1 was defined in a focus group interview with subject matter experts.

Contribution	Documentation model	Procedure model	Roles	Techniques
		<i>Case C.</i> Focusing process metrics was added to activity I.1. Activities II.1 and III.1 were added as well.		
Medium	<i>Case B.</i> Templates for the informal documentation of the results of I.3 and II.2 were designed.			<i>Case A.</i> A questionnaire with generic but adaptable questions to be used as a guideline for focus group interviews (cf. activity I.3) was designed.
Low	<i>Case A.</i> No further standardized result templates.		<i>Case C.</i> No further roles.	<i>Case B.</i> No further techniques.

Table B1-4: Contribution of cases to the design process of method components

The documentation model and the techniques will not be treated further here. Basically, in all three projects similar techniques (interviews, check lists etc.) were applied. However, these techniques have not been standardized yet, what makes them hardly comparable. The development, testing, and adaptation of concrete techniques and result documents will constitute the next steps in the research process.

As far as the procedure model is concerned, which was developed in an iterative process, no activity was eliminated and at least one activity was added as we moved from one project to the next. So we ranked all three projects with ‘high’ regarding their contribution to the development of the procedure model. The same thing applies for the role model, apart from case C where no further role was added.

1.4.1 Meta-model

Figure B1-1 shows entities and relations that are used to describe the activities of the procedure model and the results. The six entities and eight relations form just a subset of the method’s meta-model. In the context of the paper at hand, their presentation aims at clarifying dependencies between data defects, business problems, and lower process metric values, for example. The following listing provides a brief description for each entity illustrated in Figure B1-1.

- *Business Problem.* State (e.g. delivery of goods not possible) or incident (e.g. production of scrap parts) leading to decreased process performance and hence to poorer values of process metrics. A business problem poses a risk (in terms

of probability of occurrence and intensity of impact [Hubbard 2009], both of which are represented by a random variable) to a business process.

- *Business Process*. Sequence of chronologically and typologically linked tasks intended to generate a clearly defined output bringing about customer benefit. Transforms a given input (e.g. a certain material) into a desired output (e.g. a watch movement) under consideration of certain rules and by using certain resources (e.g. data). Is controlled and designed by means of metrics defined as part of an overall business strategy [Davenport 1993, Österle 1995, Höning 2009].
- *Data*. Representations of objects (e.g. customers) and relations between objects (e.g. Customer A orders Article X) based on the description of certain object characteristics [Mertens et al. 2005]. The process of choosing characteristics from the set of all existing characteristics of an object is an abstraction process, synonyms are *conceptualization*, defining an *abstract syntax* [McCarthy 1962], or *data modeling* [Mertens et al. 2005]. A data element (or attribute [Martin 1975]) is a component of a data model which cannot be further subdivided in logical terms in a given or presupposed setting [Stahlknecht/Hasenkamp 2005]. The paper considers corporate master data (e.g. customer address data, material data, or parts lists), with a focus not on data models (i.e. data classes and data elements) but on values assigned to data elements.
- *Data Defect*. Incident (e.g. a customer's address data is entered wrongly into a CRM system) leading to poorer values of data quality metrics. A data defect poses a risk (in terms of probability of occurrence and intensity of impact [Hubbard 2009], both of which are represented by a random variable) to data.
- *Data Quality Metric*. Quantitative measure (cf. *Metric*, as well denoted by *data quality measure* [English 1999, Redman 2001]) of the degree to which data possess a given quality attribute (e.g. completeness, timeliness, see [English 1999, Redman 2001] for examples). For a data quality metric, one or more measuring methods need to be provided (i.e. where the measurement is made, what data are included, what measuring device is used, and what scale is used for measuring).
- *Metric*. Quantitative measure of the degree to which an item (in the following denoted as 'entity', according to [ISO/IEC 2007]) possesses a given quality attribute (i.e. a feature or characteristic that affects an entity's quality) [IEEE 1990]. A measuring system (i.e. the measuring procedure that is implemented by a measuring device) measures a metric value at a certain measuring point and at a certain measuring frequency. The magnitude measured is mapped to a value on a scale to which a measuring unit is assigned [Muschter 1998, ISO/IEC 2007].

- Process Metric.* Quantitative measure (cf. *Metric*, as well denoted by operational measures [Kaplan/Norton 1992]) of the degree to which a business process possesses a given quality attribute (e.g. lead time or scrap rate). Results directly from process performance, e.g. as the average lead time for processing of an order, or as the number of order cancelations per month [Mende 1995]. Provides information on a process' state indicating weakpoints and allowing immediate reaction [Österle 1995].

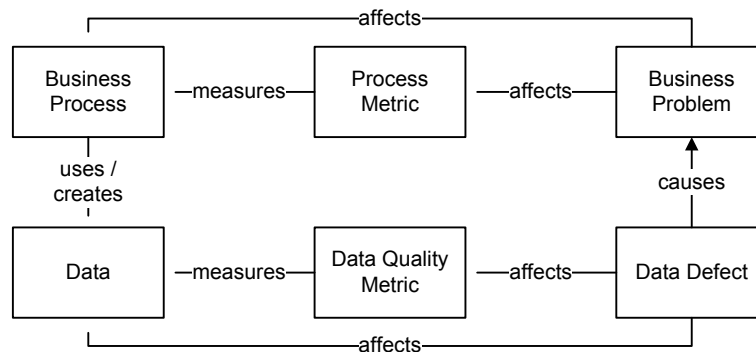


Figure B1-1: Entities and relations of a business oriented data quality metric

1.4.2 Procedure Model

Figure B1-2 shows the procedure model as it was specified upon completion of the project with Company C. The color codes of the letters under each activity box indicate whether this activity was fully used (black), partly used (gray), or not used (white) in the respective project (cf. Table B1-4 for a brief description of the activities' design process). Table B1-5 gives a brief description for each activity.

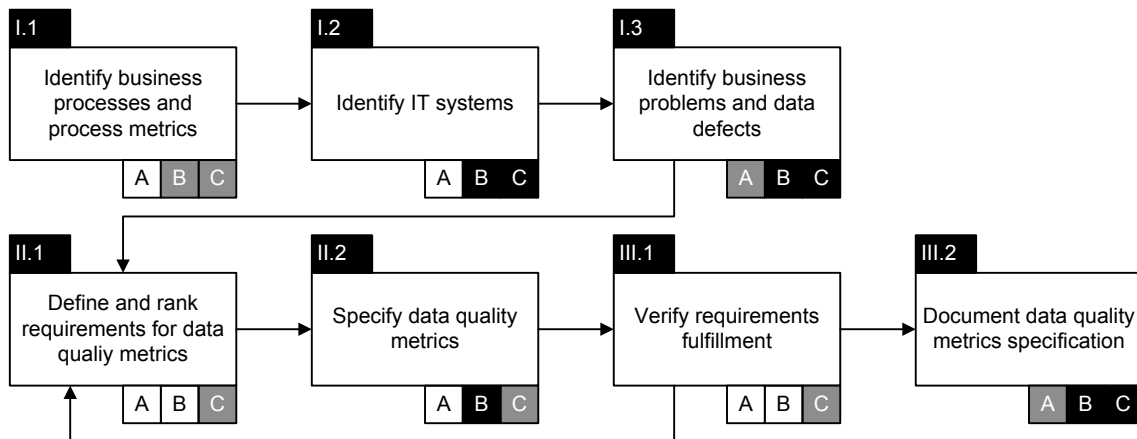


Figure B1-2: Procedure model and degree of usage of activities in each case

Name	Description
Phase I: Collect information	
I.1. Identify business processes and process metrics	I.1 is a preliminary activity for I.3 and aims at identifying business processes and process metrics to focus on during the remaining identification process. Criteria for the selection of a particular business process might be its importance for the company's business success, for example.

Name	Description
	<p>However, an important selection constraint is the availability (i.e. existence and access) of metrics and measurement values for the selected business processes to enable a comparison of process metric values and measurement values of the subsequently designed DQ metrics. Beside an informal documentation of the rationales for choosing the particular business processes and metrics, the result of activity I.1 comprises a list of contacts (i.e. process owner and process user) which might be invited to the focus group interviews (cf. activity I.3).</p>
I.2. Identify IT systems	<p>I.2 is a preliminary activity for I.3 and aims at identifying IT systems (e.g. ERP systems, CRM systems, or databases) that are supporting the identified business processes (cf. I.1). The overall objective of activity I.2. is to identify IT experts for the focus group interviews of activity I.3).</p>
I.3. Identify business problems and data defects	<p>I.3 is the main activity of phase I and aims at identifying cause-effect chains between business problems and data defects (i.e. data defects that cause business problems). The top-down search direction (i.e. first identifying critical business problems and then indentifying causing data defects) has proven to be effective in the discussed cases, but indentifying potential business problems for already known data defects might be useful as well.</p> <p>Subject matter experts from both business and IT departments should be involved in collaborative focus group interviews to enable discussions with different perspectives. [Hüner 2011] contains interview guidelines and exemplary cause-effect chains to support the focus group interviews. The result of activity I.3 is an informal documentation of cause-effect chains (i.e. business problems and data defects) and likely affected business processes, process metrics, and data classes.</p>
Phase II: Specify requirements and design data quality metrics	
II.1. Define and rank requirements for data quality metrics	<p>The result of activity II.1 is a ranked list of requirements a DQ metric must meet. The list comprises both generic (e.g. a specified measurement device and measurement points) and company specific requirements (e.g. facility to visualize metric measurements in a specific manner). [Hüner 2011] contains a list of generic requirements that have been identified in a focus group interview with subject matter experts of different companies. The list is used as a checklist in activity III.1 to verify the requirements' fulfillment.</p>
II.2. Specify data quality metrics	<p>The result of activity II.2 is a specification of at least one DQ metric. The basis for the metrics' design process are the results of activities I.3 and II.1. The activity comprises for each DQ metric</p> <ul style="list-style-type: none"> • the specification of a (subset of a) data class (e.g. customer data) that is measured by the metric, • the specification of a measurement device (e.g. a database with data analysis functionality) and a measurement point where the measurement is to be done, • the specification of a measurement scale to which a measured magnitude is to be mapped, and • the specification of a measurement procedure that is implemented by the measurement device and performs the measurement at the given measurement point at a measurement frequency.
Phase III: Approve and document results	
III.1. Verify requirements fulfillment	<p>Activity III.1 verifies the requirements defined in activity II.1. If a requirement is not met, the process starts again with II.1 in order to check the requirements' content and ranking.</p>

Name	Description
III.2. Document data quality metrics specification	The result of III.2 is a documentation of the specification of the DQ metrics (cf. activity II.2) including the identified cause-effect chains (I.3), and the requirements (activity II.1). This documentation might be used as a requirements documentation for the implementation of the DQ metrics, for example.

Table B1-5: Description of activities of the procedure model

1.4.3 Roles

Table B1-6 shows the assignment of roles to activities of the procedure model by using the RACI notation [ITGI 2007] that declares a role to be responsible, accountable, consulted, or informed regarding a particular activity. The role definitions partially base on the roles described by Wende [2007]. The whole identification and specification process is managed by the chief data steward who is involved in each activity.

- *Chief Data Steward.* The person which puts strategic DQ objectives into reality and leads the process of identifying DQ metrics. Directs business data stewards and technical data stewards.
- *Business Data Steward.* Cooperates with company departments and business units on DQ issues. One business data steward is proposed as being responsible for one specific data class (e.g. material master data, customer master data etc.). Yet other concepts are possible too (e.g. responsible for a certain department, a certain process etc.).
- *Technical Data Steward.* Has detailed knowledge about the IT systems managing and providing corporate data. One technical data steward is proposed as being responsible for one specific data class (cf. business data steward).
- *Process Owner.* Plans and monitors process objectives, and plans and initiates measures for improvement. The process owner might grant the permission to use metric measurements of the monitored processes.
- *Process User.* Execute concrete tasks belonging to a certain process. Regarding the creation and usage of data in business processes, data defects have impact on daily work of these individuals.
- *Sponsor.* A single person (e.g. the chief executive officer) or a board (e.g. management board) that has sufficient resources (both money and power) to support or prohibit an initiative like the identification and implementation of DQ metrics and that might place the order with the chief data steward for designing and implementing DQ metrics.

	Chief Data Steward	Business Data Steward	Technical Data Steward	Process Owner	Process User	Sponsor
I.1. Identify business processes and process metrics	Responsible, Accountable			Consulted		
I.2. Identify IT systems	Responsible, Accountable			Consulted		
I.3. Identify business problems and data defects	Responsible, Accountable	Consulted	Consulted		Consulted	
II.1. Define and rank requirements for data quality metrics	Responsible, Accountable	Consulted	Consulted			Informed
II.2. Specify data quality metrics	Responsible, Accountable	Consulted	Consulted			
III.1. Verify requirements fulfillment	Responsible, Accountable	Consulted	Consulted			Informed
III.2. Document data quality metrics specification	Responsible, Accountable					Informed

Table B1-6: Assignment of roles to activities of the procedure model

1.5 Conclusion and Further Research

The article at hand describes a method for identifying business oriented DQ metrics and outlines two of the method's central components (i.e. a procedure model and a role model). The method takes up a number of input variables (e.g. critical success factors for a company) and provides a specification of concrete DQ metrics that can be used as a basis for implementation of such metrics.

During the process of designing and adapting the method it became obvious that general findings from previous research (e.g. DQ metrics requirements) and company specific findings (e.g. concrete data defects identified) need to be combined. Generic approaches (e.g. DQ dimensions) can be used here as supporting tools but do not offer any applicable results for concrete business contexts without prior operationalization.

Despite the presented description of seven activities and six roles to perform these activities, the process of identifying 'DQ metrics that matter' remains a complex task that has to be supported not only by a guiding procedure, but also by a repository of already implemented DQ metrics which serve as best practices. In particular for the

process of identifying ‘relevant’ cause-effect chains between business problems and data defects within focus group interviews (cf. activity I.3) illustrative examples of correlated measurements of process and DQ metrics enable discussions and thereby support the identification process. Therefore the design and documentation of DQ metrics in real-world cases, the analysis of the identified cause-effect chains, and the derivation of generic cause-effect patterns between data defects (e.g. grouped by DQ dimensions) and business problems (e.g. grouped by commercial sectors, or supply chain reference models) constitute multiple areas for further research.

2 Specifying data quality metrics that matter to business

Titel	Specifying data quality metrics that matter to business
Autoren	Kai M. Hüner; Boris Otto; Hubert Österle Universität St. Gallen, Institut für Wirtschaftsinformatik Müller-Friedberg-Strasse 8, 9000 St. Gallen, Schweiz {kai.huener boris.otto hubert.oesterle}@unisg.ch
Publikationsorgan	Journal of the Association for Information Systems
Status	In zweiter Begutachtung

Tabelle B2-1: Bibliographische Angaben zum Beitrag „Specifying data quality metrics that matter to business“

Abstract

Data quality is critical for companies faced with ever increasing competition in the market as well as a growing number of regulatory requirements. As a consequence, intensive discussions are ongoing in both practice and science on how to make sure that corporate data is of sufficient quality, and how corporate data can be continuously monitored for quality defects. However, while the strategic relevance of data quality in general cannot be denied, a causal relation between data quality and the achievement of strategic objectives is not always easy to identify in a particular case. At the same time, company-specific success factors and IT landscapes make it virtually impossible to provide generic metrics that can be implemented without any adjustment. This paper presents a method for identifying causal relations between data defects, business problems, and strategic objectives as well as for specifying data quality metrics to monitor identified data defects. The aim of using such business-oriented data quality metrics is to identify data defects as early as possible, thereby preventing business problems caused by those data defects. Following a top-down approach, the method aims at specifying metrics solely for data defects that lead to business-critical problems. Besides the method itself, the article illustrates the entire design process which encompasses focus group inter-views, case studies, and a multi-perspective evaluation.

2.1 Introduction

Data defects can compromise the business success of companies [Kagermann et al., 2010 pp. 185-192, Lee et al. 2006, pp. 9-10]. For example, outdated, inaccurate or incomplete data may cause wrong results, delays or decreased performance of even supposedly simple processes such as the transportation of shampoo bottles, or fertilizer production. If companies are to identify and correct critical data defects in time, they should regularly and systematically monitor data quality, in particular the quality of corporate data. The following example briefly illustrates how corporate data has taken

on strategic relevance. Further details relating to the case of Bayer CropScience are provided in 2.6.1.

Bayer CropScience AG (approx. 18,700 employees and revenues of 6.5 billion euros in 2009) is a German manufacturer of agrochemical products. In 2007, the company began setting up a central Enterprise Resource Planning (ERP) system to successively replace its regional ERP systems [Ebner et al. 2011]. A strategic objective of this project is to consolidate material demands on a worldwide basis in order to improve the coordination of sales planning and material demand planning. However, the accuracy of the resulting consolidation is jeopardized by the data quality of product hierarchies. The product hierarchy is a data attribute (i.e. an 11-digit numeric string) consisting of five elements and defining where a product belongs in the organization (distribution company, business section, business region) as well as its main active ingredient and composition. This means that data defects in product hierarchies can lead on the one hand to the incorrect assignment of sales figures and products, and therefore to production volume that is not required. On the other hand, there is the risk for demand planning that a calculated demand is assigned to the wrong ingredient, with the consequence that a material demand cannot be met in subsequent production.

Although the general strategic relevance of data quality for a company's business success cannot be denied, the connection between data quality and strategic business objectives is not always easy to identify. This is mainly due to the fact that data quality is a phenomenon that is both multi-dimensional and context-dependent [Wang/Strong 1996, p. 6, 5-34]. Companies need to deal with various data classes (e.g. customer data, material data, supplier data), for which a number of data quality dimensions (e.g. accuracy, completeness, timeliness) might have to be considered. In order to be able to specify target levels for each data quality dimension, the context in which data is supposed to be used (e.g. demand planning) needs to be thoroughly analyzed. Furthermore, target level definition should be based on a cost-benefit analysis [Lee et al., 2006 pp. 13-18]. In the case of Bayer CropScience mentioned above, action to improve the product hierarchy quality should only be taken if the costs of the measure do not exceed the anticipated benefit to be gained from globally centralized demand planning.

In order to conduct a thorough cost-benefit analysis and to define target levels for each data quality dimension, it is necessary to identify chains of causal relations (i.e. causal chains) between a data defect (e.g. defective product hierarchy), a business problem (e.g. only a subset of global demands of a certain material is consolidated) and a strategic objective (e.g. accurate consolidation of global material demands). Once such causal chains have been identified, business-oriented data quality metrics (i.e. metrics to measure data quality dimensions that are mission-critical for a company) can be specified for the root cause (i.e. the data quality defect) of the causal chain.

While there have been numerous scientific studies on the issue of data quality, which elaborate on how data quality can be measured, what metrics and measuring techniques can be used, and how specific metrics can be designed, there has yet to be a contribution which presents a method for systematically identifying the above mentioned causal chains. The research objective of this paper is therefore to design a method a) to identify causal relations between data defects, business problems, and strategic objectives of a company; and b) to specify data quality metrics for monitoring the identified data defects.

In general, methods represent recommended courses of action for achieving a specific goal (e.g. to design an information system) [Brinkkemper 1996, pp. 275-276, Heym/Österle 1993, pp. 132-133, Olle et al. 1988, pp. 1-2]. For this purpose, methods specify the goal to be achieved and describe procedures, activities and techniques to facilitate the systematic achievement of that goal (see Section 2.4.2 for further definition). The multiple application of a method therefore leads to comparable results, enables the reuse of information systems (e.g. data structures, documentation templates) and facilitates estimation of the associated effort. These few features already highlight the usefulness of a method for specifying business-oriented data quality metrics. In addition to the comparability of specified metrics, the reduction in effort through repeated application is of crucial importance. The data models of information systems used by large companies can encompass several thousand data attributes (cf. material data model of an ERP system from SAP with some 800 attributes), and are also continuously changed in order to meet new legal and market requirements (cf. 80 validation rules for monitoring material data quality at Bayer CropScience in Section 2.6.1). It should therefore be possible for different people within the company to identify causal chains with a low amount of effort.

The remainder of the paper starts by presenting a brief overview of related work in the field and the concept of business-oriented data quality metrics in Sections 2.2 and 2.3 respectively. Section 2.4 outlines the methodology of the presented research, describes the method design process and briefly explains applied research techniques. Section 2.5 then presents the method as a design artifact, which is the core contribution of the paper. A multi-perspective evaluation of the method is given in Section 2.6. Finally, Section 2.7 summarizes the findings, discusses limitations of the presented research results and provides an outlook on future research to be carried out.

2.2 Definition of Fundamental Concepts

2.2.1 Data and information

Practice and science differentiate the concepts of data and information [Boisot/Canals 2004, pp. 45-62, Bourdreau/Couillard 1999, pp. 25-26, English 2002, p. 52, Tuomi 1999, pp. 45-48]. A distinction commonly drawn is that information arises through the

subjective and context-specific interpretation of data (representation of objective facts), which provides the foundation for business decisions and forms both the input and output of business processes [Price/Shanks 2005, p. 89]. However, this paper considers the purpose of data (or information) processing, which is to support the business processes of a company in the creation of goods and services. Whether the representation of business objects necessary for technical processing is denoted as data or information appears insignificant in this context. For this reason and as in the case of other works with a similar focus [McKinney Jr./Yoos 2010, p. 331, Pipino et al. 2005, p. 212, Wang 1998, p. 59], no distinction is made between these two concepts in this paper.

2.2.2 Data quality

Data quality has been a research topic in numerous studies which elaborate on lists and categories of data quality dimensions based on empirical research [Wang et al. 1995, Wang/Strong 1996], experiences from case studies [English 1999, Redman 1996] and theoretical implications [Price/Shanks 2005, Wand/Wang 1996]. Common to many of these studies is that data quality is seen as something that is mainly determined by the ‘fitness for use’ of data, i.e. whether data is of good or poor quality depends on the context and the end users of data.

2.2.3 Data quality management

Data quality management (DQM) comprises activities for the improvement of data quality [Batini/Scannapieco 2006, pp. 69-71]. Going beyond mere reactive action (e.g. identification and repairing of data defects) [Shankaranarayanan/Cai 2006, pp. 303-304], DQM works as a proactive and preventive concept, characterized by a continuous cycle consisting of activities to define, measure, analyze and improve data quality [English 1999, pp. 69-81, Eppler/Helfert 2004, 311-325, Khatri/Brown 2010, Wang 1998, 95-105]. An overview of the most relevant approaches to DQM is given by Batini et al. [2009].

2.2.4 Data quality measurement

Apart from the structural conceptualization of data quality by means of various data quality dimensions, the measurement of data quality has been a central issue of many scientific studies. In this context, data quality metrics serve to operationalize data quality dimensions as they specify the data to be measured, a measuring point, a measuring technique, and a measuring scale [Batini/Scannapieco 2006, p. 19]. However, most studies remain somewhat limited regarding the question of business impact, as they only provide calculation formulas for single data quality dimensions, e.g. for timeliness [Heinrich et al. 2007, Heinrich et al. 2009], accuracy [Gorla/Krishnan 2002] or

completeness [Cai/Ziad, 2003]. Further studies present procedures and techniques for measuring data quality by means of interviews or surveys [Huang et al. 1999, pp. 87-88, Lee et al. 2002, Nicolaou/McKnight 2006, Price et al. 2008] or by validation rules [Fan et al. 2008, Hipp et al. 2007]. Attention should also be paid to approaches modeling data production processes and to the primitive operations (i.e. cartesian product, difference, projection, selection, union) that constitute a data aggregation process [Parssian et al. 1999, Parssian et al. 2004, Parssian et al. 2009].

2.3 Literature review

For the process of specifying data quality metrics in a specific context and also for the process of identifying relevant data quality dimensions, only rough recommendations have been proposed so far [Batini et al. 2007, Caballero et al. 2008, Heinrich et al. 2009]. Although the need for ways of enabling the systematic operationalization of data quality dimensions has been identified [Batini/Scannapieco 2006, pp. 221-222], what is still missing is a systematic procedure for specifying data quality metrics and considering the business impact of data quality. Therefore, the method presented in this paper aims at the identification of causal relations between data defects, business problems, and strategic objectives as well as the specification of data quality metrics for monitoring identified data defects. In this endeavor, the method combines several of the approaches already mentioned (e.g. data quality dimensions, questionnaires and validation rules for measuring data quality), and proposes an integrated procedure model, result types specified by a meta-model, roles and techniques for facilitating the method's application (see Section 2.4.2 for definitions of these concepts).

The idea of deriving data quality metrics from strategic business objectives basically follows the concept of the balanced scorecard [Kaplan/Norton, 1992] and the Benefits Dependency Network [Peppard et al. 2007], both of which model cause-effect relations between different levels (or divisions) of a company. Table B2-2 gives an overview of how the state of the art was actually used in the method design process (see Section 2.4.2 for details regarding components of a method).

	Meta-model	Procedure model	Roles	Techniques
1	[IEEE 1990] [Lee et al. 2006] [Robbert/Fletcher 2007] [Batini et al. 2009]	[Even/Kaiser 2009] [Strong et al. 1997b] [Kahn et al. 2002] [Winkler 2004] [Batini/Scannapieco 2006]	—	[Gorla/Krishnan 2002] [Shankaranarayanan et al. 2003] [Shankaranarayanan/Cai 2006]
2	[Martin 1975] [Wand/Wang 1996] [IEEE 1998] [Even/Shankaranarayanan 2005] [Price/Shanks 2005]	[Lee et al. 2002] [Batini et al. 2007] [Heinrich et al. 2007] [Caballero et al. 2008]	—	[Parssian et al. 1999] [Raghunathan 1999] [Joshi/Rai 2000] [Parssian et al. 2004] [Winkler 2004] [Lee et al. 2006]

	Meta-model	Procedure model	Roles	Techniques
	[Caballero et al. 2007] [ISO/IEC 2007] [Even/Kaiser 2009] [Viscusi et al. 2009]			[Nicolaou/McKnight 2006] [Even et al. 2007] [Hipp et al. 2007] [Fan et al. 2008] [Parssian et al. 2009]
3	[Wang/Strong 1996]	—	[Weber et al. 2009a]	[Ballou et al. 1998] [Redman 1998] [Fisher et al. 2003] [Pierce 2004]
4	—	—	—	[Wang/Strong 1996] [Fisher/Kingma 2001] [Lee et al. 2002] [Alur et al. 2007] [Heinrich et al. 2007] [Heinrich et al. 2007] [Price et al. 2008] [Heinrich et al. 2009] [Heinrich et al. 2009]
1: Used as a design rationale. 2: Used for orientation for the design of a particular model element. 3: Used with modification as a model element. 4: Used without modification as a model element.				

Table B2-2: Publications identified by literature review

2.4 Research methodology

2.4.1 Method design process

The method presented in this paper is an outcome of design-oriented research and follows Design Science Research (DSR) as methodological paradigm. DSR aims at the design of artifacts (i.e. constructs, models, methods and instantiations) that are supposed to solve practical problems [Hevner et al. 2004, p. 76, March/Smith 1995, pp. 256-258]. The Competence Center Corporate Data Quality (CC CDQ) constitutes the research context of the method design process. The CC CDQ is a consortium research project [Österle/Otto 2010] and part of the Business Engineering research program at the University of St. Gallen. Since 2006, it has been developing artifacts to support the quality management of corporate data in collaboration with renowned partner companies.

Taking into account various requirements to be met by the design of artifacts [Gregor 2006, pp. 628-629, Hevner et al. 2004, pp. 82-90], the design process of a DSR artifact consists of several phases, namely problem identification, definition of design objectives, artifact design, and artifact evaluation. In order to ensure the clarity and repeatability of such a process, decisions made in the course of a design process should be justified and documented. For this purpose, Rossi et al. [2004, pp. 367-369] adopt the concept of design rationales from the domain of software engineering [Louri-

das/Loucopoulos, 2000, Potts 1989, Potts/Runs 1988]. During the design process of the presented method, design decisions were documented according to this approach (see Appendix D, Table D-2).

Figure B2-1 shows the design process of the presented method in terms of DSR methodology [Peffer et al. 2008, pp. 52-56] and lists research techniques that have been applied in the particular phases. The research gap was identified by subject matter experts from partner companies of the CC CDQ in two focus group interviews (A and B, see Appendix D, Table D-1), and was validated by a literature review.

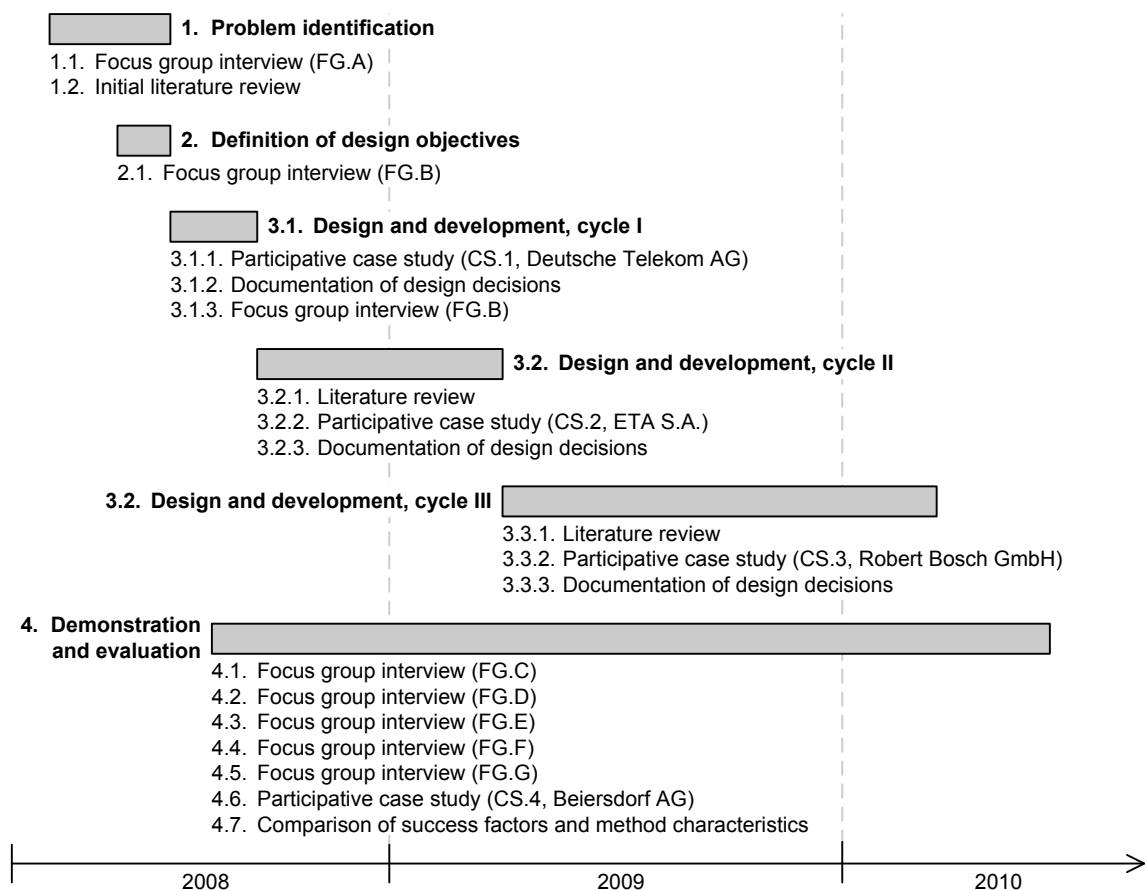


Figure B2-1: Method design process

The presented method is the result of three design cycles (see CS.1, CS.2, and CS.3 in Figure B2-1) carried out between February 2008 and June 2010. The application of various research techniques ensures the relevance of the designed artifact as well as the rigor of the research process).

- *Relevance.* In focus group interviews [Morgan/Krueger 1993], subject matter experts from partner companies of the CC CDQ defined research objectives and evaluated the research results (see Appendix D, Table D-1). In participative case studies [Baskerville 1997], the respective development status of the method was applied and evaluated. The application of these research techniques ensures that the method addresses a relevant problem [Rosemann/Vessey 2008, pp. 7-12], and that the method is practicable and useful.

- *Rigor.* The iterative design of the method in three cycles, the application of Method Engineering (see Section 2.4.2) as an established design technique in information systems research and the documentation of design decisions taken (see Appendix D, Table D-2) ensure the possibility of validating, adapting and further developing the method in future research work. Repeated literature reviews ensure the use of existing research results and classification of the method as state of research.

In DSR, design objectives guide the design and evaluation process [Peppers et al. 2008, pp. 55-56]. Table B2-3 shows the design objectives which concretize the research objective outlined in Section 2.1. These objectives were defined in a focus group interview (see Figure B2-1, FG.B), for which the guidelines of modeling [Schütte/Rotthowe 1998] and the fundamental objectives of DSR [Hevner et al. 2004] were used as input.

Design objective	Description
DO1: Clarity ⁷	The method comprises only the required number of elements.
DO2: Effectiveness (method) ⁸	The outcome of the method's application is a specification document of business-oriented data quality metrics.
DO3: Effectiveness (metrics) ⁹	Specified data quality metrics consider success factors for effective data quality measurement.
DO4: Efficiency ¹⁰	The benefit of specifying and using data quality metrics exceeds the costs of the method's application.
DO5: Flexibility ¹¹	The method is adaptable and extendable for new business contexts.
DO6: Language adequacy ¹²	The method represents the problem of specifying business-oriented data quality metrics in an adequate form (i.e. understandable to and applicable by involved roles).
DO7: Systematic design ¹³	The method instantiates the meta-model for methods (see Section 2.4.1).
DO8: Utility ¹⁴	Specified data quality metrics enable monitoring of business-critical data defects.

Table B2-3: Design objectives for the method design process

2.4.2 Method Engineering

According to Method Engineering, a method is a systematic procedure for developing information systems [Heym/Österle 1993, p. 345]. Method Engineering enables the systematic design of methods by applying methodological principles (i.e. engineered design of information systems) to the design process. In this sense, a method compris-

⁷ See principle of clarity [Schütte/Rotthowe 1998, pp. 248-249].

⁸ See effectiveness as a goal of design science research [Hevner et al. 2004, p. 85].

⁹ See effectiveness as a goal of design science research [Hevner et al. 2004, p. 85].

¹⁰ See principle of economic efficiency [Schütte/Rotthowe 1998, pp. 247-248].

¹¹ See principle of systematic design and principle of comparability [p. 249].

¹² See principle of language adequacy [pp. 246-247].

¹³ See principle of construction adequacy and principle of systematic design [245-246, 249].

¹⁴ See utility as a goal of design science research [Hevner et al. 2004, p. 80].

es both a product description and a process description [Brinkkemper 1996, pp. 275-276, Olle et al. 1988, pp. 2-3]: Design results as the method's outcome (i.e. the product) is specified by the meta-model. And the process elements (i.e. activities, roles, and techniques) describe which entities of the meta-model are instantiated in which activity, who provides the information required for the instantiation, and who carries out activities and techniques.

The meta-model for Method Engineering [Gutzwiller 1994, p. 13] (see Figure B2-2; a similar variant is proposed by Nuseibeh et al. [1996]) is the conceptual basis for the method proposed in this paper. It ensures a modular method design and thereby addresses DO5 (see Table B2-3). From a component perspective, a method comprises a meta-model (i.e. specification of design results), a procedure model (i.e. all activities), a documentation model (i.e. all design results), roles and techniques.

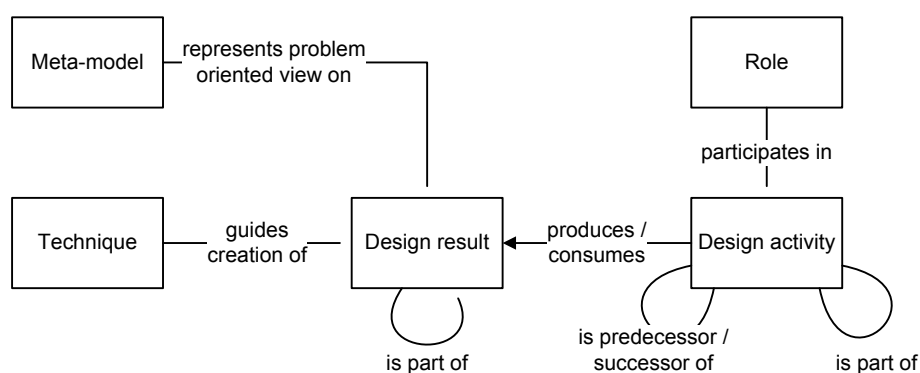


Figure B2-2: Method Engineering according to Gutzwiller [1994, p. 13]

2.4.3 Application of the method

Section 2.6.1 introduces evaluation criteria for each of the design objectives defined in Table B2-3. These criteria were used to evaluate the method with various research techniques. One of these techniques was to apply the method in a participative case study at Beiersdorf. As the following sections use this case to explain the method, the context of the case study is first outlined in brief.

Beiersdorf AG (approx. 20,300 employees and revenues of 5.7 billion euros in 2009) is a German consumer goods manufacturer. Its well-known brands include Eucerin, Nivea und tesa. In 2004, the company introduced a central product lifecycle management system for managing globally utilized product data. This also involved defining and implementing harmonized, company-wide processes for the creation and maintenance of the product data used by the system. A central organization unit located in the Corporate Supply Chain Management Division is responsible for organizing these processes.

The smooth running of cross-company supply chain processes at Beiersdorf is attributable to good organizational and technical master data management structures. Regular-

ly occurring business operations problems that counteract strategic business objectives (e.g. compliance with legal and regulatory provisions, high service level), or that cause high costs have not been regularly reported. Nevertheless, critical voices have pointed to the quality of product data, particularly with regard to their company-wide use. For example, some distribution centers have been complaining about the accuracy of weights of newly launched products (i.e. of related logistic units). Logistics data defects of this kind can result in additional costs due to repackaging and restocking if the tolerance range for pallet weights is exceeded.

Having noted these complaints and in view of a growing awareness that product data quality is of particular importance at the interfaces with external partners (e.g. customer companies, data pool providers, logistics service providers), Beiersdorf initiated a project aiming at specifying and implementing data quality metrics for monitoring these defects. Between February 2010 und June 2010, seven data quality metrics with a total of 32 validation rules were specified in seven interviews and five project meetings using the presented method [Hüner et al., 2011c].

2.5 Method design

2.5.1 Overview and structure of the method

This section explains the individual elements of the method for specifying business-oriented data quality metrics. Since procedure model, documentation model, and techniques are closely linked with one another, these partial models are not described individually but in the form of method fragments (in each case, design result, activity producing the result, roles involved and techniques used). Vignettes [Stake 1995, pp. 128-130] provide information on design results from the method application at Beiersdorf [Hüner et al. 2011c].

Figure B2-3 provides an overview of the method with the activities (left-hand side) and the respective roles involved (right-hand side). The middle column of Figure B2-3 shows the attributes of the meta-model which are instantiated by the stated roles in the respective activity and thus specify the result of the individual activities. The “*” symbol means that all the attributes of the stated meta-model entity type are instantiated (i.e. required information for the related result documents is to be collected) when the method is applied.

Activities	Results	Roles					
		Corporate Data Steward	Sponsor	Process Owner	Data User	Business Data Steward	Technical Data Steward
Phase I: Collect information							
I.1: Define the scope	ProcessActivity.Identifier	x	x	x			
	ProcessActivity.BusinessProcess	x	x	x			
	ProcessActivity.Responsibility	x		x			
	ProcessMetric.*	x		x			
	BusinessProblem.*	x		x			
I.2: Identify data and IT systems	StrategicObjective.*	x		x			
	ProcessActivity.EnterpriseSystem	x		x			
I.3: Identify causal chains	Data.*	x		x			
	StrategicObjective.Causality	x		x			
	BusinessProblem.Causality	x			x	x	x
	DataDefect.*	x			x	x	x
	PreventiveInitiative.*	x			x	x	x
ReactiveInitiative.*	x			x	x	x	
Phase II: Analyze and specify							
II.1: Define and prioritize requirements	DataQualityMetric.Requirements	x	x			x	x
II.2: Specify data quality metrics	Data.Constraints	x				x	x
	DataQualityMetric.Identifier	x				x	x
	DataQualityMetric.DataQualityDimension	x				x	x
	DataQualityMetric.MeasuringPoint	x					x
	DataQualityMetric.MeasuringProcedure	x					x
	DataQualityMetric.ScaleType	x				x	x
	DataQualityMetric.UnitOfMeasurement	x				x	x
DataQualityMetric.MeasuringFrequency	x				x	x	
Phase III: Approve and document							
III.1: Check the fulfillment of requirements		x				x	x
III.2: Document metrics specifications		x	x				

Figure B2-3: Procedure model with relationships to meta-model and roles

The activities of the procedure model support a step-by-step instantiation process of the meta-model (i.e. the specification of result elements, see Section 2.5.2) as they specify what information is to be identified, specified and documented by whom and in what order. The fundamental structure of the method consists of the following three phases:

1. *Phase I: Collect information.* The knowledge required to identify causal chains is distributed across a large number of roles in the company. This phase therefore encompasses activities with a high degree of communication. Information is collected from different roles and consolidated into a scope for application of the method as well as into causal chains.

2. *Phase II: Analyze and specify.* The aim is to specify all the information required to implement at least one data quality metric. The prime task is to formalize identified data defects by means of validation rules (or another measuring technique) and to link the rules (e.g. through calculation formulas) to form metrics.
3. *Phase III: Approve and document.* The aim is to document the specifications of the data quality metrics and the connection to business (i.e. the causal chains) so that the metrics can be implemented.

The three phases are described in more detail in section 2.5.4.

2.5.2 Meta-model

Figure B2-4 shows the meta-model of the method with attributes and relationships of the design objects of the documentation model. Foreign key attributes are not explicitly listed in the meta-model and have to be added for implementation purposes. Key design objects are data defects and data quality metrics. The data defect is part of the causal chain also comprising a business problem and a strategic objective. It represents a certain state of data used in a process activity and causing a business problem. As the data defect is not anticipated by the process activity, a business problem occurs together with a deviation of a process metric's actual value from a desired value.

For the implementation of data quality metrics, the measuring technique and the measuring scale are particularly important. The measuring technique measures a metric value by mapping the result of a measurement performed at a measuring point onto a scale [IEEE 1998, pp. 2-9]. In order to be able to compare the measured values of a data quality metric with those of a process metric (e.g. to show any negative influence of poor data quality on process performance), the level of measurement of the process metric must be taken into account when specifying the data quality metric.

2.5.3 Roles

The roles of the method are based on a reference model for data governance [Weber et al. 2009, pp. 9-13]. The central role is corporate data steward (cf. Figure B2-3). He or she is responsible for configuration (e.g. selection of techniques) and application of the method, coordinates cooperation between the people and boards involved, and documents the results. This role is usually assumed by the head of the company's DQM organization. In large organizations, however, this person may also delegate the task to someone else in his or her team. Further description of the method assumes a DQM organization headed by the corporate data steward (one person) and a sponsor (one person, superior of the Corporate Data Steward) outside of the DQM organization. The sponsor defines goals for the method application, ensures the availability of the neces-

sary resources and approves the business concept (result document of Activity III.2, cf. section 2.5.4.3).

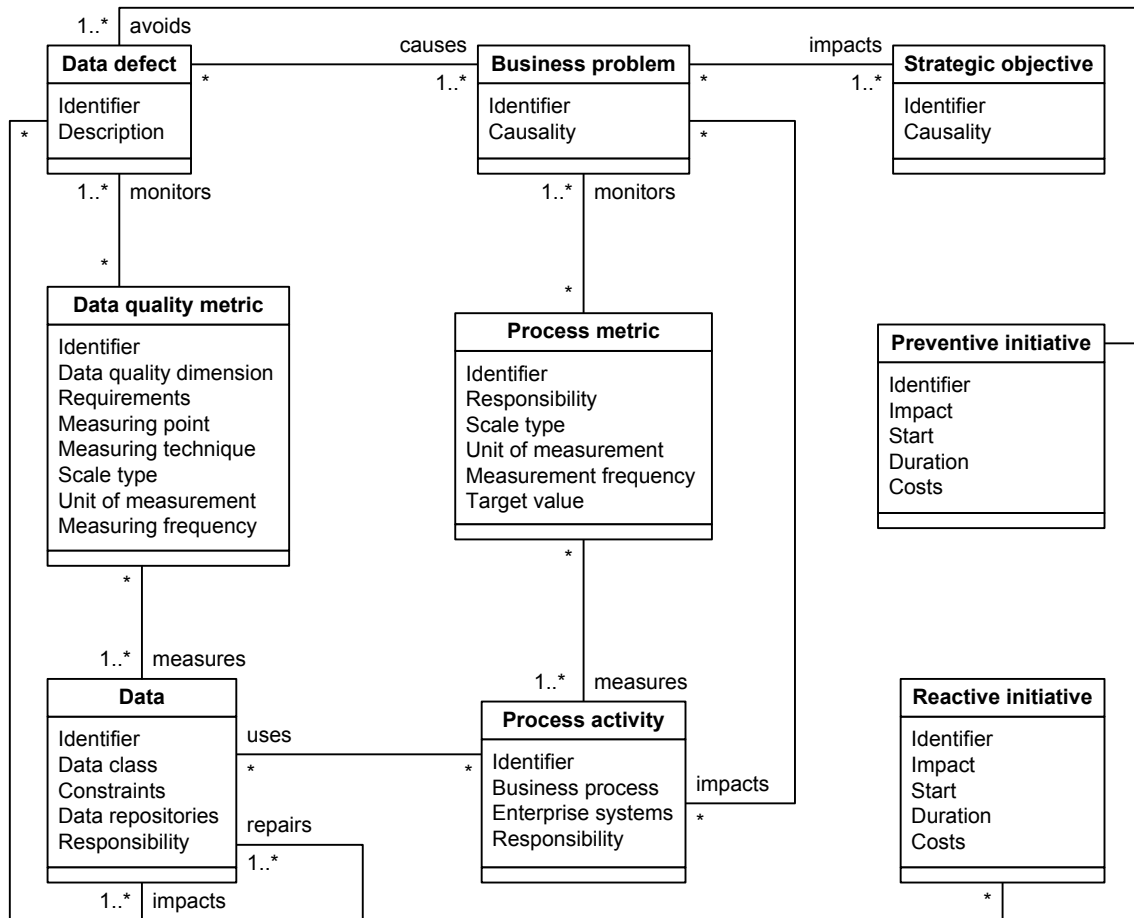


Figure B2-4: Meta-model of the method

Process owners are useful interlocutors when identifying suitable scopes. Typical tasks of a process owner are the planning and monitoring of process goals, the specification and implementation of process metrics, and the planning and initiation of improvement measures. In terms of the method application, the process owner’s task is to assess which process activities in particular could be jeopardized by data defects and which process metrics might clearly show the impacts of possible data defects. Data users are company employees who create and use data in their day-to-day work (i.e. execute work procedures described in process activities), the emphasis being on use in the context of applying the method. As the employees are directly confronted with data defects in their work (e.g. missing, inaccurate, out-of-date data elements), they can describe the immediate impacts of defective data and are therefore important interlocutors when identifying causal chains.

Business data stewards are part of a company’s DQM organization and report to the corporate data steward. They are responsible for working with business departments and maintain, for example, business metadata or business data quality requirements to be met by application systems. Although other responsibilities are possible (e.g. per

department, per process), further description of the method is based on the responsibility of a business data steward for one data object type in each case (e.g. material master, customer master). A technical data steward represents the same for the IT department that a business data steward represents for the business department.

2.5.4 Procedure model and techniques

2.5.4.1 Phase I: Collect information

The prime task of the corporate data steward and sponsor in Activity I.1 is to identify a scope which on the one hand focuses on a major part of the company's value chain and for which on the other hand there are clear causal chains, i.e. clearly recognizable relationships between data defects and business problems (see Table B2-4). The method proposes interviews for the identification of an appropriate scope. To support the interviews, a collection of causal chains and documentation templates (provided as supportive material for several techniques) may be used. For example, data quality simulation is a technique allowing the simulation of causal relations in order to be able to explicate assumed interrelations (see Table D-3 in Appendix D for further information about the method's techniques). The result of this activity is a list of scopes.

The goal is to identify process activities or business transactions where business problems are known or presumed, the cause of which is presumed to be data defects. For each process activity, at least one process metric should be identified for which measured values are available. In addition, the (monetary) impact of the business problems is to be documented.		
Result Document List of scopes	Roles <ul style="list-style-type: none"> • Sponsor • Corporate data steward • Process owner 	Techniques <ul style="list-style-type: none"> • Interview/workshop • Data quality simulation
Checklist <ul style="list-style-type: none"> • Process activities and business transactions with known or presumed business problems (potentially caused by data defects) have been identified. • Process metrics for monitoring the performance of the identified process activities have been identified. • Business problems and data defects as potential cause have been described. • Costs arising as a result of identified business problems have been described. 		

Table B2-4: Define the scope (Activity I.1)

For each process activity (i.e. as part of every scope) focused on, at least one process metric should be identified for which measured values are available. Furthermore, the business problem is to be described along with the impact of the problem on the process metric and the achievement of the corporate objective focused on. In addition to the focus on concrete process activities as the scope, it is also possible to concentrate on data object types (e.g. product data). This focus is useful if concrete data defects are known, or if the DQM organization calls for a focus on specific data objects.

The head of the DQM organization at Beiersdorf was in charge of applying the method with external support. As sponsor, the head of Supply Chain Management, which DQM also belongs to organizationally, supported the measure (project budget, personnel resources, assistance with requests for company-wide interviews).

In its search for critical data defects, Beiersdorf considered the entire supply chain and limited the scope by focusing on product data as well as intra-organizational and inter-organizational interfaces. The reasons for this definition of the scope were on the one hand the fundamentally good quality of the master data (little criticism, few known problems) and on the other, occasional but concrete criticism of the quality of logistical product data (in particular weights). Strategic corporate objectives (compliance, cost reduction, high service level) have not been refined by concrete process metrics.

Vignette B2-1: Scope at Beiersdorf

The goal of Activity I.2 is to identify the IT systems used in the individual scopes and to obtain a description of the analyzed data which is as precise as possible (see Table B2-5). In interviews with the respective process owners, the data used in a given process activity and the IT systems which deliver or use the data (databases or business applications respectively) are to be identified for each previously identified scope. In the case of data, not only the data object type (e.g. customer data), but also a detailed limitation of the presumably defective data objects (e.g. “all male customers”, “all active products offered in Latin America”) is to be documented.

<p>The goal is to identify the data used in the process activities and the IT systems which store (databases) and/or use (business applications) the data for each scope previously identified, and to concretize these by limiting the data object types already documented. In addition, the type of possible data defects is to be limited by the selection of analyzed data quality dimensions.</p>		
<p>Result document List of scopes, data, IT systems and data quality dimensions</p>	<p>Roles</p> <ul style="list-style-type: none"> • Corporate data steward • Process owner 	<p>Techniques</p> <ul style="list-style-type: none"> • Interview/workshop
<p>Checklist</p> <ul style="list-style-type: none"> • For each scope, the analyzed data have been described as precisely as possible (i.e. data object type and limitations). • For each scope, the databases and business applications used have been identified. • For each scope, the data quality dimensions to be focused on have been selected. 		

Table B2-5: Identify data and IT systems (Activity I.2)

In addition to limiting the data analyzed, the corporate data steward should select several data quality dimensions for the identification of causal chains in order to ensure a focused discussion. The method does not prescribe any selection of data quality dimensions, but provides various lists as a guide. The corporate data steward should se-

lect approximately four to eight dimensions from these suggestions and formulate guiding questions by way of explanation for each dimension.

Beiersdorf has selected approx. 100 data attributes (grouped into 20 data clusters) from the company's product data model consisting of approx. 800 attributes. The strategy of the DQM organization is stated to center on "the timely provision of accurate, complete and consistent data". In order to focus the discussion in planned workshops, the four data quality dimensions timeliness, accuracy, consistency and completeness have therefore been selected.

Vignette B2-2: Selected data attributes and data quality dimensions at Beiersdorf

The goal of Activity I.3 is to identify causal chains with business-critical data defects for which the corporate data steward will subsequently specify data quality metrics in Activity II.1 to monitor them (see Table B2-6). The identification process should be performed top down, i.e. the corporate data steward should try to identify data defects on the basis of known business problems.

The goal is to identify causal chains, i.e. cause-and-effect relationships between data defects, business problems and strategic corporate objectives. For this purpose, business problems are to be identified for each process activity within the scope, and for the each business problem, the impact (e.g. costs) and the data defects causing it.

Result Document	Roles	Techniques
List of causal chains	<ul style="list-style-type: none"> • Business data steward • Corporate data steward • Data user • Technical data steward • Process owner 	<ul style="list-style-type: none"> • Data quality analysis • Data quality simulation • Interview/workshop
Checklist <ul style="list-style-type: none"> • Causal chains have been described in full (i.e. data defects, business problems and impacts). • Causalities (business problem / impact on strategic corporate objective, data defect / business problem) have been described precisely (i.e. quantified as far as possible). 		

Table B2-6: Identify causal chains (Activity I.3)

Interviews or workshops with various experts within the company support the identification process. The technique data quality simulation facilitates the discussion in the interviews and illustrates the correlations discussed by means of simulated DQM scenarios. If the identification of data defects on the basis of business problems (top-down approach) does not lead to the desired result, or if the corporate data steward wants to identify potential data defects in preparation for interviews (bottom-up approach), the technique data quality analysis enables the identification of statistically different and thereby potentially defective data elements.

Applying the method has shown that not just causal chains are identified in Activity I.3, but that possible initiatives for improving data quality are discussed as well. While such information is not necessarily relevant in order to specify data quality metrics, it

is nonetheless documented so that it is available for potential follow-up projects (see Section 2.5.2, preventive and reactive initiatives).

In addition to the corporate data steward and external support, the project team at Beiersdorf, who applied the method, includes a business data steward for product data as well as a technical data steward. The technical data steward assisted with detailed expert knowledge of the functionality of the central product lifecycle management system. A total of 16 data users identified causal chains in seven workshops. Prior to the interviews, the sponsor communicated the necessity for the measure and, in so doing, ensured the availability and active participation of the interviewees.

Vignette B2-3: Identification of causal chains at Beiersdorf

2.5.4.2 Phase II: Analyze and specify

The goal of Activity II.1 is to define a list of requirements, i.e. a company-specific list of requirements to be met by a data quality metric, ranked in order of importance (see Table B2-7). After specifying the metrics in Activity II.2, the corporate data steward checks that the requirements have been met (see Activity III.1). Since these requirements represent what is expected of the data quality metrics, the corporate data steward should agree the list of requirements with the sponsor. Prioritizing the individual requirements also assists the approval and, if necessary, the adaptation of the list of requirements in Activity II.2.

The goal is to define a list of requirements, i.e. a company-specific list of requirements to be met by a data quality metric, ranked in order of importance.		
Result Document List of requirements	Roles <ul style="list-style-type: none"> • Business data steward • Corporate data steward • Sponsor • Technical data steward 	Techniques <ul style="list-style-type: none"> • Interview/workshop
Checklist <ul style="list-style-type: none"> • The defined requirements include all corporate guidelines (e.g. existing metrics systems, reporting templates, standards). • The list of requirements does not contain any conflicting requirements. • The prioritization reflects the importance of the individual requirements. 		

Table B2-7: Define and prioritize requirements (Activity II.1)

One example of such a requirement is that the values of all data quality metrics must be aggregated to one value of one single metric so it can be used for e.g. a balanced scorecard applied on a company-wide basis [Kaplan/Norton 1992]. If this requirement is to be met, it is important to make sure that not only the scale type of a relevant process metric is taken into account when specifying the scale type for a data quality metric. The scale types of existing data quality metrics as well as those of the balanced scorecard metrics must also be considered. Another requirement relates to the use of

specified metrics not just for company-internal purposes, but also for comparison with other companies (benchmarking). If this requirement is to be met, the chief data steward has to make sure that in the process of specifying validation rules only attributes of standardized data models (e.g. fields of an unmodified ERP system) are taken into consideration as company-specific adaptations (e.g. added attributes) cannot be used for measurements in other companies.

The goal of Activity II.2 is to specify data quality metrics for monitoring the previously identified data defects. For this purpose, it is necessary to describe the databases for whose data elements the metrics values are to be measured (measuring points), the data to be validated, the timing of the measurements (measuring frequency), a measuring scale, a unit of measure and a measuring technique (see Table B2-8).

<p>The goal is to specify at least one data quality metric for each previously identified causal chain. For this purpose, it is necessary to describe the databases for whose data elements the metrics values are to be measured (measuring points), the data to be validated, the timing of the measurements (measuring frequency), a measuring scale, a unit of measure and a measuring technique. In terms of measuring techniques, the method proposes validation rules and supports the design of rules and calculation formulas by describing two techniques.</p>		
<p>Result Document List of data quality metrics specifications</p>	<p>Roles</p> <ul style="list-style-type: none"> • Business data steward • Corporate data steward • Technical data steward 	<p>Techniques</p> <ul style="list-style-type: none"> • Data quality simulation • Interview/workshop • Specification of calculation formulas • Specification of validation rules
<p>Checklist</p> <ul style="list-style-type: none"> • Each data quality metric has been specified in full, i.e. the data to be validated, a measuring point, a measuring scale, a unit of measure, a measuring frequency and a measuring technique have all been described. • The specification of the measuring techniques takes into account all previously documented causal chains, unless reasons are stated as to why a causal chain has not been included. • All specified measuring techniques can be implemented at the specified measuring points. 		

Table B2-8: Specify data quality metrics (Activity II.2)

A prime task of the corporate data steward is the specification of measuring techniques, i.e. the design of procedures to investigate whether or not the examined data show one or more of the business-critical data defects identified in Activity I.3. Two types of measuring technique can be distinguished:

- *Data user surveys*. The quality of data can be evaluated by the people who use it in interviews or surveys [Lee et al. 2002, Price et al. 2008]. An interview guideline or a questionnaire is therefore a possible measuring technique that is to be selected in particular when the identified data defects have to be assigned to data quality dimensions such as objectivity, believability or ease of understanding, which are difficult to interpret in technical terms.

- *Validation by means of rules.* The quality of a data object can be measured by examining specific characteristics of the data elements or attributes [Fan et al. 2008, Loshin 2008, pp. 171-197, Olson 2003, pp. 121-254]. If a data object shows all the required characteristics, its quality is 100%. Data characteristics (e.g. last time a data element was modified, completeness of a data element) can be formulated as rules and validated automatically. Validation rules of this kind for verifying the characteristics required for a data object (in a given context) are therefore a possible measuring technique.

Both techniques are fundamentally suitable for checking data for possible defects. However, since interviews or surveys tie up resources for each measurement and consequently cannot be conducted frequently, the presented method focuses on using validation rules as measuring technique. The technique specification of validation rules supports the corporate data steward with the formulation of validation rules by means of rule patterns and examples.

Analysis of the identified causal chains at Beiersdorf resulted in 32 validation rules for 7 problem areas (e.g. dangerous goods indicators, quality of bills of materials, customs information). For each problem area, a calculation formula (same calculation pattern for all metrics) aggregates the results obtained with the assigned validation rules into one metric value on a percentage scale.

Vignette B2-4: Specification of validation rules at Beiersdorf

2.5.4.3 Phase III: Approve and document

Activity III.1 checks whether the requirements defined in Activity II.1 to be met by the data quality metrics specified are fulfilled (see Table B2-9). If high-priority requirements are not fulfilled, the requirements, their prioritization and the metrics specifications have to be reviewed and adapted (i.e. return to Activity II.1).

The goal is to check whether the requirements defined in Activity II.1 are fulfilled. If high-priority requirements are not fulfilled, the requirements, their prioritization and the metrics specifications have to be reviewed and adapted (i.e. return to Activity II.1).		
Result document List of evaluated requirements	Roles <ul style="list-style-type: none"> • Business data steward • Corporate data steward • Technical data steward 	Techniques <ul style="list-style-type: none"> • Interview/workshop
Checklist <ul style="list-style-type: none"> • All high-priority requirements have been fulfilled. • For all unfulfilled requirements, a reason has been stated for each metric that does not fulfill the requirement. 		

Table B2-9: Check the fulfillment of requirements (Activity III.1)

Activity III.2 is the last activity of the procedure model and therefore delivers the result of applying the method. The goal of this activity is to document the specifications

of the data quality metrics as an implementable business concept, i.e. in the form of a document on the basis of which the specified measuring techniques and scales can be implemented (see Table B2-10). Activity III.2 (and with it the entire method) concludes with the release of the documented metrics specification (if necessary, after adjustments) by the sponsor.

The goal is to document the specifications of the data quality metrics as an implementable business concept, i.e. in the form of a document on the basis of which the specified measuring techniques and scales can be implemented.		
Result document Business concept for implementing data quality metrics	Roles <ul style="list-style-type: none"> • Sponsor • Corporate data steward 	Techniques <ul style="list-style-type: none"> • Documentation • Data quality simulation
Checklist <ul style="list-style-type: none"> • All specified data quality metrics have been documented, i.e. in particular measuring techniques and scale for each metric. • A causal chain has been documented for each data quality metric, i.e. the impact of the data defects it monitors. 		

Table B2-10: Document metrics specifications (Activity III.2)

2.6 Evaluation

2.6.1 Overall evaluation approach

The goal of evaluation in a DSR process is to verify whether, and how well, previously defined design objectives (see Table B2-11) have been achieved [Hevner et al. 2004, pp. 85-87, Peffers et al. 2008, p. 56]. Table B2-11 therefore describes evaluation criteria for each defined design objective, which are based on general criteria for evaluating conceptual models [Frank 2007, pp. 123-136].

Design objective	Evaluation criteria	Evaluation technique
DO1: Clarity	EC1.1: Clarity of the method's structuring	Focus group interview (see Figure B2-1, FG.G)
	EC1.2: Sufficient depth versus complexity of the method's components (e.g. meta-model, roles)	Focus group interview (FG.G)
DO2: Effectiveness (method)	EC2.1: Availability of a specification document for data quality metrics	Participative case study (Beiersdorf, CS.4)
DO3: Effectiveness (metrics)	EC3.1: Matching of success factors and method characteristics	Comparison of success factors (Bayer CropScience) with method characteristics
DO4: Efficiency	EC4.1: Costs and benefits of the method's application	Focus group interview (FG.G) Participative case study (CS.4)
	EC4.2: Utility of templates and tools provided by the method	Focus group interview (FG.G) Participative case study (CS.4)
DO5: Flexibility	EC5.1: Adaptability of the method	Iterative method design process (3 participative case studies, CS.1,

Design objective	Evaluation criteria	Evaluation technique
		CS.2, CS.3)
	EC5.2: Clarity of the method's design process	Focus group interview (FG.G)
DO6: Language adequacy	EC6.1: Comprehensibility of the method's terminology and documentation	Focus group interview (FG.G) Iterative method design process (CS.1, CS.2, CS.3) Participative case study (CS.4)
	EC6.2: Application of the complete method	Participative case study (Beiersdorf, CS.4)
	EC6.3: Contribution of involved roles to the method's application	Focus group interview (FG.G) Participative case study (CS.4)
	EC6.4: Acceptance of the method	Focus group interview (FG.G) Participative case study (CS.4)
DO7: Systematic design	EC7.1: Compliance with the meta-model of Method Engineering	Focus group interview (FG.G)
DO8: Utility	EC8.1: Utilization (i.e. implementation) of specified data quality metrics	Participative case study (CS.4)
	EC8.2: Relevance of the method's purpose	Focus group interview (FG.G)

Table B2-11: Evaluation techniques and evaluation criteria

The following three evaluation techniques were used to evaluate the method (see Table B2-11).

- *Focus group interview.* On February 9, 2010, 11 subject matter experts from 10 multinational companies (see Table D-4, Appendix D), each with 3 to 20 years' experience in master data management and DQM, evaluated the method in a focus group interview. After a 45-minute presentation of the method, the panel was asked to discuss and evaluate the method with the aid of several guiding questions (see Table D-5, Appendix D).
- *Participative case study.* Between February 2010 and June 2010, the method was successfully applied at Beiersdorf (see Section 2.5.4). The result, obtained in seven interviews and five project meetings, is the specification of seven data quality metrics with a total of 32 validation rules [Hüner et al. 2011c]. For certain criteria, the iterative method design process is also cited as an evaluation technique since components of the method have been successfully applied in three participative case studies during the course of the method's design process (see Figure B2-1).
- *Comparison of success factors with method characteristics.* Since the beginning of 2008, Bayer CropScience has been using a data quality index (DQI) for measuring the quality of their material data. The DQI is an example where a business-oriented data quality metric has been successfully implemented in an organization. The DQI value simply indicates the ratio of defective data

records to the total data records, with a data record being defective if at least one validation rule has identified a data defect. In the meantime, the DQI has become a well-established data quality metric comprising more than 80 validation rules. Table D-6 (see Appendix D) shows factors identified in retrospect as being critical for effective data quality measurement at Bayer Crop Science [Ebner et al. 2011]. These factors are compared with characteristics of the method in order to evidence the effectiveness of metrics which are specified by the method.

2.6.2 Evaluation results

The clarity of the method's structuring (see Table B2-12, EC1.1) was confirmed by the focus group. There was particular praise for the Method Engineering meta-model as it structures the method into understandable components. In contrast, EC1.2 was discussed quite controversially in the focus group interview. Some experts mentioned the method's complexity as a problem, but were unable to propose any concrete simplifications. One proposal relating to the roles suggested further explication of a customer view (for data users) and a data supplier view (for data stewards) in order to support the identification of causal chains. Another concrete demand referred to the provision of reference metrics (i.e. generic data quality metrics) for company-independent use (e.g. compliance with regulatory provisions and laws).

Evaluation criteria	Evaluation technique	Evaluation results
EC1.1: Clarity of the method's structuring	Focus group interview	Approved
EC1.2: Sufficient depth versus complexity of the method's components (e.g. meta-model, roles)	Focus group interview	Controversial discussion, but approved in the absence of alternatives
EC2.1: Availability of a specification document for data quality metrics	Participative case study (Beiersdorf)	Approved
EC3.1: Matching of success factors and method characteristics	Comparison of success factors (Bayer CropScience) with method characteristics	Partially approved
EC4.1: Costs and benefits of the method's application	Focus group interview	Approved, but considered to be less important
	Participative case study (Beiersdorf)	Approved
EC4.2: Utility of templates and tools provided by the method	Focus group interview	Approved, but with request for extension
	Participative case study (Beiersdorf)	Approved
EC5.1: Adaptability of the method	Iterative method design process	Approved
EC5.2: Clarity of the method's design process	Focus group interview	Approved, but rated as irrelevant
EC6.1: Comprehensibility of the me-	Focus group interview	Approved

Evaluation criteria	Evaluation technique	Evaluation results
thod's terminology and documentation	Iterative method design process (3 participative case studies)	Approved
	Participative case study (Beiersdorf)	Approved
EC6.2: Application of the complete method	Participative case study (Beiersdorf)	Approved
EC6.3: Contribution of involved roles to the method's application	Focus group interview	Approved
	Participative case study (Beiersdorf)	Approved
EC6.4: Acceptance of the method	Focus group interview	Approved
	Participative case study (Beiersdorf)	Approved
EC7.1: Compliance with the meta-model of Method Engineering	Focus group interview	Approved
EC8.1: Utilization (i.e. implementation) of specified data quality metrics	Participative case study (Beiersdorf)	Approved
EC8.2: Relevance of the method's purpose	Focus group interview	Approved

Table B2-12: Evaluation results

As already described, it was possible to demonstrate the effectiveness of the method in the participative case study at Beiersdorf (EC2.1). Comparison of the success factors of the DQI developed at Bayer CropScience and the characteristics of the presented method show that the method addresses the success factors, the characteristics of data quality metrics (SF1, SF4, SF7, and SF9) and their specification process (SF2, and SF6). Even if this comparison does not actually prove the effectiveness of specified data quality metrics (EC3.1), it nonetheless shows that the method does take important characteristics of effective data quality measurements into account in the specification of data quality metrics.

Due to the high relevance assigned to the issue of measuring data quality, the need for a thorough cost-benefit analysis when applying the method (EC4.1) was rated as less important by some focus group experts. Particularly experts from companies operating in highly regulated markets (e.g. pharmaceuticals and chemicals) referred to numerous examples of data defects possibly causing infringement of regulatory or legal provisions (e.g. mandatory documents for admission of crop protection products, completeness and correctness of expiry date formats), which they are already monitoring. However, the related measurements are not conducted at an early stage of the particular product's supply chain (e.g. when data is created or used for the first time) but towards the end (e.g. by quality assurance just before products are shipped to customers). In such cases (i.e. when data quality metrics were able to identify data defects in a timely manner), the respective experts said that cost-benefit analysis was superfluous. In the case of Beiersdorf, the costs of applying the method were less than the potential costs

of a specific data defect type, whose causes could be repaired as a result of the causal chain identification process. However, no cost and benefit analysis was conducted at Beiersdorf prior to the method's application.

While the focus group confirmed the utility of templates and tools provided by the method (EC4.2), the experts called for the examples of causal chains and data quality metrics to be further developed. In particular, it was mentioned that generic metrics for non-company-specific use cases (e.g. industry standards and regulatory requirements) would simplify the specification process of metrics for a certain company. In the case of Beiersdorf, techniques and templates supported the method's application, and during workshops led to good progress in identifying causal chains and specifying validation rules and metrics.

The adaptability of the method (EC5.1) is evidenced by the documentation of the iterative design process (see Table D-2, Appendix D), during the course of which the method was extended and adapted in three design cycles (see Section 2.4.1). Although the clarity of the design process (EC5.2) was confirmed by the focus group, it was not seen as decisive for evaluation of the method. The experts wanted to see further examples demonstrating the method's outcome rather than clarity of the method design process.

The comprehensibility of the method (EC6.1) was also demonstrated during its development stage by applying individual components in three participative case studies. The focus group and application of the entire method at Beiersdorf have corroborated this positive assessment. In both cases, a brief introduction to the terminology and the procedure was sufficient to enable content to be discussed with the experts and design results to be produced collaboratively. EC6.2 was proven by applying the method at Beiersdorf. The communication process expected to be initiated between all the roles involved in the method application process (EC6.3, EC6.4) was judged to be positive by the focus group. Some experts mentioned the lack of communication between business and IT departments regarding data quality requirements as a possible reason for business problems. In their opinion, joint workshops to identify causal chains could help solve this problem. Application of the method at Beiersdorf also confirmed this assessment. All the interviews reflected the constructive collaboration between all the roles involved.

Compliance of the method to the meta-model of Method Engineering (EC7.1) was clearly corroborated by the focus group. The method's utility has been demonstrated by the effective implementation of specified data quality metrics at Beiersdorf (EC8.1). In addition, all experts in the focus group confirmed the relevance of data quality measurements and methodical support for the design of company-specific data quality metrics (EC8.2). They said that their companies would strive for permanent measurement of their master data quality and search for support in this complex task.

Also, alignment of data quality metrics with strategic company objectives and process performance objectives was considered an important design principle by the experts.

2.7 Discussion and outlook

2.7.1 Research results and limitations

This paper presents a method for specifying business-oriented data quality metrics. The procedure model defined by the method describes the identification of causal chains between data defects, business problems and strategic objectives of a company as well as the specification of metrics for monitoring the identified data defects. In addition to the method and its design process, the paper explains the application of the method in a participative case study at Beiersdorf, discussion of the method in a focus group interview, and a comparison of characteristics of the method with success factors for data quality measurements. This multidimensional evaluation verifies whether, and how well, the method fulfills the previously defined design objectives.

Application of the method at Beiersdorf led to useful results with an acceptable level of effort, and shows that the operationalization of identified company-specific data defects by means of validation rules enables the specification of practicable data quality metrics for business. In particular, applying the metrics method ensures the business relevance of the specified metrics and makes technical implementation of the specified validation rules possible. A positive side effect of applying the method (in particular the organization of interdepartmental workshops for identifying causal chains) is improved communication between business departments and IT on the subject of data quality and possible improvement measures.

For the purposes of applying the method, it should be borne in mind that business-oriented data quality metrics as described in this paper cannot be specified exclusively by means of the presented procedure. This method should merely be seen as one possibility for specifying business-oriented data quality metrics. The described case of Bayer CropScience shows that comparable validation rules for measuring data quality were defined during the course of a project aimed at harmonizing business processes. Bayer CropScience now monitors some 80 rules for material data. The monitoring of customer and supplier data is also envisaged.

Furthermore, the method represents a complex artifact that requires some effort in order to be applied effectively. Therefore, in each individual case where the method is to be used it is important to carefully consider which elements of the method are likely to be of benefit and which elements can be regarded as irrelevant. For example, business-oriented data quality metrics might be specified by a single person in a single activity (e.g. supported by techniques and templates of the method, see Table D-3, Appendix D), provided that this person has access to all the information required for this pur-

pose. Thus, the method should not be seen as a monolithic artifact, but rather as a collection of artifacts, each of which can be used in its own right.

2.7.2 Outlook for future research

While the individual characteristics of the respective company are to be taken into account when specifying data quality metrics, it is nonetheless possible to use previously specified metrics as a template in order to reduce the effort required for the specification process. Moreover, metrics for monitoring general or industry-specific data defects can be specified and used by several companies. Legal requirements in respect of accounting, data storage or documentation can be used as the starting point. For this reason, extending the collection of causal chains and data quality metrics is an objective for further research.

Investigating the effectiveness of data quality management measures also provides possibilities for empirical research. In companies that already use data quality metrics to measure the quality of their data, the effect of data cleansing or improved data maintenance processes on the measured values obtained with data quality metrics and on performance of the respective business processes can be examined. Research findings relating to the cost and effectiveness of individual measures would be of interest for practice.

Up to now, the presented method has only been used for identifying defects in structured data within companies (mostly referred to as master data, e.g. customer address data, material data). Also using the method for unstructured data (e.g. comments and posts in social media) can therefore be a goal for further research. While the concept of the measuring technique proposed by the method (i.e. validation rules) does not mandatorily require a relational data model, it does, however, call for data attributes that have been specified in some way as a basis for specifying validation rules. If no such structure exists, specifying validation rules might be difficult. However, text mining algorithms which extract the information from unstructured texts and store it in a defined data structure for measurements might be helpful in such cases.

The long-term application of business-oriented data quality metrics in companies should lead to an improvement in the quality of the data monitored as well as to fewer business problems. The ultimate objective of measuring data quality measurement is, after all, not the measurement itself but the improvement in data quality. In the case of empirical investigations, however, it should be borne in mind that data defects are not the only cause of business problems.

3 Methode zur Spezifikation geschäftsorientierter Datenqualitätskennzahlen

Titel	Methode zur Spezifikation geschäftsorientierter Datenqualitätskennzahlen
Autoren	Kai M. Hüner Universität St. Gallen, Institut für Wirtschaftsinformatik Müller-Friedberg-Strasse 8, 9000 St. Gallen, Schweiz kai.huener@unisg.ch
Publikationsorgan	Arbeitsbericht, Institut für Wirtschaftsinformatik, Universität St. Gallen, St. Gallen 2008.
Status	Veröffentlicht

Tabelle B3-1: Bibliographische Angaben zum Beitrag „Methode zur Spezifikation geschäftsorientierter Datenqualitätskennzahlen“

Zusammenfassung

Der Qualität von Daten (insbesondere von Stammdaten) kommt in Unternehmen zunehmend grössere Bedeutung zu. Ursachen sind die steigende Zahl behördlicher und gesetzlicher Auflagen sowie die wachsende Bedeutung von Informationssystemen zur Entscheidungsunterstützung, die Anforderungen an die Qualität von Daten (z. B. Aktualität, Genauigkeit, Vollständigkeit) stellen. Effektives Datenqualitätsmanagement benötigt Kennzahlen, die die Erfüllung solcher geschäftsorientierter Datenqualitätsanforderungen überwachen und Datendefekte frühzeitig erkennen. Die individuelle Gestaltung von Datenqualitätskennzahlen unter Berücksichtigung unternehmensspezifischer Geschäftsprozesse, Führungssysteme und Anwendungslandschaften gewährleistet den Geschäftsbezug der Kennzahlen und ermöglicht Datenqualitätsmessungen, die geschäftskritische Datendefekte rechtzeitig erkennen.

Der Bericht beschreibt eine Methode zur Identifikation von Kausalitäten zwischen Datendefekten, Geschäftsproblemen und strategischen Unternehmenszielen sowie zur Spezifikation von Kennzahlen für die Überwachung der identifizierten Datendefekte. Die Methode folgt einem Top-Down-Vorgehen und spezifiziert Kennzahlen nur für solche Datendefekte, die zu kritischen Geschäftsproblemen führen können. Als Messverfahren schlägt die Methode Validierungsregeln vor. Ein Messsystem (nicht im Fokus der Methode) kann durch die Implementierung dieser Regeln Eigenschaften von Daten überprüfen und dadurch Datendefekte erkennen.

Der Beitrag beschreibt die Methode als Ergebnis gestaltungsorientierter Forschung und erläutert verschiedene Methodenelemente durch Beispiele der Methodenanwendung in drei Unternehmen. Die Beispiele verdeutlichen, dass nicht die Nutzung aller Methodenelemente notwendig ist, sondern die Methode für eine konkrete Anwendung konfiguriert werden kann. Zur Unterstützung weiterer Anwendungen der Methode enthält der Anhang der Arbeit verschiedene durch Fallstudien, Fokusgruppeninterviews und eine Literaturrecherche gesammelte Beispiele für Anforderungen an Daten-

qualitätskennzahlen, Datenqualitätsdimensionen, Kausalketten, Kostenarten defekter Daten und Validierungsregeln.

3.1 Einleitung

Hohe Datenqualität ist ein kritischer Erfolgsfaktor für Unternehmen, um zunehmend starkem Wettbewerb im Markt sowie vermehrten regulatorischen Anforderungen begegnen zu können. Die folgenden Beispiele verdeutlichen die strategische Bedeutung qualitativ guter Unternehmensdaten.

Das Standardisierungsgremium Global Standards One (GS1) spezifiziert u. a. die Global Trade Item Identification Number (GTIN) zur weltweit eindeutigen Identifikation von Produkten [GS1 2010]. Externe Dienstleister bieten basierend auf diesem von GS1 bereitgestellten Nummernsystem sogenannte Daten-Pools zur multilateralen Synchronisation von Produktstammdaten an [Nakatani et al. 2006].

Einzelhandelsunternehmen und Konsumgüterhersteller profitieren von der Nutzung von Daten-Pools, die eine überbetriebliche Synchronisation von Stammdaten unterstützen [Schemm 2008, S. 98-103]. Hersteller, die Stammdaten zu ihren Produkten in einem solchen Pool bereitstellen, müssen jedoch jede warenwirtschaftlich relevante Komponente eines Artikels kennzeichnen und dabei Richtlinien von GS1 (z. B. welche Eigenschaften von Produkten mit gleicher GTIN wie stark variieren dürfen) beachten. So müssen Produkte mit gleichem Inhalt (z. B. Shampoo-Flaschen), deren Gewicht sich aber um mehr als 20 Prozent unterscheidet, verschiedene GTINs haben. Die Einhaltung solcher Vorgaben ist notwendig, um Probleme beim Transport (Paletten überschreiten zulässiges Gewicht aufgrund höherer Produktgewichte) und beim Kunden (Produkte passen bei abweichenden Abmessungen nicht in vorgesehene Regale) zu vermeiden.

Die Bereitstellung ressourcenintensiver Telekommunikationsprodukte wie IP-TV (über das Internet übertragene Fernsehprogramme) erfordert Transparenz und Flexibilität der genutzten Netzinfrastruktur: Bei einer Bestellung ist möglichst schnell zu entscheiden, ob das bestellte Produkt mit der bestehenden Infrastruktur angeboten werden kann oder in welchem Zeitraum erforderliche Infrastrukturerweiterungen realisierbar sind. Bei dieser Verfügbarkeitsprüfung müssen Netzinfrastruktur-, Kunden- und Produktdaten in Beziehung gesetzt und analysiert werden um zu entscheiden, ob eine angefragte Leitung grundsätzlich und unter Berücksichtigung der bereits verkauften (d. h. vertraglich zugesicherten) Bandbreite in der Lage ist, das angefragte Produkt bereitzustellen.

Falsche oder zu späte Antworten auf eine Verfügbarkeitsanfrage können dazu führen, dass Kunden einen anderen Anbieter wählen. Schlechte Datenqualität (z. B. unvollständige Dokumentation der Netzinfrastruktur) kann somit unmittelbar die Gewinnung von Neukunden gefährden.

Trotz der offensichtlichen strategischen Bedeutung von Datenqualität in den geschilderten Beispielen ist ein Zusammenhang zwischen Datenqualität und der Erreichung von Unternehmenszielen nicht immer einfach zu identifizieren. Ein Grund dafür ist die Multidimensionalität und Kontextabhängigkeit von Datenqualität [Wang/Strong 1996, S. 6]: Verschiedene Datenqualitätsdimensionen (DQ-Dimensionen, z. B. Aktualität, Genauigkeit, Vollständigkeit) sind für verschiedene Datenobjekttypen (z. B. Kundendaten, Produktdaten) zu unterscheiden. Und zur Festlegung von Sollwerten für einzelne DQ-Dimensionen ist der jeweilige Verwendungskontext der Daten (z. B. die Prüfung der Verfügbarkeit einer bestimmten Bandbreite auf einer Telekommunikationsleitung) zu analysieren. Bei der Suche nach aussagekräftigen Sollwerten ist ausserdem eine Kosten-Nutzen-Betrachtung erforderlich [Lee et al. 2006, S. 13-18]: Massnahmen zur Qualitätsverbesserung der im Beispiel beschriebenen Infrastrukturdaten sollten beispielsweise nur dann durchgeführt werden, wenn die Kosten der Massnahmen den erwarteten Wert zusätzlicher Neukunden nicht übersteigen.

Voraussetzung für die beschriebene Kosten-Nutzen-Betrachtung sowie die Definition von Sollwerten für DQ-Dimensionen ist die Identifikation von Ursache-Wirkungs-Beziehungen (Kausalitäten) zwischen Datendefekten (z. B. fehlende Infrastrukturdaten), Geschäftsproblemen (z. B. falsche Auskunft zur Verfügbarkeit eines Produkts) und Kosten (z. B. Wert eines Neukunden). Auf Basis dieser Kausalketten können geschäftsorientierte Datenqualitätskennzahlen (DQ-Kennzahlen) spezifiziert werden, d. h. Kennzahlen zur Messung von DQ-Dimensionen, die kritisch sind für den Geschäftserfolg eines Unternehmens.

Es gibt verschiedene Forschungsarbeiten, die Datenqualität allgemein sowie DQ-Kennzahlen im Besonderen untersuchen und konkrete Messverfahren vorschlagen (vgl. Abschnitt 3.4). Ausserdem existieren Arbeiten mit Anforderungen an DQ-Kennzahlen und mit Handlungsempfehlungen für die Gestaltung solcher Kennzahlen. Es fehlt jedoch ein strukturiertes Vorgehen zur Identifikation der genannten Kausalitäten und zur Spezifikation von Kennzahlen für die Überwachung der identifizierten Datendefekte. Aus dieser Lücke ergibt sich das Forschungsziel der Gestaltung einer Methode a) zur Identifikation von Kausalitäten zwischen Datendefekten, Geschäftsproblemen und strategischen Zielen und b) zur Spezifikation von DQ-Kennzahlen für die Überwachung der identifizierten Datendefekte.

Das Forschungsziel ist insbesondere für Unternehmen mit einem unternehmensweiten Datenqualitätsmanagement (DQM) von Interesse: Hier gilt es, nicht nur einzelne Kausalitäten wie die Anfangs erläuterten Beispiele zu identifizieren, sondern die Methode

mehrfach anzuwenden und verschiedene, miteinander in Beziehung stehende Kausalitäten zu identifizieren: Eine Qualitätsverbesserung logistischer Produktdaten kann sich nicht nur positiv auf Transportprozesse auswirken, sondern ausserdem die Anzahl von Beschwerden und Reklamationen reduzieren.

Nach der Erläuterung konzeptioneller Grundlagen in Abschnitt 3.2 und einer kurzen Beschreibung des Forschungsprozesses in Abschnitt 3.3 stellt Abschnitt 3.4 ähnliche Arbeiten vor und diskutiert deren Ergebnisse in Bezug auf das Forschungsziel dieser Arbeit. Abschnitt 3.5 stellt die Methode zur Spezifikation geschäftsorientierter DQ-Kennzahlen als Forschungsergebnis vor. Methodenelemente und Beispiele der Methodenanwendung werden dabei gemeinsam beschrieben, um den Zweck der einzelnen Elemente besser illustrieren zu können. Abschnitt 3.6 diskutiert Einschränkungen der Methodenanwendung sowie weiteren Forschungsbedarf.

3.2 Grundlagen

Die vorgestellte Methode ist Ergebnis gestaltungsorientierter Forschung, die das Metamodell des Methodenkerns des Business Engineering (BE-Methodenkern) als konzeptionellen Bezugsrahmen nutzt. Abschnitt 3.2.1 gibt daher einen kurzen Überblick zu BE und ordnet die Konzepte der Methode in das Metamodell des BE-Methodenkerns ein. Ziel der Nutzung von DQ-Kennzahlen (Ergebnis der Methodenanwendung) ist es, Datenqualität objektiv zu bewerten und vergleichbare Datenqualitätsmessungen (DQ-Messungen) als präventive DQM-Massnahme zu ermöglichen. Abschnitt 3.2.2 erläutert daher kurz das der Arbeit zugrunde liegende Verständnis von Datenqualität und DQM.

3.2.1 Business Engineering als Bezugsrahmen

BE ist ein systematisches Vorgehen, das Unternehmen bei der Durchführung von Transformationsprojekten (z. B. strategische Neuausrichtung, Gestaltung strategiekonformer Geschäftsprozesse) unterstützt. BE fokussiert dabei nicht isolierte Aspekte der Unternehmenstransformation (z. B. strategische Führung, Aufbauorganisation), sondern strukturiert eine komplexe Gestaltungsaufgabe auf verschiedenen Modellierungsebenen durch beherrschbare Aktivitäten mit definierten Ergebnissen und Beziehungen [Österle 1995, Scheer 1998].

Der St. Galler Ansatz des BE [Österle 1995, Österle/Blessing 2003, S. 80-84, Österle/Blessing 2005, S. 12-15] ist der methodische Bezugsrahmen der vorliegenden Arbeit. Der Ansatz strukturiert Transformationsprojekte auf den Modellierungsebenen *Strategie*, *Prozess* und *System* durch Methoden (vgl. Abschnitt 3.3.2) und spezifiziert Gestaltungsobjekte der Ergebnisdokumente (Ergebnisse der Anwendung einer Methode) in einem Metamodell. HÖNING [2009] identifiziert den Kern (d. h. ähnliche Elemente) verschiedener BE-Ansätze, strukturiert diesen durch Konzepte des Methoden-

in Abbildung B3-1 dargestellten Konzepte durch die Ergebnisdokumente. Abbildung B3-1 zeigt lediglich die Beziehungen zentraler Konzepte der vorgestellten Methode zu Gestaltungsobjekten des BE-Methodenkerns. Anhang E.1 beschreibt die erweiterten Metaentitätstypen, die Gestaltungsobjekttypen des BE-Methodenkerns spezifiziert HÖNING [2009, S. 252-257].

3.2.2 Datenqualität und Datenqualitätsmanagement

Datenqualität ist Gegenstand verschiedener Forschungsarbeiten mit verschiedenen Forschungsansätzen. Einige Arbeiten schlagen Listen und Kategorien mit DQ-Dimensionen als Ergebnis empirischer Forschung vor [Wang/Strong 1996], wohingegen andere Arbeiten Erfahrungen aus Berufs- und Beratungspraxis beschreiben [Redman 1996, English 1999, Loshin 2001] oder DQ-Dimensionen als Ergebnis theoretischer Schlussfolgerungen [Wand/Wang 1996, Price/Shanks 2005] vorstellen.

DQM umfasst Massnahmen zur Verbesserung der Qualität von Daten [Batini/Scannapieco 2006, S. 69-71], geht aber über die rein reaktive Verbesserung der Datenqualität (d. h. Identifikation und Bereinigung von Datendefekten) hinaus [Shankaranarayanan/Cai 2006, S. 303-304]. Vielmehr beinhaltet DQM die proaktive und präventive Verbesserung der Datenqualität durch einen kontinuierlichen Kreislauf zur Definition, Messung, Analyse und Verbesserung der Datenqualität sowie die Gestaltung der dazu erforderlichen Rahmenbedingungen [Wang et al. 1998, English 1999, S. 69-81, Eppler/Helfert 2004]. Eine Übersicht zu verschiedenen Ansätzen für DQM liefern BATINI ET AL. [2009].

3.3 Forschungsmethodik

Die Beschreibung des Entstehungsprozesses der vorgestellten Methode als Artefakt [Simon 1996] ermöglicht die Reproduzierbarkeit und Bewertung der Methode als Ergebnis wissenschaftlicher Forschung. Dabei ist zu beachten, dass die vorliegende Arbeit das Gestaltungsziel der Wirtschaftsinformatik und weniger das Erkenntnisziel [Frank 1997, Becker et al. 2003] fokussiert: Forschungsziel ist die Gestaltung einer Methode zur Spezifikation geschäftsorientierter DQ-Kennzahlen. Zwar ist für die Erreichung dieses Ziels auch Erkenntnisgewinn, z. B. die Identifikation der im Metamodell der Methode (vgl. Abschnitt 3.5.1) beschriebenen Zusammenhänge zwischen Datendefekten, Geschäftsproblemen und DQ-Kennzahlen, notwendig [vom Brocke 2003, S. 2]. Im Vordergrund steht jedoch die Methode als Forschungsergebnis und der Nutzen der durch ihre Anwendung gestalteten Ergebnisse. Abschnitt 3.5.2 beschreibt dazu Ergebnisse der Methodenanwendung in drei partizipativen Fallstudien [Baskerville 1997] die zeigen, dass die Anwendung der Methode mit vertretbarem Aufwand nützliche Ergebnisse erzeugen kann.

3.3.1 Forschungsprozess

Die vorgestellte Methode ist das Ergebnis gestaltungsorientierter Forschung, die Design Science Research als forschungsmethodischem Paradigma folgt. Ziel von Design Science Research ist die Gestaltung von Artefakten als Lösung praktischer Probleme [March/Smith 1995, S. 256-258, Hevner et al. 2004, S. 76]. Vorgeschlagene Artefakttypen sind Konstrukte, Modelle, Methoden (Typ des vorgestellten Artefakts, vgl. Abschnitt 3.3.2) und Instanziierungen.

Den Forschungskontext des Gestaltungsprozesses der Methode bildet das Kompetenzzentrum Corporate Data Quality¹⁵ (CC CDQ). Das CC CDQ ist Teil des Forschungsprogramms Business Engineering¹⁶ der Universität St. Gallen und entwickelt seit 2006 als Kooperation des Instituts für Wirtschaftsinformatik¹⁷ mit verschiedenen Unternehmen¹⁸ Lösungen zur Unterstützung des Qualitätsmanagements konzernweit genutzter Daten [Österle/Otto 2010, Otto et al. 2010, Otto/Österle 2010].

Gemäss verschiedenen Anforderungen an die Gestaltung von Artefakten [Hevner et al. 2004, S. 82-90] beinhaltet der Prozess zur Gestaltung eines Artefakts Phasen der Problemidentifikation, der Definition von Gestaltungszielen, der Gestaltung von Artefakten als Problemlösung und der Evaluation der Artefakte [Gregor 2006, S. 628-629, Peffers et al. 2008, S. 52-56]. Fokusgruppeninterviews [Morgan/Krueger 1993, Rosemann/Vessey 2008] mit Fachexperten verschiedener Unternehmen des CC CDQ und partizipative Fallstudien [Baskerville 1997] haben die Phasen der Problemidentifikation, Zieldefinition und Evaluation des Gestaltungsprozesses der vorgestellten Methode unterstützt. Beispiele der Methodenanwendung in den Fallstudien erläutern in Abschnitt 3.5.2 einzelne Elemente der Methode. Abschnitt 3.6 erläutert die Evaluation der Methode in einem Fokusgruppeninterview.

3.3.2 Methoden-Engineering

Eine Methode ist nach dem Verständnis des Methoden-Engineering ein systematisches Vorgehen zur Entwicklung von Informationssystemen (IS) [Heym/Österle 1993, S. 345]. Methoden-Engineering ermöglicht die Gestaltung von Methoden unter Anwendung der Prinzipien von Methoden, also dem ingenieurmässigen Entwurf von IS [Heym/Österle 1993, S. 347, Gutzwiller 1994, S. 11]. Eine Methode beinhaltet eine Produktbeschreibung und eine Prozessbeschreibung [Olle et al. 1988, S. 2-3, Heym/Österle 1993, S. 349, Brinkkemper 1996, S. 275-276]. Die Prozesselemente leiten dabei die Gestaltung des Ergebnisses der Methode an: Sie definieren, in welcher Reihenfolge welche Elemente des Metamodells zu instanzieren sind und wer die für

¹⁵ <http://cdq.iwi.unisg.ch>

¹⁶ <http://www.iwi.unisg/behsg/>

¹⁷ <http://www.iwi.unisg>

¹⁸ <http://cdq.iwi.unisg.ch/index.php?id=144>

die Instanziierung benötigten Informationen liefert oder die jeweiligen Aktivitäten durchführt.

Die vorgestellte Methode ist als Instanziierung eines von GUTZWILLER [1994, S. 13] vorgeschlagenen Metamodells (s. Abbildung B3-2) gestaltet worden, dass die Elemente der folgenden Teilmodelle einer Methode definiert. Eine vergleichbare Strukturierung beschreiben NUSEIBEH ET AL. [1996].

- *Vorgehensmodell.* Eine Aktivität ist eine Verrichtungseinheit, die darauf abzielt, ein oder mehrere definierte Ergebnisse zu gestalten. Aktivitäten können hierarchisch strukturiert und Teil einer Ablauffolge sein. Alle Ablauffolgen gemeinsam bilden das Vorgehensmodell der Methode.
- *Rollen.* Menschen oder Gremien, die eine bestimmte Rolle bekleiden, nehmen Aktivitäten wahr. Eine Rolle fasst aus Sicht eines Auftraggebers verschiedene Aktivitäten zusammen. Rollen sind an Aktivitäten immer in einer bestimmten Form beteiligt (z. B. *verantwortlich* oder *beratend*).
- *Dokumentationsmodell.* Aktivitäten verwenden Ergebnisse und erzeugen oder modifizieren Ergebnisse. Wie Aktivitäten können Ergebnisse hierarchisch strukturiert sein. Alle Ergebnisse gemeinsam bilden das Dokumentationsmodell der Methode.
- *Metamodell.* Das Metamodell spezifiziert die Gestaltungsobjekte der Ergebnisse. Im Sinne einer formalen Sprache definiert es Syntax und Semantik der Ergebnisse.
- *Techniken.* Techniken beschreiben, wie Ergebnisse erzeugt werden. Im Gegensatz zum Vorgehensmodell, das festlegt, wann und in welcher Abfolge welche Ergebnisse zu gestalten sind, beschreiben Techniken, wie Ergebnisse gestaltet werden.

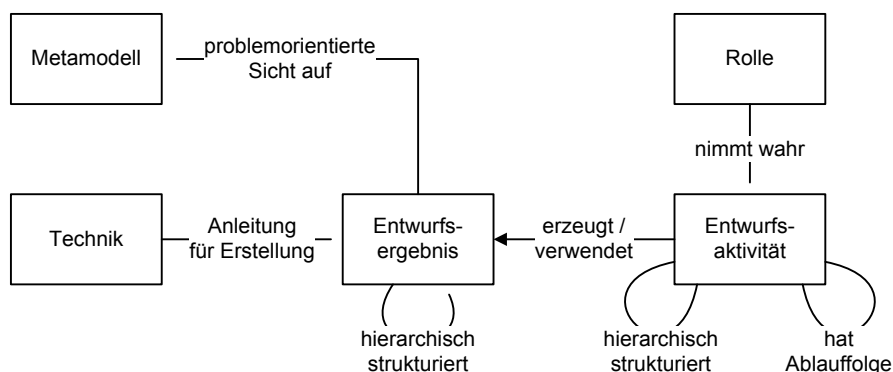


Abbildung B3-2: Metamodell des Methoden-Engineering [Gutzwiller 1994, S. 13]

3.4 Stand der Forschung

Neben der Differenzierung verschiedener DQ-Dimensionen (vgl. Abschnitt 3.2.2) sind DQ-Messungen, also die Bewertung einzelner Dimensionen durch objektiv ermittelte Messwerte, Gegenstand verschiedener Forschungsarbeiten. Die folgenden Abschnitte stellen Ansätze vor, die konkrete DQ-Kennzahlen, Messverfahren, Vorgehen zur Identifikation von Datendefekten oder Berechnungsvorschriften für DQ-Kennzahlen beschreiben.

Einige der diskutierten Beiträge unterscheiden die Konzepte *Daten* und *Information* und entsprechend *Daten-* und *Informationsqualität*. Die vorgestellte Methode betrachtet *Daten* als in Datenbehältern gespeicherte und von Datennutzern genutzte Repräsentation von Geschäftsobjekten (vgl. Abschnitt 3.2.1). *Daten* und *Information* werden dabei nicht unterschieden, wie in anderen Arbeiten mit ähnlichem Fokus ebenfalls [Wang et al. 1998, Bourdreau/Couillard 1999, S. 59, Pipino et al. 2002, S. 212, McKinney Jr./Yoos 2010, S. 331]. Für eine differenzierte Diskussion zur Unterscheidung der Konzepte sei auf andere Arbeiten verwiesen [Bourdreau/Couillard 1999, Spiegler 2000, Boisot/Canals 2004, McKinney Jr./Yoos 2010].

3.4.1 Batini et al.

BATINI/SCANNAPIECO [2006, S. 19-35] beschreiben verschiedene DQ-Dimensionen und erläutern diese durch Berechnungsvorschriften. Ausserdem beschreiben sie verschiedene Kosten-Nutzen-Klassifikationen [S. 89-94] und erläutern die Notwendigkeit, geschäftskritische Datendefekte zu identifizieren und durch DQ-Kennzahlen zu überwachen [S. 184-185]. Eine vergleichende Analyse verschiedener Ansätze zur Messung und Verbesserung von Datenqualität erläutern BATINI ET AL. [2009]. BATINI ET AL. [2007] beschreiben ein methodisches Vorgehen mit den folgenden vier Phasen, um Datendefekte (org. data quality risk) in Datenobjekten eines Unternehmens zu identifizieren und zu überwachen:

- *Data Quality Risk Prioritization*. Modellierung von Organisationseinheiten, Prozessen und genutzten Daten sowie der Beziehungen, für die Datendefekte bekannt sind oder vermutet werden.
- *Data Quality Risk Identification*. Identifikation konkreter Geschäftsprobleme mit Datendefekten als Ursache.
- *Data Quality Risk Measurement*. Qualitative und quantitative Messung der Datenqualität (nicht weiter differenziert) und Bewertung der Auswirkung von Datendefekten.
- *Data Quality Risk Monitoring*. Periodische DQ-Messung, d. h. Überwachung und Bewertung der identifizierten Datendefekte.

3.4.2 Caballero et al.

CABALLERO ET AL. [2007] definieren ein Metamodell für DQ-Messungen und -Kennzahlen, das auf dem ISO/IEC-Standard 15939 (vgl. Abschnitt 3.4.7) aufbaut. In einer weiteren Arbeit nutzen CABALLERO ET AL. [2008] das Metamodell zur Beschreibung einer Methodik für DQ-Messungen, die in vier Phasen mit 16 Aktivitäten ein Vorgehen zur Etablierung von DQ-Messungen, zur Planung eines DQ-Messprozesses, zur Durchführung von DQ-Messungen und zur Evaluation bestehender Messprozesse beschreibt. Die vier Phasen beinhalten

- *Establish and Sustain the Data Quality Measurement Commitment.* Aufbau einer für die Spezifikation, Nutzung und Pflege von DQ-Kennzahlen verantwortlichen Gruppe von Mitarbeitern.
- *Plan the Data Quality Measurement Process.* Identifikation von Geschäftsproblemen, betroffener Prozessaktivitäten, Daten und Rollen, Identifikation und Priorisierung von Anforderungen an genutzte Daten, Identifikation messbarer Eigenschaften der Daten, Spezifikation von DQ-Kennzahlen zur Überwachung der Eigenschaften und Auswahl der zu prüfenden Datenobjekte.
- *Perform the Data Quality Measurement Process.* Integration von Prozessen für DQ-Messungen in bestehende DQM-Prozesse, Spezifikation von Datenqualitätsberichten zur Kommunikation der Messergebnisse, Durchführung der DQ-Messungen und Kommunikation der Messergebnisse.
- *Evaluate Data Quality Measurement Process.* Kontinuierliche Verbesserung der DQ-Kennzahlen und -Messungen.

3.4.3 DAMA

Die Data Management Association (DAMA) definiert neun Funktionen für Datenmanagement [DAMA 2009, S. 291-317] und beschreibt in der Funktion DQM die Definitionen von DQ-Kennzahlen [S. 298-300] und die Definition von Validierungsregeln (org. data quality business rules) [S. 300-302] als zwei von zwölf Aktivitäten. Die Definition von DQ-Kennzahlen umfasst die folgenden sechs Schritte.

- Auswahl eines zuvor identifizierten Geschäftsproblems.
- Identifikation zugehöriger Datenobjekttypen, Datennutzungsprozesse und Datenpflegeprozesse.
- Auflistung der an die identifizierten Daten gestellten Anforderungen (s. Anhang E.5.1).

- Identifikation beeinträchtigter DQ-Dimensionen für jede Anforderung, Spezifikation von Validierungsregeln um zu prüfen, ob Datenobjekte die Anforderungen erfüllen.
- Beschreibung des Messprozesses für jede Validierungsregel, d. h. wie die durch die Regeln definierten Eigenschaften eines Datenobjekts zu prüfen sind.
- Spezifikation eines Schwellwerts für jede Validierungsregel.

3.4.4 Larry P. English

ENGLISH [1999, S. 24] definiert Datenqualität (org. information quality) als „consistently meeting knowledge worker and end-customer expectations“ und ergänzt damit die durch WANG/STRONG [Wang/Strong 1996, S. 6] geprägte Kontextabhängigkeit von Datenqualität um das Element der konsistenten Bereitstellung von Daten. Konsistenz umfasst dabei a) die Erfüllung der Anforderungen aller (nicht nur bestimmter) Datennutzer und b) die Bereitstellung gleicher Daten (auch aus verschiedenen Datenbehältern) für gleiche Geschäftsobjekte.

Weiter definiert ENGLISH [1999, S. 19] Information als Aggregat aus Daten, klarer Bedeutung der Daten (Definition) und verständlicher Darstellung (Präsentation) und gibt für die drei Komponenten verschiedene DQ-Dimensionen (s. Anhang E.2.1) an. Die Differenzierung findet sich auch in der vorgeschlagenen Methodik *Total Quality data Management* (TQdM) zur nachhaltigen Verbesserung von Datenqualität [English 1999, S. 69-82], die Prozesse für Qualitätsmessungen von Datendefinitionen (vgl. Abbildung B3-3, P1) und Daten (P2) unterscheidet. Zur Messung der Auswirkung schlechter Datenqualität (org. nonquality information costs, P3) definiert English [1999, S. 209-212] verschiedene Kostenarten defekter Daten (s. Anhang E.3.1). Abbildung B3-4 zeigt die einzelnen Aktivitäten der Prozesse für DQ-Messungen.

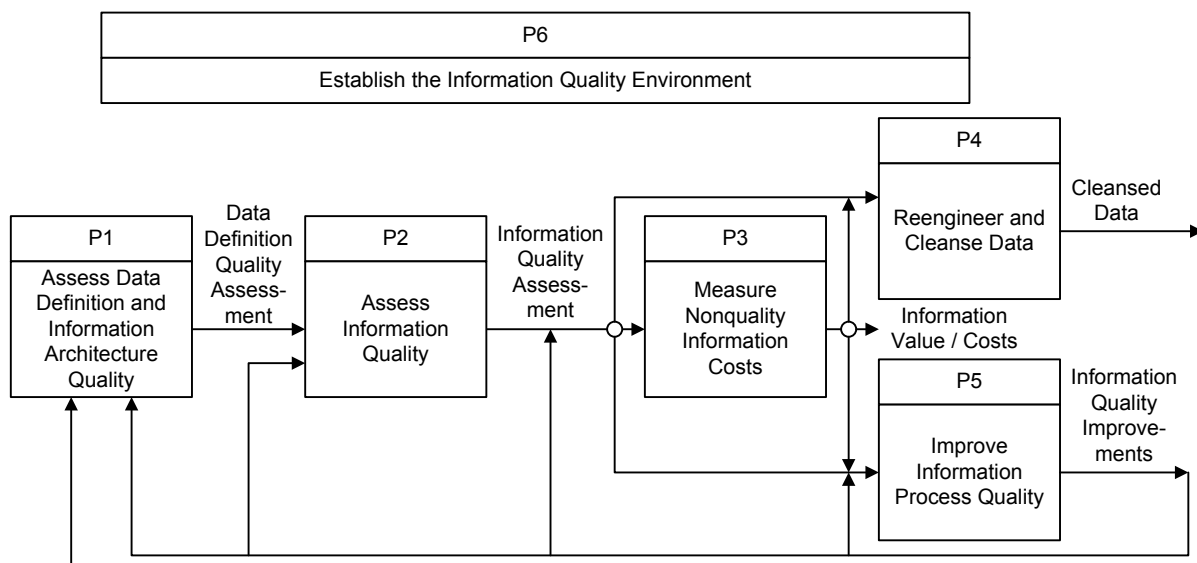


Abbildung B3-3: Prozesse der TQdM-Methodik [English 1999, S. 70]

Da Datendefinitionen und Informationsarchitektur (im Folgenden vereinfachend *Metadaten*) die Basis für bestimmte DQ-Messungen in P2 bilden, sind einzelne Aktivitäten von P1 vor dem Start von P2 durchzuführen (vgl. Abbildung B3-4, S1.1, S1.2 und S2.1). Neben den Aktivitäten für die eigentlichen DQ-Messungen (S1.4, S1.5, S1.6 und S2.7) umfassen die Messprozesse Aktivitäten zur Einschränkung eines Betrachtungsbereichs (S1.2, S2.1), zur Identifikation von Anspruchsgruppen (S1.3), zur Zielwert- und Kennzahlenspezifikation (S2.2) sowie zur Gegenüberstellung von Kosten und Nutzen der Datenhaltung und -nutzung [English 1999, S. 118-136, 155-197].

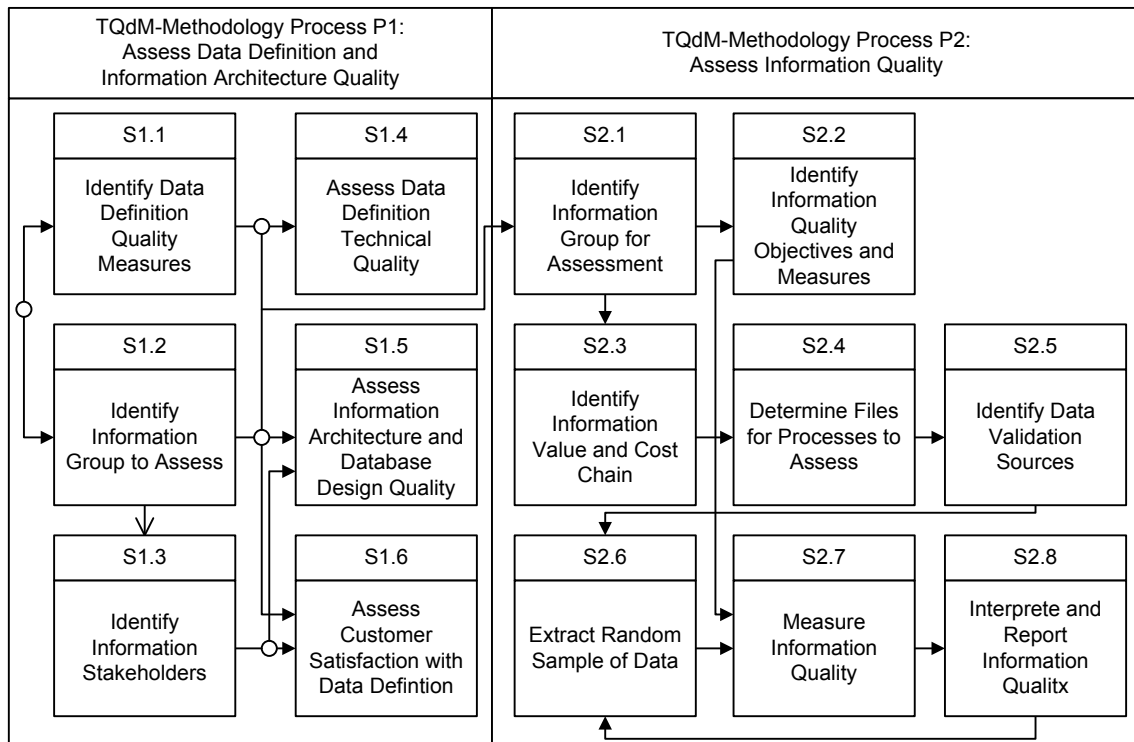


Abbildung B3-4: Prozess zur Messung von Datenqualität [English 1999, S. 72-76]

3.4.5 Heinrich et al.

KAISER ET AL. [2007, S. 3-4] definieren sechs Anforderungen an DQ-Kennzahlen (s. Anhang E.5.2, ebenfalls beschrieben von HEINRICH ET AL. [2007, 2008]) und definieren Kennzahlen für die Dimensionen *Aktualität* (org. *Timeliness* bzw. *Currency*) [Heinrich et al. 2007, Heinrich et al. 2009], *Richtigkeit* (org. *Correctness*) und *Vollständigkeit* (org. *Completeness*) [Heinrich et al. 2008].

3.4.6 IBM

ALUR ET AL. [2007, S. 4-19] beschreiben ein methodisches Vorgehen zur Identifikation von Datendefekten. In der ersten Phase (*Data Profiling*) werden Häufigkeitsverteilungen gleicher Datenelemente berechnet, Primär- und Fremdschlüsselintegrität analysiert und die Konsistenz von Datentypen und -formaten geprüft. Die Ergebnisse dieser Phase (z. B. Liste ungewöhnlicher Datenelemente, Liste mit nötigen Datentypkorrek-

turen, Primär- und Sekundärschlüssel-Listen) dienen als Input für die zweite Phase (*Rule Validation*), in der Validierungsregeln definiert und geprüft werden.

3.4.7 ISO / IEEE

Der Standard 15939 der International Organization for Standardization und der International Electrotechnical Commission (ISO/IEC) [DIN/EN/ISO 2000] definiert einen Messprozess für System- und Software-Entwicklung (vgl. Abschnitt 3.4.2 für Adaption auf DQ-Messungen). Das Prozessmodell umfasst Aktivitäten zur Spezifikation benötigter Informationen für Messungen, zur Durchführung von Messungen und zur Prüfung der Messergebnisse. Neben dem Prozessmodell enthält der Standard ein Glossar verwendeter Begriffe (vgl. Abschnitt 3.4.2, Adaption der Terminologie für DQ-Messungen).

Der Standard 1061 des Institute of Electrical and Electronics Engineers (IEEE) [1998] definiert ein methodisches Vorgehen zur Identifizierung, Implementierung, Analyse und Validierung von Qualitätsmetriken für Software und Software-Entwicklungsprozesse. Neben der Methodik selbst definiert der Standard zahlreiche für die Beschreibung der Methodik notwendigen Begriffe basierend auf dem IEEE-Glossar für Software-Engineering [IEEE 1990]. Die beiden IEEE-Standards bilden die konzeptionelle Basis für den oben genannten ISO/IEC-Standard.

3.4.8 David Loshin

LOSHIN [2001, S. 48-49] definiert Datenqualität entsprechend der Definition von WANG/STRONG [1996] als „fitness for use“ und differenziert dieses Verständnis durch die Aspekte *Defektfreiheit* (org. *Freedom of Defects*) und *Anforderungserfüllung* (org. *Possesses Desirable Characteristics*) sowie verschiedene DQ-Dimensionen (s. Anhang E.2.2), die durch fünf Kategorien strukturiert sind [2001, S. 102-123]. Für DQ-Messungen beschreibt LOSHIN [2001, S. 171-197] neun Kategorien mit Validierungsregeln (org. *Data Quality Rule*, s. Anhang E.6.4) sowie verschiedene Beispiele (s. Anhang E.2.2) zur Messung der einzelnen DQ-Dimensionen [Loshin 2001, S. 210-228].

Ausserdem beschreibt LOSHIN [2001, S. 93-99] ein methodisches Vorgehen, um Kosten für Geschäftsprobleme zu identifizieren, deren Ursache Datendefekte sind. Das Vorgehen orientiert sich an Datennutzungsprozessen (org. *information chain*) und umfasst die folgenden Aktivitäten:

- *Datennutzungsprozesse modellieren*. Modellierung von Prozessaktivitäten und Datenflüssen (z. B. Geschäftsprozess *Rechnungsversand* mit Datenpflege durch Kundendienst, Pflege der Rechnungsposten durch Abrechnungsmanagement und Versand der Rechnung durch Postdienstleister).

- *Mitarbeiter interviewen.* Identifikation von Geschäftsproblemen in Interviews mit Mitarbeitern pro Aktivität des Datennutzungsprozesses.
- *Kunden interviewen.* Identifikation von Datendefekten in Interviews mit Kunden pro Aktivität, in der Kunden Daten nutzen (z. B. Rechnung, Marketingmassnahme).
- *Datendefekte isolieren.* Annotierung des Datennutzungsprozesses mit identifizierten Datendefekten.
- *Auswirkungen identifizieren.* Zuordnung von Geschäftsproblemen zu den einzelnen Datendefekt/Aktivitäts-Paaren.
- *Bewertung der Geschäftsprobleme.* Bewertung der einzelnen Aktivität/Geschäftsproblem-Paare durch verursachte Kosten oder unternehmensspezifische Metriken für Geschäftsprobleme.

Zur Unterstützung der Methodik beschreibt LOSHIN [2001, S. 74-93] verschiedene durch Datendefekte verursachte Geschäftsprobleme (s. Anhang E.3.2).

3.4.9 Jack E. Olson

OLSON [2003, S. 24-32] definiert Datenqualität als Erfüllung aufgabenspezifischer Anforderungen (org. „data has quality if it satisfies the requirements of its intended use“) und differenziert die DQ-Dimension *Richtigkeit* (org. *Accuracy*, „whether the data values stored for an object are the correct values“) durch sechs weitere Dimensionen (s. Anhang E.2.3), die grundsätzlich die Aspekte *falsches Datenelement* und *falsches Datenformat* unterscheiden. Dabei werden verschiedene Aspekte wie fehlende, inkonsistente oder ungenaue Datenelemente und falsche Datenformate, die andere Autoren durch Dimensionen wie *Konsistenz*, *Genauigkeit* oder *Vollständigkeit* beschreiben, berücksichtigt. Ziel dieser Differenzierung ist jedoch nicht eine neue Definition von Datenqualität, sondern die Unterscheidung messbarer und nicht messbarer Datendefekte (vgl. Abbildung B3-5).

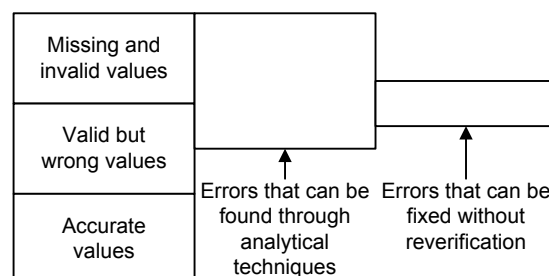


Abbildung B3-5: Messbare Datendefekte [Olson 2003, S. 35]

Zur Messung von Datenqualität (org. data profiling) beschreibt OLSON [2003, S. 121-254] fünf verschiedene Verfahren, die jeweils unterschiedliche Datendefekttypen adressieren (s. Anhang E.6.5). Alle Verfahren formulieren mögliche Datendefekte als

Regeln und identifizieren bei einer Messung die Datenobjekte, die bestimmte Regeln nicht erfüllen (vgl. Abbildung B3-6).

Neben der ausführlichen Beschreibung regelbasierter Messverfahren beschreibt OLSON [2003, S. 103-118] verschiedene Kostenarten (s. Anhang E.3.3) und Nutzenpotentiale verbesserter Datenqualität. Er betont dabei die Notwendigkeit vorbereitender Projekte zur Identifikation und Bewertung kritischer Datendefekte (org. assessment project). Diese haben zwar keinen unmittelbaren Nutzen, erhöhen aber den Nutzen folgender Projekte mit konkreten Verbesserungsmaßnahmen (org. implementation projects) durch die Auswahl effektiver Massnahmen und die Priorisierung der Defekte [Olson 2003, S. 110-112].

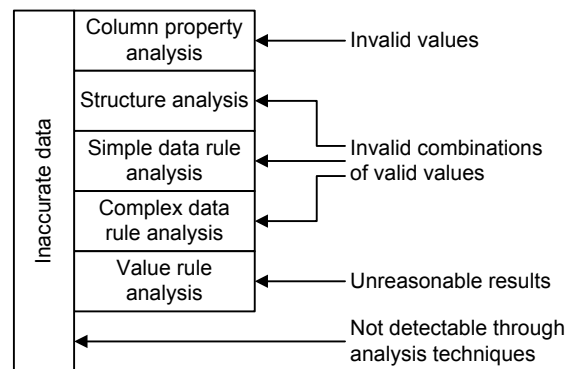


Abbildung B3-6: Regelbasierte Verfahren für DQ-Messungen [Olson 2003, S. 131]

3.4.10 Price et al.

PRICE/SHANKS [2005] definieren Datenqualität aus semiotischer Sicht und definieren die folgenden drei Kategorien für DQ-Dimensionen:

- *Syntactic Data Quality*. Diese Dimensionen bewerten die Konformität von Datenelementen und deren Spezifikation.
- *Semantic Data Quality*. Diese Dimensionen bewerten die Repräsentation der Realwelt (z. B. Geschäftsobjekte) durch Datenobjekte.
- *Pragmatic Data Quality*. Diese Dimensionen bewerten die Erfüllung von Erwartungen an Daten eines bestimmten Datennutzers bei der Bearbeitung einer bestimmten Aufgabe (subjektive Datenqualität).

Zur Messung subjektiver Datenqualität definieren PRICE ET AL. [2008] 34 Aussagen (s. Anhang E.6.3) und beschreiben deren Verwendung als Messverfahren.

3.4.11 Thomas C. Redman

REDMAN [1996, S. 245-266] definiert Datenqualität durch in drei Kategorien gegliederte DQ-Dimensionen für Qualitätsmessungen von Metadaten, Datenelementen und der Darstellung von Daten (s. Anhang E.2.3). Er fokussiert dabei die DQ-Dimension

Genauigkeit (org. *Accuracy*) [Redman 1996, S. 17-33, 2005], die er als Verhältnis der Anzahl korrekter Datenelemente zur Anzahl aller Datenelemente definiert [Redman 1996, S. 256].

3.4.12 Wang et al.

WAND/WANG [1996, S. 91f] definieren vier Datendefekttypen (org. *data deficiency*) und die vier intrinsischen DQ-Dimensionen *Vollständigkeit* (org. *complete*), *Eindeutigkeit* (org. *unambiguous*), *Bedeutung* (org. *meaningful*) und *Richtigkeit* (org. *correct*). Ein Datendefekt ist dabei eine Abweichung zwischen der unmittelbaren Realitätswahrnehmung eines Nutzers (org. „user’s direct view of a real-world system“) und der Wahrnehmung eines IS (z. B. Daten) als Realitätsrepräsentation (org. „user’s view of a real-world system as inferred from an IS representing the real-world system“) [Wand/Wang 1996, S. 88].

- *Incomplete representation*. Das IS ist nicht in der Lage, einen bestimmten Zustand des Realwelt-Systems zu repräsentieren.
- *Ambiguous representation*. Ein Zustand des IS repräsentiert mehrere Zustände des Realwelt-Systems.
- *Meaningless states*. Ein Zustand des IS repräsentiert einen unmöglichen Zustand des Realwelt-Systems.
- *Garbling*. Der als Repräsentation eines Zustands des Realwelt-Systems erzeugte Zustand eines IS ist falsch, d. h. der Zustand des Realwelt-Systems wird durch einen Zustand des IS repräsentiert, der bei korrekter Repräsentation einen anderen Zustand des Realwelt-Systems repräsentieren würde.

WANG/STRONG [1996] haben 15 DQ-Dimensionen (s. Anhang E.2.5) und eine Gliederung durch vier Kategorien in einer zweistufigen Befragung (Identifikation möglicher Dimensionen, Bewertung der gesammelten Dimensionen) von Datennutzern identifiziert. Durch die Betrachtung von Daten als Produkt und die Definition von Datenqualität als „the data’s fitness for use“ (vgl. Abschnitt 3.2.2) prägen die Autoren das dieser Arbeit zugrunde liegende Datenqualitätsverständnis. Weitere Beiträge konkretisieren diese Perspektive [Strong et al. 1997b, Wang 1998, Wang et al. 1998].

BALLOU ET AL. [1998] greifen die Betrachtung von Daten als Produkt auf und definieren ein Metamodell für Datenproduktionssysteme (org. *information manufacturing systems*), d. h. ein Prozessmodell zur Beschreibung von Datenlebenszyklen (Spezifikation, Nutzung und Pflege). Am Beispiel einer DQ-Kennzahl für die DQ-Dimension *Aktualität* (org. *Timeliness*) erläutern BALLOU ET AL. [1998, S. 467-472], wie DQ-Messwerte für die verschiedenen Metamodellelemente zu berechnen sind. SHANKARANARAYANAN ET AL. [2000, 2003] entwickeln den Ansatz weiter.

HUANG ET AL. [1999, S. 60-66] beschreiben die Notwendigkeit, DQ-Kennzahlen unternehmensspezifisch zu entwickeln, und definieren die folgenden drei Kennzahlenkategorien (keine im Vergleich zu WANG/STRONG [1996] neuen Kennzahlen).

- *Subjektive DQ-Kennzahlen.* Kennzahlen zur Messung der von Datennutzern empfundenen Datenqualität.
- *Objektive anwendungsunabhängige DQ-Kennzahlen.* Kennzahlen zur Messung von z. B. Integrität oder Vollständigkeit.
- *Anwendungsspezifische DQ-Kennzahlen.* Kennzahlen zur Prüfung, ob z. B. die Anforderungen einer datennutzende Anwendung erfüllt sind.

Zur Unterstützung der Messungen beschreiben HUANG ET AL. [1999, S. 66-89] zwei Applikationen zur Messung objektiver und anwendungsspezifischer Kennzahlen und einen Fragenkatalog zur Messung subjektiver DQ-Kennzahlen (vgl. Anhang E.6.1).

LEE ET AL. [2002, S. 143-144] definieren 65 Aussagen (vgl. Anhang E.6.2) für DQ-Messungen in Organisationen entlang der DQ-Dimensionen von WANG/STRONG [1996]. Der Fragebogen ist in fünf Organisationen mit insgesamt 261 Personen erfolgreich genutzt worden (vierseitiger Ausdruck, Fragen nicht pro Dimension organisiert sondern gemischt, pro Frage Likert-Skala mit elf Bewertungsstufen beschriftet mit „stimme überhaupt nicht zu“ (org. „not at all“), „5“ und „stimme vollständig zu“ (org. „completely“).

PIPINO ET AL. [2002, S. 213-215] definiert die folgenden drei Kategorien für die Berechnung von Messwerten von DQ-Kennzahlen und ordnet die 15 DQ-Dimensionen von WANG/STRONG [1996] exemplarisch zu.

- *Simple Ratio.* Verhältnis der Anzahl gewünschter Ergebnissen zur Anzahl aller Ergebnisse. Zugeordnete Dimensionen: *Free-of-error*, *Completeness*, *Consistency*, *Concise Representation*, *Relevancy* und *Ease of Manipulation*.
- *Min or Max Operation.* Minimaler oder maximaler Wert von Messwerten für verschiedene Attribute oder Datenobjekte. Zugeordnete Dimensionen: *Believability* (Min), *Appropriate Amount of Data* (Min), *Timeliness* (Max) und *Accessibility* (Max).
- *Weighted Average.* Gewichteter Durchschnitt von Messwerten für verschiedene Attribute oder Datenobjekte. Anwendbar auf die für die Min/Max-Operation genannten Dimensionen.

PIPINO ET AL. [2005, S. 44-49] beschreiben Berechnungsvorschriften (org. measurement scales) zur Messung der DQ-Dimensionen *Vollständigkeit* (org. *Completeness*), *Richtigkeit* (org. *Correctness*) und *Aktualität* (org. *Timeliness* und *Currency*). LEE ET AL. [2006, S. 53-66] erläutern ebenfalls verschiedene Berechnungsvorschriften zur

Messung von DQ-Dimensionen und beschreiben die Anwendung eines Fragenkatalogs [Lee et al. 2002] zur Bewertung von Datenqualität.

3.4.13 Vergleichende Bewertung vorgestellter Ansätze

Die vorgestellte Methode soll a) die Identifikation von Kausalitäten zwischen Datendefekten, Geschäftsproblemen und strategischen Unternehmenszielen und b) die Spezifikation von DQ-Kennzahlen zur Überwachung der identifizierten Datendefekte unterstützen. Die folgenden Kriterien differenzieren dieses Forschungsziel und ermöglichen die Bewertung der vorgestellten Ansätze im Vergleich untereinander und bzgl. des Forschungsziels.

- *Methodenstruktur (K1 – K5)*. Ziel der in Abschnitt 3.3.2 erläuterten Struktur einer Methode ist es, die Komplexität einer Gestaltungsaufgabe (hier: Spezifikation geschäftsorientierter DQ-Kennzahlen) durch Strukturierung der Aufgabe zu reduzieren. Auch wenn keiner der vorgestellten Ansätze eine nach diesem Methodenverständnis vollständige Methode beschreibt, sind einzelne Aktivitäten (K1), Ergebnisse (K2), Metamodell-Elemente (K3), Rollen (K4) und Techniken (K5) vergleichbar und haben die Gestaltung der Methode unterstützt.
- *Pragmatische Definition und Darstellung (K6)*. Ziel der Methodenanwendung ist es, Kennzahlen zur Überwachung geschäftskritischer Datendefekte zu spezifizieren. Dies setzt die Identifikation geschäftskritischer Datendefekte voraus, bei der fachliche Nutzer von Daten wichtige Gesprächspartner sind. Um in Interviews oder Workshops Datendefekte und mögliche Auswirkungen erläutern zu können, ist eine verständliche und problemorientierte Beschreibung des Phänomens *Datenqualität* erforderlich.
- *Kennzahlen mit Geschäftsbezug (K7)*. Ziel ist die Gestaltung von geschäftsorientierten DQ-Kennzahlen, d. h. von Kennzahlen zur Überwachung von Datendefekten, die kritische Geschäftsprobleme verursachen können. Der Gestaltungsprozess sollte daher Aktivitäten beinhalten, die Kausalitäten zwischen Datendefekten und Geschäftsproblemen identifizieren und bewerten.
- *Implementierungsfähige Kennzahlen (K8)*. Ziel ist eine detaillierte Dokumentation der DQ-Kennzahlen, so dass eine prototypische Implementierung der Messverfahren und die Interpretation der Messwerte auf einer Messskala möglich sind. Dazu ist insbesondere die konkrete Beschreibung eines Messverfahrens (z. B. Fragebogen, Validierungsregeln) erforderlich.

Tabelle B3-20 zeigt die Bewertung der vorgestellten Ansätze und Ergebnisse entlang der acht Kriterien auf einer fünfstufigen Ordinalskala. Die maximale Bewertung bedeutet, dass der jeweilige Ansatz bzw. das jeweilige Ergebnis bei der Gestaltung der Methode ohne grössere Anpassungen wiederverwendet worden ist.

Beitrag	(K1) Vorgehen	(K2) Dokumentationsmodell	(K3) Metamodell	(K4) Rollen	(K5) Techniken	(K6) Definition und Darstellung	(K7) Geschäftsbezug	(K8) Implementierbarkeit
BATINI/SCANNAPIECO [2006]	○	○	○	○	◐	◐	◐	◐
BATINI ET AL. [2007]	◐	○	○	○	○	○	◐	○
BATINI ET AL. [2009]	◐	○	○	○	○	○	◐	○
CABALLERO ET AL. [2007]	○	○	◐	○	○	○	○	○
CABALLERO ET AL. [2008]	◐	◐	○	○	◐	○	○	○
DAMA [2009]	◐	○	○	○	○	◐	○	○
ENGLISH [1999]	◐	○	○	◐	◐	◐	◐	○
KAISER ET AL. [2007]	○	○	○	○	◐	○	○	◐
HEINRICH ET AL. [2009]	◐	○	○	○	○	○	○	◐
ALUR ET AL. [2007]	◐	○	○	○	●	○	○	○
ISO/IEC [2000]	○	○	◐	○	○	○	○	○
IEEE [1998]	○	○	◐	○	○	○	○	○
IEEE [1990]	○	○	◐	○	○	○	○	○
LOSHIN [2001]	◐	○	○	○	○	◐	◐	◐
REDMAN [1996]	○	○	○	○	○	◐	○	◐
OLSON [2003]	◐	○	○	○	◐	◐	◐	◐
PRICE/SHANKS [2005]	○	○	○	○	○	◐	○	○
PRICE ET AL. [2008]	○	○	○	○	◐	○	○	◐
WAND/WANG [1996]	○	○	○	○	○	◐	○	○
WANG/STRONG [1996]	○	○	○	○	○	●	○	○
BALLOU ET AL. [1998]	○	○	○	○	◐	○	○	◐
HUANG ET AL. [1999]	○	○	○	○	◐	◐	◐	◐
LEE ET AL. [2002]	◐	○	○	○	◐	○	○	◐
PIPINO ET AL. [2002]	○	○	○	○	◐	◐	○	◐
PIPINO ET AL. [2005]	○	○	○	○	◐	◐	○	◐
LEE ET AL. [2006]	○	○	○	○	○	◐	○	◐

○: Ansatz/Ergebnis berücksichtigt jeweiligen Aspekt nicht, ◐: Ansatz/Ergebnis berücksichtigt jeweiligen Aspekt, ◑: Ansatz/Ergebnis ist als Vorlage für die Gestaltung der Methode geeignet, ◒: Ansatz/Ergebnis ist mit Modifikationen für Verwendung in der Methode geeignet, ●: Ansatz/Ergebnis wird in der Methode verwendet

Tabelle B3-2: Bewertung vorgestellter Ansätze zur Messung von Datenqualität

3.5 Methodenentwurf und Anwendung der Methode

Dieser Abschnitt erläutert die einzelnen Elemente der Methode zur Spezifikation geschäftsorientierter DQ-Kennzahlen. Da Vorgehensmodell, Dokumentationsmodell,

Rollen und Techniken eng miteinander verzahnt sind, werden diese Teilmodelle nicht einzeln, sondern in den Abschnitten 3.5.2.1 bis 3.5.2.7 in Form von Methodenfragmenten (jeweils Entwurfsergebnis, erzeugende Aktivität, beteiligte Rollen und genutzte Techniken) beschrieben [Bucher 2009, S. 49]. Techniken und Rollen werden dann beschrieben, wenn die jeweilige Rolle bzw. Technik zum ersten Mal an der Ausführung einer Aktivität beteiligt ist bzw. angewendet werden kann.

3.5.1 Metamodell

Ziel der Anwendung der vorgestellten Methode ist

- die Identifikation von Kausalketten (d. h. von zusammenhängenden Ursache-Wirkungs-Beziehungen) zwischen Datendefekten, kritischen Geschäftsproblemen und strategischen Unternehmenszielen sowie
- die Spezifikation von DQ-Kennzahlen zur Überwachung der identifizierten Datendefekte.

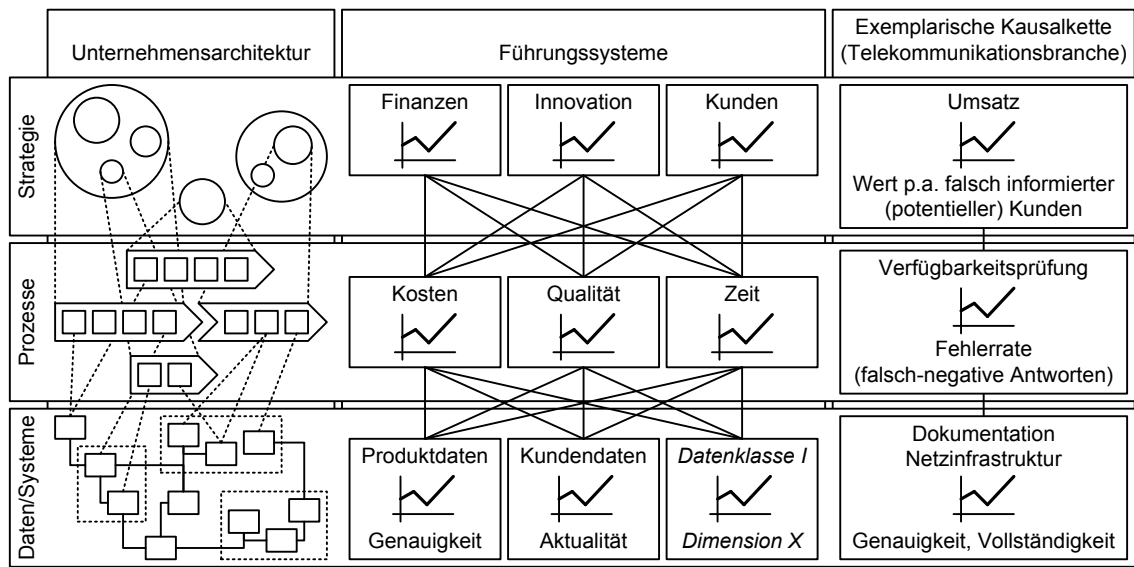


Abbildung B3-7: Kausalketten im Kontext einer Unternehmensarchitektur

Abbildung B3-7 verdeutlicht das Konzept der Kausalketten durch die Einordnung der Konzepte auf verschiedenen Ebenen einer Unternehmensarchitektur. Die vereinfachte Darstellung zeigt den Zusammenhang zwischen der von Geschäftsprozessen eines Unternehmens erbrachten Leistung und der Erreichung strategischer Geschäftsziele sowie die Nutzung von Daten bei der Leistungserbringung (vgl. Abbildung B3-7, links). Für die Identifikation konkreter Kausalketten (vgl. Abbildung B3-7, rechts) sind alle Beziehungen zwischen strategischen Zielen, Prozessaktivitäten und genutzten Daten zu berücksichtigen und die Ketten auszuwählen, die

- ein Geschäftsproblem enthalten, dass die Erreichung eines strategischen Unternehmensziels gefährdet, und
- deren Geschäftsprobleme Datendefekte als klar erkennbare Ursache haben.

Eine Kennzahl, die einen auf diese Weise identifizierten Datendefekt überwacht, hat über die Kausalkette einen klar definierten Bezug zur Wertschöpfung des Unternehmens und ist daher eine geschäftsorientierte DQ-Kennzahl.

Da strategische Erfolgsfaktoren meist durch Prozesskennzahlen oder strategische Führungssysteme (z. B. Balanced Scorecard [Kaplan/Norton 1992]) überwacht werden (vgl. Abbildung B3-7, Mitte), sind diese Überwachungsstrukturen sowohl bei der Identifikation der Kausalketten, als auch bei der Spezifikation von DQ-Kennzahlen zu berücksichtigen (vgl. Abschnitt 3.5.2.5).

Abbildung B3-8 zeigt das Metamodell der Methode mit Attributen und Beziehungen der Gestaltungsobjekte des Dokumentationsmodells. Zentrale Gestaltungsobjekte sind Datendefekte und DQ-Kennzahlen. Die übrigen Gestaltungsobjekte dienen der Dokumentation des Identifikationsprozesses und des Geschäftsbezugs der identifizierten Datendefekte. Die Bedeutung dieser Geschäftsobjekte wird im Kontext der einzelnen Methodenfragmente erläutert (vgl. Abschnitte 3.5.2.1 – 3.5.2.7).

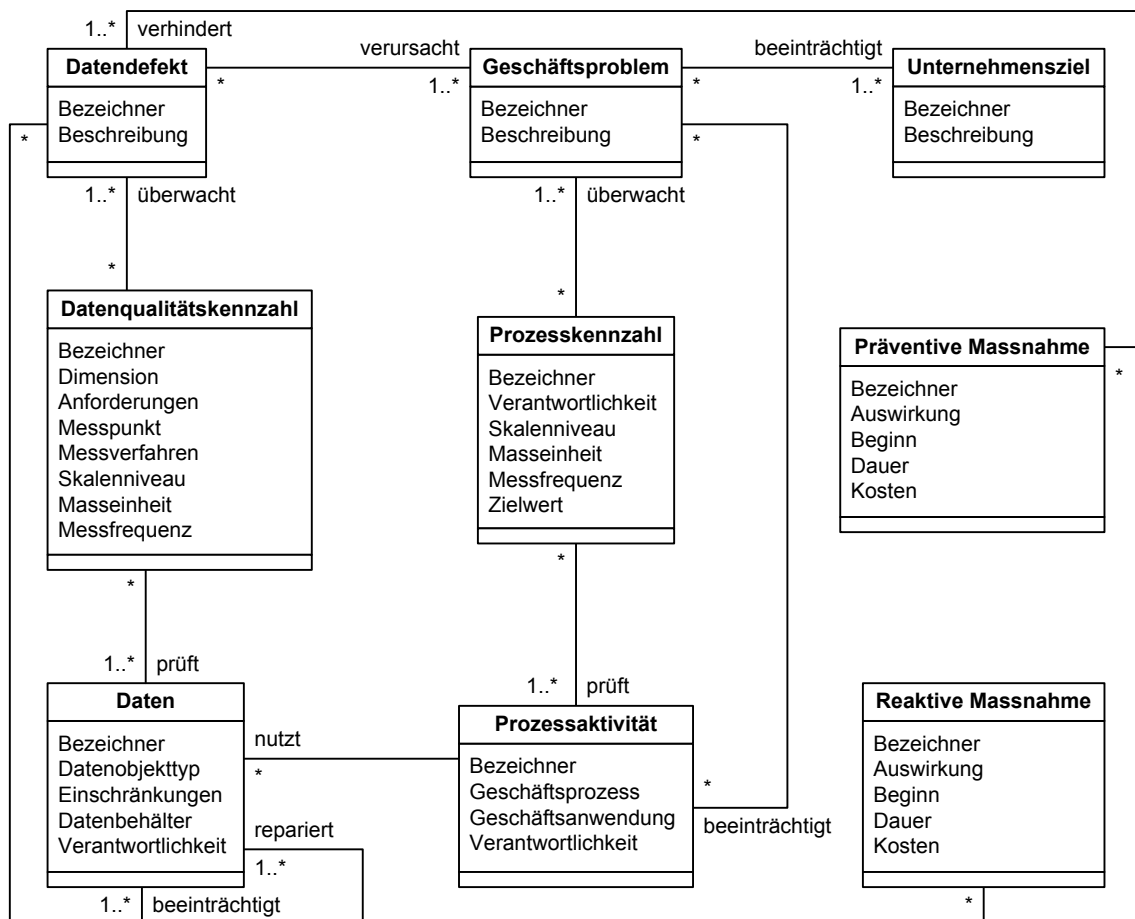


Abbildung B3-8: Metamodell der Methode

Die Definition des Konzepts *Datendefekt* basiert auf der in Abschnitt 3.2.2 erläuterten Definition von Datenqualität, nach der die Qualität von Daten dann gut ist, wenn sie die Erwartungen eines Datennutzers bei der Ausführung einer Aufgabe erfüllt. Ein Datendefekt ist nach diesem Verständnis eine Eigenschaft des Datenobjekts, des Da-

tenobjekttyps oder der Darstellung von Datenelementen, die gestellte Anforderungen nicht erfüllt. Erwartet beispielsweise die Prozessaktivität *Rechnungsversand* vollständige Adressdaten für den Postversand von Rechnungen, ist ein fehlender Strassenname ein Datendefekt, der die Prozessleistung (z. B. gemessen durch den Anteil nicht zustellbarer Rechnungen) verringert und zu einer Abweichung der Messwerte einer Prozesskennzahl von deren Sollwert führt.

Eine DQ-Kennzahl hat die Aufgabe, Datendefekte zu überwachen, d. h. die Eigenschaften der Datenobjekte, der Datenobjekttypen oder der Darstellung von Datenelementen, die nicht den Anforderungen entsprechen, frühzeitig (d. h. vor der Verursachung eines Geschäftsproblems) zu erkennen. Für die Spezifikation von DQ-Kennzahlen sind daher die an Daten gestellten Anforderungen zu identifizieren und in (möglichst maschinell ausführbare) Messverfahren zu überführen. Die vorgestellte Methode beschreibt dafür eine Technik zur Formalisierung von Anforderungen durch Validierungsregeln (vgl. Abschnitt 3.5.2.5.1).

Für die Implementierung der DQ-Kennzahlen sind insbesondere das Messverfahren und das Skalenniveau von Bedeutung. Das Messverfahren misst einen Messwert, indem es das Ergebnis einer an einem Messpunkt durchgeführten Messung auf einer Skala abbildet. Um Messwerte einer DQ-Kennzahl mit Messwerten einer Prozesskennzahl vergleichen zu können (z. B. um einen negativen Einfluss schlechter Datenqualität auf die Prozessleistung zu zeigen), muss das Skalenniveau der Prozesskennzahl bei der Spezifikation der DQ-Kennzahl berücksichtigt werden: Vergleichende Aussagen (z. B. Korrelationen) sind nur für gleiche Skalenniveaus zulässig.

3.5.2 Methodenfragmente

Die Beschreibung der vorgestellten Methode durch Methodenfragmente ist durch sieben Aktivitäten in einem Vorgehensmodell strukturiert. Da Vorgehensmodell, Dokumentationsmodell, Rollen und Techniken eng miteinander verzahnt sind, werden diese Teilmodelle nicht einzeln, sondern in den Abschnitten 3.5.2.1 bis 3.5.2.7 in Form von Methodenfragmenten (jeweils Entwurfsergebnis, erzeugende Aktivität, beteiligte Rollen und genutzte Techniken) beschrieben [Bucher 2009, S. 49]. Abbildung B3-9 zeigt eine Übersicht der Methode mit den Aktivitäten (linke Seite und den jeweils beteiligten Rollen (rechte Seite). Die mittlere Spalte von Abbildung B3-9 zeigt die Attribute des Metamodells, die die angegebenen Rollen in der jeweiligen Aktivität instanziierten und die somit das Ergebnis der einzelnen Aktivitäten bilden. Das „*-Symbol bedeutet, dass jeweils alle Attribute des angegebenen Metamodell-Entitätstyps instanziiert werden. Grundsätzlich gliedert sich die Methode in die folgenden drei Phasen.

Aktivitäten	Ergebnisse	Rollen					
		Konzern-Datensteward	Auftraggeber	Prozesseigentümer	Datennutzer	Fachlicher Datensteward	Technischer Datensteward
Phase I: Informationen sammeln							
I.1: Betrachtungsbereich definieren	Prozessaktivität.Bezeichner	x	x	x			
	Prozessaktivität.Geschäftsprozess	x	x	x			
	Prozessaktivität.Verantwortlichkeit	x		x			
	Prozesskennzahl.*	x		x			
	Geschäftsproblem.*	x		x			
	Geschäftsziel.*	x		x			
I.2: Daten und IT-Systeme identifizieren	Prozessaktivität.Geschäftsanwendung	x		x			
	Daten.*	x		x			
I.3: Kausalketten identifizieren	Geschäftsziel.Beeinträchtigung	x		x			
	Geschäftsproblem.Ursache	x			x	x	x
	Datendefekt.*	x			x	x	x
	PräventiveMassnahme.*	x			x	x	x
	ReaktiveMassnahme.*	x			x	x	x
Phase II: Analysieren und spezifizieren							
II.1: Anforderungen definieren und priorisieren	Datenqualitätskennzahl.Anforderungen	x	x			x	x
II.2: Datenqualitätskennzahlen spezifizieren	Daten.Einschränkungen	x				x	x
	Datenqualitätskennzahl.Bezeichner	x				x	x
	Datenqualitätskennzahl.Dimension	x				x	x
	Datenqualitätskennzahl.Messpunkt	x					x
	Datenqualitätskennzahl.Messverfahren	x					x
	Datenqualitätskennzahl.Skalenniveau	x				x	x
	Datenqualitätskennzahl.Masseinheit	x				x	x
	Datenqualitätskennzahl.Messfrequenz	x				x	x
Phase III: Überprüfen und dokumentieren							
III.1: Erfüllung der Anforderungen prüfen		x				x	x
III.2: Spezifikation dokumentieren		x	x				

Abbildung B3-9: Vorgehensmodell mit Beziehungen zu Meta- und Rollenmodell

- *Informationen sammeln.* Das zur Identifikation von Kausalketten benötigte Wissen ist auf zahlreiche Rollen im Unternehmen verteilt. Diese Phase umfasst daher Aktivitäten mit hohem Anteil an Kommunikation, in denen Informationen von verschiedenen Rollen gesammelt und zu einem Betrachtungsbereich für die Anwendung der Methode und zu Kausalketten verdichtet werden.
- *Analyse und Spezifikation.* Ziel ist die Spezifikation aller zur Implementierung mindestens einer DQ-Kennzahl benötigten Informationen. Zentrale Aufgabe ist die Formalisierung identifizierter Datendefekte durch Validierungsregeln (oder ein anderes Messverfahren) und die Verknüpfung der Regeln (z. B. durch Berechnungsvorschriften) zu Kennzahlen.

- *Überprüfen und dokumentieren.* Ziel ist die Dokumentation der Spezifikationen der DQ-Kennzahlen und des Geschäftsbezugs (d. h. der Kausalketten), so dass die Kennzahlen implementiert und ihre Messwerte mit der Leistung von Geschäftsprozessen in Beziehung gesetzt werden können.

Die Abschnitte 3.5.2.1 bis 3.5.2.7 enthalten neben der Beschreibung der Methode Beispiele, die die Instanziierung bzw. Nutzung einzelner Methodenelemente in drei Unternehmen schildern. Tabelle B3-21 skizziert die anonymisierten Unternehmen und den jeweiligen Projektkontext.

	Konsumgüter AG	Strom AG	Telefon AG
Branche	Konsumgüterproduktion	Stromversorgung	Telekommunikation
Umsatz (2009)	6 Mrd. EUR	750 Mio. CHF	60 Mrd. EUR
Mitarbeiter (2009)	20'000	1'400	260'000
Projektkontext			
Zeitraum	01/2010 – 06/2010	11/2009 – 06/2010	06/2009 – 02/2010
Projektteam	5 Personen 17 Interviewpartner	4 Personen 9 Workshop-Teilnehmer	3 Personen 2 Interviewpartner
Geschäftsprozesse	Distribution, Marketing, Qualitätsmanagement, Produktion	Abrechnung, Mahnung, Rechnungsversand, Marketing	Verfügbarkeitsprüfung Breitband-Internet
Datenobjekttyp	Produktdaten	Geschäftspartner-, Vertrags-, Vertragskontodaten	Netzinfrastuktur-, Kundendaten

Tabelle B3-3: Anwendungsunternehmen (anonymisiert) und Projektkontext

3.5.2.1 Betrachtungsbereich definieren

Ziel von Aktivität I.1 ist die Definition möglicher Betrachtungsbereiche (vgl. Tabelle B3-22).

Aktivität I.1: Betrachtungsbereich definieren Ziel ist die Identifikation von Prozessaktivitäten oder Geschäftsvorfällen, für die Geschäftsprobleme bekannt sind oder vermutet werden, deren Ursache vermutlich Datendefekte sind. Für jede Prozessaktivität sollte mindestens eine Prozesskennzahl identifiziert werden, für die Messwerte verfügbar sind. Ausserdem ist der (monetäre) Effekt der Geschäftsprobleme zu dokumentieren.		
Ergebnisdokument Liste mit Betrachtungsbereichen	Rollen <ul style="list-style-type: none"> • Auftraggeber • Konzern-Datensteward • Prozesseigner 	Techniken <ul style="list-style-type: none"> • Interview/Workshop • Datenqualitätssimulation
Checkliste <ul style="list-style-type: none"> <input type="checkbox"/> Prozessaktivitäten und Geschäftsvorfälle mit bekannten oder vermuteten Geschäftsproblemen (potenziell verursacht durch Datendefekte) sind identifiziert. <input type="checkbox"/> Prozesskennzahlen zur Leistungsüberwachung identifizierter Prozessaktivitäten sind identifiziert. <input type="checkbox"/> Geschäftsprobleme und Datendefekte als potentielle Ursache sind beschrieben. <input type="checkbox"/> Kosten, die durch identifizierte Geschäftsprobleme entstehen, sind beschrieben. 		

Tabelle B3-4: Methodenfragment zu Aktivität I.1

Ein Betrachtungsbereich fokussiert einen Ausschnitt der Unternehmensarchitektur (z. B. Prozessaktivitäten, Geschäftsobjekte, Datenobjekttypen) eines Unternehmens, in dem Geschäftsprobleme bekannt sind oder vermutet werden. Ziel der weiteren Methodenanwendung ist es, in diesen Betrachtungsbereichen Kausalketten (insbesondere Geschäftsprobleme und verursachende Datendefekte) zu identifizieren und zur Überwachung der Datendefekte DQ-Kennzahlen zu spezifizieren.

Die Telefon AG hat als Betrachtungsbereich die Verfügbarkeitsprüfung für IP-TV definiert (Prozessaktivität). Über eine Website können potentielle Kunden prüfen, ob an einer bestimmten Adresse IP-TV verfügbar ist (positive Antwort) oder nicht (negative Antwort). Eine falsch-negative Antwort, also eine negative Antwort trotz tatsächlicher Verfügbarkeit von IP-TV an der angefragten Adresse, reduziert deutlich die Wahrscheinlichkeit, die anfragende Person als Neukunden zu gewinnen (Wert eines Neukunden ca. 500 EUR p.a.), und erschwert dadurch die Erreichung strategischer Unternehmensziele (Gewinnung von Neukunden, Umsatzsteigerung). Der Anteil falsch-negativer Antworten ist daher eine Prozesskennzahl zur Überwachung der Leistung der Verfügbarkeitsprüfung. Es wird ein deutlicher Zusammenhang zwischen der Richtigkeit einer Antwort der Verfügbarkeitsprüfung und der Qualität (Genauigkeit, Vollständigkeit) der Netzinfrastrukturdaten vermutet.

Für jede fokussierte Prozessaktivität (d. h. als Bestandteil jedes Betrachtungsbereichs) sollte mindestens eine Prozesskennzahl identifiziert werden, für die Messwerte verfügbar sind. Ausserdem ist das Geschäftsproblem (im Beispiel eine falsch-negative Antwort der Verfügbarkeitsprüfung) sowie die Auswirkung des Problems auf die Prozesskennzahl und die Erreichung des fokussierten Unternehmensziels zu beschreiben.

Die Konsumgüter AG hat bei der Suche nach kritischen Datendefekten die gesamte Supply Chain betrachtet und den Betrachtungsbereich durch die Fokussierung von Produktdaten sowie inner- und überbetrieblicher Schnittstellen eingeschränkt. Gründe für diese Definition des Betrachtungsbereichs sind einerseits die grundsätzlich gute Qualität der Stammdaten (wenig Kritik, wenig bekannte Probleme) und andererseits vereinzelt aber konkrete Kritik an der Qualität logistischer Produktdaten (insbesondere Gewichte). Strategische Unternehmensziele (Compliance, Kostenreduktion, hohes Service Level) sind nicht durch konkrete Prozesskennzahlen differenziert worden.

Neben der Fokussierung konkreter Prozessaktivitäten als Betrachtungsbereich ist auch eine Fokussierung von Datenobjekttypen (im Beispiel Produktdaten) möglich. Diese Fokussierung ist dann sinnvoll, wenn konkrete Datendefekte bekannt sind oder die

DQM-Organisation (z. B. Abteilung der IT-Organisation, nach Datenobjekttypen gegliedertes Shared Service Center) die Fokussierung bestimmter Datenobjekte nahelegt.

Die Strom AG hat den Betrachtungsbereich auf die Geschäftsobjekte Kunde, Vertrag und Vertragskonto sowie verschiedene Geschäftsvorfälle aus Abrechnungs-, Kundendienst- und Marketingprozessen eingeschränkt. Grund für diese Auswahl ist das Ziel, auf Basis der später spezifizierten DQ-Kennzahlen DQ-Messungen mit anderen Energieversorgern vergleichen zu können (Benchmarking). Daher ist der Betrachtungsbereich auf Geschäftsobjekte und Geschäftsvorfälle eingeschränkt worden, die vermutlich bei vielen Energieversorgern ähnlich strukturiert und somit vergleichbar sind.

Grundsätzlich ist schon bei der Definition des Betrachtungsbereichs das übergeordnete Ziel von DQ-Messungen zu berücksichtigen. Zwar kann auch die Spezifikation von DQ-Kennzahlen als DQM-Werkzeug alleiniges Ziel der Methodenanwendung sein. Meist sind jedoch konkrete Anforderungen wie die Verbesserung der Leistung einer bestimmten Prozessaktivität (vgl. Telefon AG), die Konkretisierung bekannter Datendefekte und Explikation des Geschäftsbezugs (vgl. Konsumgüter AG) oder unternehmensübergreifendes Benchmarking (vgl. Strom AG) Treiber für die Spezifikation von DQ-Kennzahlen.

3.5.2.1.1 Rollen

Die Rollen der Methode basieren auf einem Referenzmodell für Data Governance [2009, S. 9-13, Weber 2009, S. 106-114]. Die zentrale Rolle ist der Konzern-Datensteward (vgl. Tabelle B3-22). Er ist verantwortlich für die Konfiguration (z. B. Auswahl von Techniken) und Anwendung der Methode, koordiniert die Zusammenarbeit der beteiligten Personen und Gremien und dokumentiert die Ergebnisse. Meist nimmt der Leiter der DQM-Organisation diese Rolle wahr, in grossen DQM-Organisationen kann dieser aber auch Auftraggeber sein und die Aufgaben des Konzern-Datenstewards an eine andere Person der DQM-Organisation delegieren. Die weitere Beschreibung der Methode geht von einer DQM-Organisation unter der Leitung des Konzern-Datenstewards (eine Person) und einem Auftraggeber (eine Person, Vorgesetzter des Konzern-Datenstewards) ausserhalb der DQM-Organisation aus.

Der Auftraggeber definiert Ziele der Methodenanwendung, gewährleistet die Verfügbarkeit der benötigten Ressourcen und gibt das Fachkonzept (Ergebnisdokument von Aktivität III.2, vgl. Abschnitt 3.5.2.7) frei. Zentrale Aufgabe von Konzern-Datensteward und Auftraggeber in Aktivität I.1 ist die Identifikation eines Betrachtungsbereichs, der einerseits einen wichtigen Teil der Wertschöpfung des Unternehmens fokussiert und für den andererseits klare Kausalketten, also deutlich erkennbare Beziehungen zwischen Datendefekten und geschäftsrelevanten Problemen, existieren.

Der Leiter der DQM-Organisation der Konsumgüter AG hat die Durchführung der Methode mit Hilfe externer Unterstützung geleitet. Als Auftraggeber hat der Leiter des Supply Chain Managements, zu dem organisatorisch auch das DQM gehört, die Massnahme gefördert (Projektbudget, Personalressourcen, Unterstützung der Anfragen für unternehmensweite Interviews).

Da neben spezifischen Kenntnissen zu Geschäftsprozessen, Geschäftsmodell und Aufbauorganisation des Unternehmens auch generelle Erfahrung bei der Suche nach Kausalketten ein Erfolgsfaktor der Methodenanwendung ist, kann externe Unterstützung des Konzern-Datenstewards sinnvoll sein.

Prozesseigentümer [DIN/EN/ISO 2000, S. 22] dienen als Gesprächspartner bei der Identifikation geeigneter Betrachtungsbereiche. Typische Aufgaben eines Prozesseigentümers sind die Planung und Kontrolle von Prozesszielen, die Pflege von Prozesskennzahlen und die Planung und Initiierung von Verbesserungsmassnahmen. Seine Aufgabe bei der Methodenanwendung ist es einzuschätzen, welche Prozessaktivitäten besonders durch Datendefekte gefährdet und welche Prozesskennzahlen Auswirkungen möglicher Datendefekte deutlich zeigen könnten.

3.5.2.1.2 Technik Interview/Workshop

Ziel eines Interviews ist es, in einem Gespräch mit einer oder mehreren Personen Informationen zu einem bestimmten Sachverhalt oder eine Bewertung eines bestimmten Sachverhalts zu sammeln. Ein Workshop dient im Vergleich zu einem Interview weniger dem Sammeln von Informationen, sondern mehr der Erarbeitung oder Diskussion von Ergebnissen (z. B. Definition von Anforderungen). Es ist sinnvoll, einen Workshop mit mehreren Personen durchzuführen, da Diskussionen zwischen den Teilnehmern neue Perspektiven schaffen und die Identifikation von ansonsten verborgenen Zusammenhängen (z. B. abteilungsübergreifende Kausalitäten zwischen Datendefekten und Geschäftsproblemen) erleichtern. Die Methode stellt Interviewleitfäden, Dokumentationsvorlagen und Beispielergebnisse (z. B. exemplarische Kausalketten) zur Unterstützung der vorgesehenen Interviews und Workshops bereit.

Da in keiner der betroffenen Aktivitäten exklusiv Interviews oder Workshops zu empfehlen sind, fasst die Methode die beiden Optionen in einer Technik *Interview/Workshop* zusammen und beschreibt pro Aktivität ihre Anwendung. Um die Lesbarkeit zu verbessern, wird immer entweder ein Interview oder ein Workshop beschrieben. In allen Fällen leitet der Konzern-Datensteward den Workshop bzw. führt das Interview und versucht, durch gezielte Fragen und Diskussionen alle benötigten Informationen und Bewertungen zu sammeln und zu dokumentieren. Weitere Techniken der Methode (z. B. die Datenqualitätssimulation, vgl. Abschnitt 3.5.2.1.3) können ein Interview oder einen Workshop unterstützen.

Für die Identifikation möglicher Betrachtungsbereiche ist zuerst ein Interview mit dem Auftraggeber zu führen und dabei

- eine Liste mit aktuell laufenden Projekten,
- eine Liste mit bereits abgeschlossenen oder gescheiterten Projekten und
- eine Liste mit Geschäftsproblemen

zu erarbeiten, für die Datendefekte oder DQM-Massnahmen bereits diskutiert wurden oder aktuell diskutiert werden. Ziel dieses Interviews ist es, Kenntnis von bereits vorhandenen Analysen, Spezifikationen oder anderen Projektergebnissen zu erhalten und Prozesseigentümer für weitere Interviews zu identifizieren. Zur Unterstützung des Interviews stellt die Methode Listen typischer Geschäftsprobleme und Nutzenpotentiale verbesserter Datenqualität (s. Anhang E.3), exemplarische Kausalketten (s. Anhang E.4.1) und eine Dokumentationsvorlage (s. Anhang E.7.1) bereit. Falls konkrete Datendefekte oder Geschäftsprobleme der Grund zur Anwendung der Methode sind, kann dieses Interview durch eine kurze Abstimmung der Ziele der Methodenanwendung ersetzt werden.

In Interviews mit einzelnen Prozesseigentümern sind die zuvor nur skizzierten Betrachtungsbereiche zu konkretisieren und pro Betrachtungsbereich

- eine Prozessaktivität,
- mindestens ein Geschäftsproblem,
- mindestens eine Prozesskennzahl, auf deren Messwerte sich das Geschäftsproblem auswirkt,
- die Auswirkung der Geschäftsprobleme auf die Prozessleistung und
- Datendefekte als mögliche Ursache jedes Geschäftsproblems

zu beschreiben. Die Materialien aus den Anhängen E.3, E.4.1 und E.7.1 können dabei unterstützen.

3.5.2.1.3 Technik Datenqualitätssimulation

Diese Technik ermöglicht die Simulation von Datenqualitätsmesswerten auf Basis eines modellierten DQM-Szenarios. Ein solches Szenario umfasst verschiedene Kausalketten (d. h. Datendefekte, DQ-Kennzahlen, Geschäftsprobleme) und DQM-Massnahmen und definiert Annahmen zur Auftrittswahrscheinlichkeit von Datendefekten, zur Verursachung von Geschäftsproblemen und zur Wirkung von DQM-Massnahmen [Otto et al. 2010].

Abbildung B3-10 zeigt ein Werkzeug zur Modellierung und Simulation von DQM-Szenarien. Das Werkzeug ermöglicht die Definition von Datendefekten, wobei für jeden Datendefekt eine Auftrittswahrscheinlichkeit angegeben wird (vgl. Schieberegler in Abbildung B3-10). Ein Szenario umfasst ausserdem Geschäftsprobleme und die Zuordnung der definierten Datendefekte als Ursache dieser Geschäftsprobleme. Ein Simulationsschritt simuliert das Auftreten der Datendefekte und die Anzahl der verursachten Geschäftsprobleme. Reaktive DQM-Massnahmen können Datendefekte reparieren und so die Anzahl der verursachten Geschäftsprobleme im nächsten Simulati-

onsschritt reduzieren. Die Simulation eines konkreten Szenarios für die Telefon AG wird in Abschnitt 3.5.2.7 beschrieben.

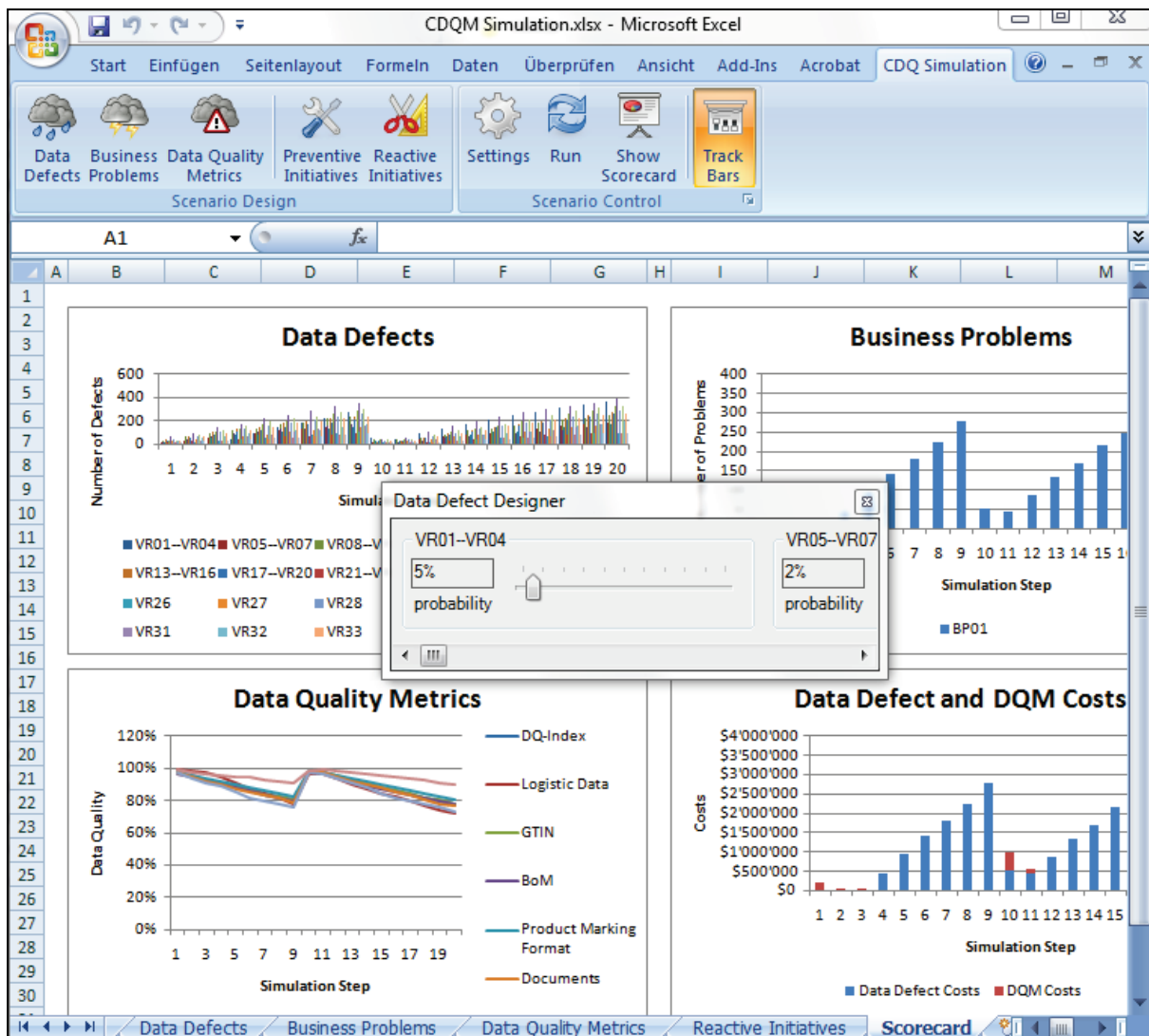


Abbildung B3-10: Werkzeug zur Simulation von Datenqualitätsmanagement-Szenarien

Bei der Definition möglicher Betrachtungsbereiche kann das in Abbildung B3-10 dargestellte Werkzeug zur Unterstützung der Interviews mit Auftraggeber und Prozesseigentümern genutzt werden, um diskutierte Zusammenhänge zu veranschaulichen und exemplarische Kausalketten zu erklären. Die Methode stellt für die in Anhang E.4.1 beschriebenen Kausalketten modellierte Szenarien bereit, die mit dem Werkzeug (s. Abbildung B3-10) angepasst und simuliert werden können. Für eine detailliertere Beschreibung der Technik und dem verwendeten Simulationsverfahren sei auf OTTO ET AL. [2010] verwiesen.

3.5.2.2 Daten und IT-Systeme identifizieren

Ziel von Aktivität I.2 ist die Identifikation der in den einzelnen Betrachtungsbereichen genutzten IT-Systeme und die möglichst exakte Beschreibung der betrachteten Daten (vgl. Tabelle B3-23). In Interviews (vgl. Technik *Interview/Workshop*) mit den jeweils

zuständigen Prozesseigentümern sind für jeden zuvor identifizierten Betrachtungsbereich die in der jeweiligen Prozessaktivität genutzten Daten und die IT-Systeme, die die Daten bereitstellen (Datenbehälter) bzw. nutzen (Geschäftsanwendungen), zu identifizieren. Für die Daten ist nicht nur der Datenobjekttyp (z. B. Kundendaten), sondern eine möglichst detaillierte Einschränkung der vermutlich defekten Datenobjekte (z. B. „alle männlichen Kunden aus St. Gallen“, „alle aktiven in der Schweiz angebotenen Produkte“) zu dokumentieren. Zur Unterstützung der Interviews stellt die Methode eine Dokumentationsvorlage (s. Anhang E.7.2) bereit.

Aktivität I.2: Daten und IT-Systeme identifizieren Ziel ist es, für jeden zuvor identifizierten Betrachtungsbereich die in den Prozessaktivitäten genutzten Daten und die IT-Systeme, die die Daten speichern (Datenbehälter) bzw. nutzen (Geschäftsanwendungen), zu identifizieren und durch Einschränkungen der bereits dokumentierten Datenobjekttypen zu konkretisieren. Ausserdem ist die Art möglicher Datendefekte durch die Auswahl betrachteter DQ-Dimensionen einzuschränken.		
Ergebnisdokument Liste mit Betrachtungsbereichen, Daten, IT-Systemen und DQ-Dimensionen	Rollen <ul style="list-style-type: none"> • Konzern-Datensteward • Prozesseigentümer 	Techniken <ul style="list-style-type: none"> • Interview/Workshop
Checkliste <ul style="list-style-type: none"> <input type="checkbox"/> Pro Betrachtungsbereich sind die betrachteten Daten möglichst exakt (d. h. Datenobjekttyp und Einschränkungen) beschrieben. <input type="checkbox"/> Pro Betrachtungsbereich sind genutzte Datenbehälter und Geschäftsanwendungen identifiziert. <input type="checkbox"/> Pro Betrachtungsbereich sind fokussierte DQ-Dimensionen ausgewählt. 		

Tabelle B3-5: Methodenfragment zu Aktivität I.2

Die Konsumgüter AG hat ca. 100 Datenattribute (gruppiert durch 20 Daten-Cluster) aus dem ca. 800 Attribute umfassenden Produktdatenmodell des Unternehmens ausgewählt. Die Strategie der DQM-Organisation gibt als Leitmotiv „timely provision of accurate, complete and consistent data“ vor. Um die Diskussion in geplanten Workshops zu fokussieren, sind daher die vier DQ-Dimensionen Aktualität, Genauigkeit, Konsistenz und Vollständigkeit ausgewählt worden.

Neben der Einschränkung der betrachteten Daten sollte der Konzern-Datensteward einige DQ-Dimensionen für die Kausalkettenidentifikation in Aktivität I.3 auswählen, um eine fokussierte Diskussion zu gewährleisten. Schon die Auswahl der in Abschnitt 3.4 vorgestellten Ansätze zeigt jedoch, dass keine allgemeingültige Definition von Datenqualität existiert, sondern verschiedene Autoren unterschiedliche, sich teilweise überschneidende Listen mit DQ-Dimensionen, vorschlagen. Zwar ist die von WANG/STRONG [1996] vorgeschlagene Liste mit 15 Dimensionen (vgl. Anhang E.2.5) der wohl bekannteste Ansatz. Andere Arbeiten [Olson 2003, Price/Shanks 2005] sind jedoch ebenfalls insbesondere unter Berücksichtigung von DQ-Messungen, also der Operationalisierung von DQ-Dimensionen durch objektiv messbare Eigenschaften von Datenobjekten, als Ziel der Methodenanwendung zu beachten.

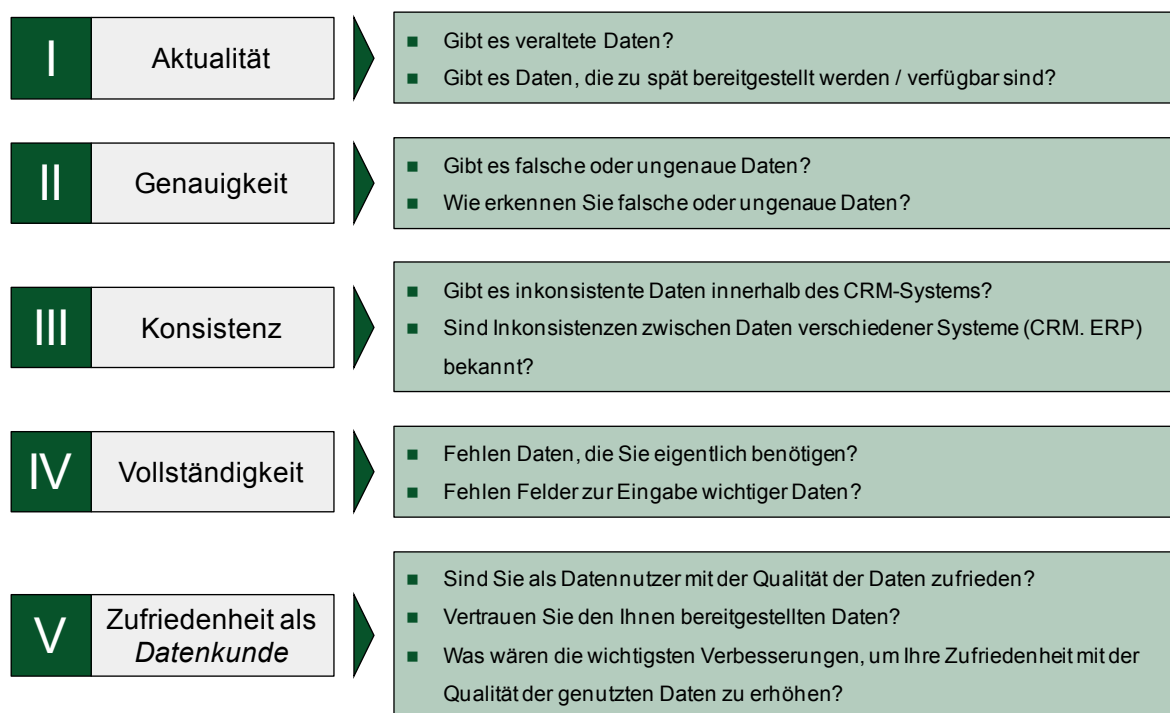


Abbildung B3-11: Leitfragen für ausgewählte DQ-Dimensionen

Die Methode gibt keine Auswahl von DQ-Dimensionen vor, sondern stellt in E.1 verschiedene Listen zur Orientierung bereit. Der Konzern-Datensteward sollte aus diesen Vorschlägen ca. vier bis acht Dimensionen auswählen und pro Dimension Leitfragen (vgl. Abbildung B3-11) zur Erläuterung formulieren. Die Anhänge E.6.1, E.6.2 und E.6.3 sowie eine Dokumentationsvorlage aus Anhang E.7.2 können dabei unterstützen. Falls die DQM-Organisation eine DQM-Strategie formuliert hat, ist diese unbedingt zu beachten, da dort berücksichtigte DQ-Dimensionen (vgl. Konsumgüter AG) als Erfolgsfaktoren für die Leistungserstellung des Unternehmens bereits mit der Unternehmensstrategie abgestimmt sind.

3.5.2.3 Kausalketten identifizieren

Ziel dieser Aktivität ist es, Kausalketten mit geschäftskritischen Datendefekten zu identifizieren, zu deren Überwachung der Konzern-Datensteward in Aktivität II.1 DQ-Kennzahlen spezifiziert (vgl. Tabelle B3-24). Der Identifikationsprozess sollte *Top-Down* verlaufen, d. h. der Konzern-Datensteward sollte versuchen, ausgehend von bekannten Geschäftsproblemen verursachende Datendefekte zu identifizieren.

Die Technik *Interview/Workshop* unterstützt den Identifikationsprozess durch Interviews mit verschiedenen Experten des Unternehmens (vgl. Abschnitt 3.5.2.3.1). Die Technik *Datenqualitätssimulation* kann die Diskussion in Interviews unterstützen und diskutierte Zusammenhänge anhand simulierter DQM-Szenarien veranschaulichen. Falls die Identifikation von Datendefekten ausgehend von Geschäftsproblemen (Top-Down-Vorgehen) nicht zum gewünschten Ergebnis führt oder der Konzern-Datensteward potentielle Datendefekte als Vorbereitung von Interviews identifizieren

möchte (Bottom-Up-Vorgehen), ermöglicht die Technik *Datenqualitätsanalyse* (vgl. Abschnitt 3.5.2.3.2) die Identifikation statistisch auffälliger Datenelemente.

Aktivität I.3: Kausalketten identifizieren Ziel ist die Identifikation von Kausalketten, d. h. von Ursache-Wirkungs-Beziehungen zwischen Datendefekten, Geschäftsproblemen und strategischen Unternehmenszielen. Dazu sind für jede Prozessaktivität der Betrachtungsbereiche Geschäftsprobleme zu identifizieren sowie für jedes Geschäftsproblem dessen Auswirkung (z. B. Kosten) und verursachende Datendefekte.		
Ergebnisdokument Liste mit Kausalketten	Rollen <ul style="list-style-type: none"> • Fachliche Datenstewards • Konzern-Datensteward • Datennutzer • Technische Datenstewards • Prozesseigentümer 	Techniken <ul style="list-style-type: none"> • Datenqualitätsanalyse • Datenqualitätssimulation • Interview/Workshop
Checkliste <input type="checkbox"/> Kausalketten sind komplett (d. h. Datendefekte, Geschäftsprobleme und Auswirkungen) beschrieben. <input type="checkbox"/> Kausalitäten (Geschäftsproblem / Auswirkung auf strategisches Unternehmensziel, Datendefekt / Geschäftsproblem) sind exakt (d. h. möglichst quantifiziert) beschrieben.		

Tabelle B3-6: Methodenfragment zu Aktivität I.3

3.5.2.3.1 Rollen

Die Identifikation von Kausalketten ist Aufgabe des Konzern-Datenstewards. Da dieser jedoch meist nicht über das benötigte Wissen zur gesamten Wertschöpfung des Unternehmens mit Prozess- und Datenarchitektur sowie zugehörigen Führungssystemen (vgl. Abschnitt 3.5.1) verfügt, muss er Informationen von verschiedenen Experten sammeln und konsolidieren.

Fachliche Datenstewards sind Teil der DQM-Organisation eines Unternehmens und berichten an den Konzern-Datensteward (vgl. Abschnitt 3.5.2.1.1). Sie sind zuständig für die Zusammenarbeit mit Fachbereichen und pflegen beispielsweise fachliche Metadaten oder fachliche Datenqualitätsanforderungen an Anwendungssysteme. Auch wenn andere Zuständigkeiten möglich sind (z. B. pro Abteilung, pro Prozess), geht die weitere Beschreibung der Methode von der Zuständigkeit eines fachlichen Datenstewards für jeweils einen Datenobjekttyp (z.B. Materialstamm, Kundenstamm) aus.

Für die IT-Abteilung ist ein technischer Datensteward das, was ein fachlicher Datensteward für den Fachbereich ist. Auch hier unterstellt die weitere Beschreibung der Methode einen technischen Datensteward pro Datenobjekttyp. Ein technischer Datensteward hat detaillierte Fachkenntnisse zu den Datenbehältern und Anwendungssystemen, die die jeweiligen Daten speichern bzw. die auf ihnen operieren.

Datennutzer sind Mitarbeiter des Unternehmens, die Daten in ihrer täglichen Arbeit (d. h. der Ausführung von in Prozessaktivitäten beschriebenen Arbeitsabläufen) erzeugen und nutzen (wobei die Nutzung bei der Anwendung der Methode im Vordergrund steht). Da sie bei ihrer Arbeit unmittelbar mit Datendefekten konfrontiert sind (z. B.

fehlende, ungenaue, veraltete Datenelemente) können sie die unmittelbaren Auswirkungen defekter Daten beschreiben und sind daher wichtige Gesprächspartner bei der Identifikation von Kausalketten.

Neben dem Konzern-Datensteward und externer Unterstützung hat das Projektteam der Konsumgüter AG, das die Methode angewendet hat, einen für Produktdaten zuständigen fachlichen Datensteward sowie einen technischen Datensteward umfasst. Der technische Datensteward hat mit detailliertem Fachwissen zur Funktionalität des zentralen Product Lifecycle Management (PLM)-System der Konsumgüter AG unterstützt. In sieben Workshops haben insgesamt 16 Datennutzer Kausalketten identifiziert. Der Auftraggeber hat im Vorfeld der Interviews die Notwendigkeit der Massnahme kommuniziert und dadurch die zeitliche Verfügbarkeit und aktive Mitarbeit der Interview-Partner gewährleistet.

3.5.2.3.2 Technik Datenqualitätsanalyse

Die Technik *Datenqualitätsanalyse* nutzt statistische Verfahren zur Identifikation ungewöhnlicher (d. h. signifikant vom Mittelwert abweichender oder seltener) Datenelemente. Verschiedene Softwarehersteller bieten Anwendungen zur Unterstützung dieser Technik [Alur et al. 2007]. Sie sollte, soweit der Konzern-Datensteward die Anwendung für sinnvoll hält, zur Vorbereitung eines Workshops (vgl. Abschnitt 3.5.2.3.3) genutzt werden, um Anhaltspunkte für mögliche Kausalketten zu finden. Für die identifizierten Abweichungen ist jeweils zu prüfen, ob es sich tatsächlich um einen Datendefekt handelt und ob dieser Ursache eines bekannten Geschäftsproblems ist (Bottom-Up-Vorgehen).

3.5.2.3.3 Anwendung der Technik Interview/Workshop in Aktivität I.3

Das Wissen zu den einzelnen Aspekten ist jedoch meist auf mehrere Personen verteilt (technische Datenstewards haben detailliertes Wissen zu Datendefekten, fachliche Datenstewards und Datennutzer zu Geschäftsproblemen). Und ausserdem erstrecken sich die Prozesse oft über verschiedene Abteilungen (z. B. Adressdateneingabe durch ein Call Center, Adressdatenbereinigung durch die DQM-Organisation und Adressdatennutzung beim Rechnungsversand und im Marketing). Daher sollte der Konzern-Datensteward versuchen, in einem Workshop möglichst verschiedene Rollen und Mitarbeiter aus unterschiedlichen Abteilungen zusammenzuführen und eine Diskussion anzuregen.

Geschäftspartner	Anrede	Name	Strasse / Nr.	Zusatz	PLZ	Ort
	Postfach Nr.	Postfach PLZ	Postfach Ort	Telefon	Fax	Natel
	Email	Kundensegment	Bankschlüssel	Bankkonto	Kontrollschlüssel	
	IBAN					

	Problem	Häufigkeit?	Wie problematisch?	Mögliche Probleme
Veraltete Daten?	Beschreibung ...	<input type="checkbox"/> (0 - 5)	<input type="checkbox"/> (0 - 5)	<ul style="list-style-type: none"> ■ Frauen werden als „Herr“ angeschrieben und umgekehrt. ■ Keine Anrede vorhanden. ■ ...
Falsche Daten?	Beschreibung ...	<input type="checkbox"/> (0 - 5)	<input type="checkbox"/> (0 - 5)	
Inkonsistente Daten?	Beschreibung ...	<input type="checkbox"/> (0 - 5)	<input type="checkbox"/> (0 - 5)	
Fehlende Daten?	Beschreibung ...	<input type="checkbox"/> (0 - 5)	<input type="checkbox"/> (0 - 5)	

Weitere Fragen ?

- Sind Sie als Datennutzer mit der Qualität der Daten zufrieden?
- Vertrauen Sie den Ihnen bereitgestellten Daten?
- Was wären die wichtigsten Verbesserungen, um Ihre Zufriedenheit mit der Qualität der genutzten Daten zu erhöhen?

Abbildung B3-12: Workshop-Leitfaden zur Identifikation von Kausalketten

Die Strom AG hat einen Workshop mit insgesamt neun Mitarbeitern aus den Bereichen Abrechnung Grosskunden, Abrechnung Tarifikunden, DQM, Kundendienst, Marketing, Rechnungsstellung und Rechnungsversand durchgeführt. Für zehn Geschäftsvorfälle (z. B. Briefkampagne, Mahnung, Rechnungserstellung) haben die Teilnehmer Geschäftsprobleme diskutiert und Defekte von Geschäftspartner-, Vertrags- und Vertragskontodaten identifiziert und bewertet.

Zur Unterstützung eines Workshops sollte der Konzern-Datensteward in einer Präsentation den Betrachtungsbereich (z. B. betrachtete Attribute, Datenobjekttypen und Geschäftsvorfälle) so aufbereiten, dass er die Diskussion fokussiert. Die Konsumgüter AG hat beispielsweise einen Leitfaden mit fachlich gruppierten Attributen erstellt und diese dann pro fokussiertem Geschäftsvorfall diskutiert (s. Abbildung B3-12).

Die Konsumgüter AG hat in den sieben durchgeführten Workshops verschiedene Kausalketten identifiziert und ausserdem die Häufigkeit und Kritikalität (unmittelbar durch z. B. Mehraufwand oder fehlerhafte Produktion verursachte Kosten, Risiko von Gesetzesverstössen und Beeinträchtigung des Service Level) bewertet. Abbildung B3-13 zeigt einen Ausschnitt der verwendeten Dokumentationsvorlage. Durch die Bewertung hat die Konsumgüter AG sieben Problembereiche identifizieren können, die das in Aktivität II.2 spezifizierte Kennzahlensystem strukturiert haben.

	Problem	Frequency rating (0-5)	Impact rating (0-5)	Costs	Service level	Legal compliance
3	Material Master, Overview	0.50	0.33	0.23	0.06	
4	Material Master FERT, Basic Data	2.00	0.75	0.00	0.00	
5	Wrong data? Description aus PLM wird nicht akzeptiert	5	2	0	0	
6	Missing data? Kurbezeichnung (DE) im lokalen System	3	1	0	0	
7	Out-dated data?	0	0	0	0	
8	Inconsistent data?	0	0	0	0	
9	Consumer satisfaction?					
10	Trust in data?					
11	Primary Improvements? Zentral vorgegebene Beschreibung für Artikel (DE, EN), um					
12	Material Master FERT, Logistical Data (prim. level)	0.75	0.50	1.50	0.00	
21	Material Master FERT, Logistical Data (sec. level)	0.50	0.50	1.50	0.00	
30	Material Master FERT, GTINs	0.50	0.00	0.75	0.00	
39	Material Master FERT, EH&S Data	1.25	1.00	0.00	0.75	
48	Material Master FERT, Storage Data	0.50	1.00	0.00	0.00	

Abbildung B3-13: Dokumentationsvorlage zur Bewertung von Kausalketten

Bei der Bewertung von Kausalketten ist eine äquidistante und über verschiedene Workshops vergleichbare Interpretation der Bewertungsskala sicherzustellen, beispielsweise durch eine Erläuterung, wie die Bewertungsstufen einer verwendeten Likert-Skala zu interpretieren sind. Bewertungsdimensionen sollte der Konzern-Datensteward so wählen, dass sie einfach verständlich sind und die Unternehmensziele repräsentieren (z. B. *Häufigkeit eines Geschäftsproblems*, *Beeinträchtigung von Kundenzufriedenheit*, *Produktqualität oder Service Level*, *Risiko von Gesetzesverstößen*, *zusätzlicher Arbeitsaufwand*, *Lagerkosten*, *Produktionsausschuss* oder *Transportkosten*). Um die Vergleichbarkeit der Bewertungen zu gewährleisten, sollte der Konzern-Datensteward in einer Interpretationshilfe alle Dimensionen auf Kosten (oder ein anderes einheitliches Kriterium) abbilden und für jede Bewertungsstufe die ungefähren Kosten angeben.

Die Art der Technikanwendung ist stark von spezifischen Rahmenbedingungen des jeweiligen Unternehmens beeinflusst (z. B. Verfügbarkeit und Verteilung von Mitarbeitern) und kann daher nicht von der Methode vorgegeben werden. Die folgenden drei Anwendungsarten sind bereits erfolgreich durchgeführt worden (s. Anhang E.7.5 für Dokumentationsvorlagen).

- *Aufgabenspezifische Interviews*. Interviews mit einem oder mehreren Mitarbeitern des Unternehmens, die eine bestimmte Tätigkeit (z. B. Call Center-Agent, Kundendienstmitarbeiter) ausführen. Bei diesen Interviews steht nicht die Diskussion abteilungs- oder rollenübergreifender Kausalketten, sondern die Konkretisierung bekannter Geschäftsprobleme und Datendefekte im Vordergrund. Anhang E.4.2 stellt dazu Leitfragen bereit.
- *Abteilungsspezifischer Workshop*. Workshop mit mehreren Mitarbeitern einer Fachabteilung (z. B. Finanzabteilung, Marketing, Produktentwicklung), einer Unterstützungsfunktion (z. B. Entscheidungsunterstützung, Qualitätsmanage-

ment) oder eines Produktionswerks. Die Durchführung dieser Workshop-Art ist sinnvoll, wenn Mitarbeiter geographisch weit verteilt arbeiten oder eine grössere Zahl an Mitarbeitern und Unternehmensteilen einbezogen werden soll. Zur Unterstützung stellt Anhang E.4.1 exemplarische Kausalketten bereit. Zusätzlich bietet die Technik *Datenqualitätssimulation* die Möglichkeit, diskutierte Kausalketten zu simulieren.

- *Abteilungsübergreifender Workshop*. Workshop mit mehreren Mitarbeitern aus unterschiedlichen Abteilungen. Diese Workshop-Art ist zu bevorzugen, da sie eine den gesamten Betrachtungsbereich abdeckende Diskussion ermöglicht. Der Konzern-Datensteward sollte die Teilnehmerzahl jedoch auf 8 bis 15 Personen beschränken, um trotz der willkommenen Breite eine fokussierte Diskussion zu gewährleisten.

Die Anwendung der Methode hat gezeigt, dass neben den gesuchten Kausalketten in einem Workshop auch mögliche Massnahmen zur Datenqualitätsverbesserung diskutiert werden. Diese Informationen sind zwar nicht unmittelbar für die Spezifikation von DQ-Kennzahlen notwendig, der Konzern-Datensteward sollte sie aber trotzdem zur Unterstützung möglicher späterer Projekte dokumentieren (vgl. Abbildung B3-8 und Dokumentationsvorlage in Anhang E.7.5).

3.5.2.4 Anforderungen definieren und priorisieren

Ziel von Aktivität II.1 ist die Definition eines Anforderungskatalogs, also einer unternehmensspezifischen, nach Wichtigkeit geordneten Liste mit Anforderungen, die eine DQ-Kennzahl zu erfüllen hat (vgl. Tabelle B3-25). Nach der Spezifikation der DQ-Kennzahlen in Aktivität II.2 überprüft der Konzern-Datensteward in Aktivität III.1 die Erfüllung der Anforderungen. Da die Anforderungen die an die DQ-Kennzahlen gestellten Erwartungen repräsentieren, sollte der Konzern-Datensteward den Anforderungskatalog mit dem Auftraggeber abstimmen. Eine Priorisierung der einzelnen Anforderungen unterstützt zudem die Prüfung und ggf. die Anpassung des Anforderungskatalogs in Aktivität II.2.

Richtlinien der Telefon AG fordern, Kennzahlen grundsätzlich in eine im gesamten Unternehmen genutzte Balanced Scorecard zu integrieren. Diese Anforderung impliziert die Notwendigkeit, bei der Spezifikation der DQ-Kennzahlen (insbesondere bei der Auswahl von Skalenniveau und Masseinheit) Struktur und Masseinheiten des Kennzahlensystems der Balanced Scorecard zu berücksichtigen. Die Telefon AG hat daher die Verwendung einer Prozentskala (optimal: 100%) für die Gestaltung der DQ-Kennzahlen als Anforderung definiert.

Aktivität II.1: Anforderungen definieren und ordnen Ziel ist die Definition eines Anforderungskatalogs, also einer unternehmensspezifischen, nach Wichtigkeit geordneten Liste mit Anforderungen, die eine DQ-Kennzahl zu erfüllen hat.		
Ergebnisdokument Liste mit Anforderungen	Rollen <ul style="list-style-type: none"> • Fachliche Datenstewards • Konzern-Datensteward • Auftraggeber • Technische Datenstewards 	Techniken <ul style="list-style-type: none"> • Interview/Workshop
Checkliste <ul style="list-style-type: none"> <input type="checkbox"/> Die definierten Anforderungen berücksichtigen alle Unternehmensrichtlinien (z. B. bereits existierende Kennzahlensysteme, Berichtsvorlagen, Normen). <input type="checkbox"/> Der Anforderungskatalog enthält keine widersprüchlichen Anforderungen. <input type="checkbox"/> Die Priorisierung repräsentiert die Wichtigkeit der einzelnen Anforderungen. 		

Tabelle B3-7: Methodenfragment zu Aktivität II.1

Typische Anforderungen betreffen die Darstellung von Messwerten, spätere Einsatzbereiche oder die Struktur des Kennzahlensystems. Zur Unterstützung der Aktivität beinhaltet die Methode eine Liste mit Anforderungen an DQ-Kennzahlen (s. Anhang E.5). Diese Anforderungen können die Diskussion in einem Workshop (vgl. Technik *Interview/Workshop*, Abschnitt 3.5.2.1.2) zwischen Konzern-Datensteward und verschiedenen Data Stewards (die für die von den Kausalketten betroffenen Datenobjekttypen zuständigen Stewards) unterstützen. Anhang E.7.4 stellt eine Vorlage zur Dokumentation der Anforderungen bereit.

Die Strom AG hat den Vergleich mit anderen Unternehmen (Benchmarking) als Ziel für den Einsatz der DQ-Kennzahlen formuliert. Um die spätere Vergleichbarkeit von DQ-Messungen zu gewährleisten, sollten Validierungsregeln nur Attribute standardisierter Datenmodelle (z. B. Datenfelder einer nicht modifizierten Standard-Software) betrachten.

Die Spezifikation von DQ-Kennzahlen in der durch die Methode vorgeschlagenen Form, d. h. mit Identifikation von Kausalketten (vgl. Abschnitte 3.5.2.1 und 3.5.2.3), der Verwendung von Validierungsregeln als Messtechnik (vgl. Abschnitt 3.5.2.5.1) und der Abbildung von Messwerten auf einer Prozentskala (vgl. 3.5.2.5.2), gewährleistet grundsätzlich die Erfüllung verschiedener Anforderungen.

- *Aggregierbarkeit.* Die Verwendung einer einheitlichen Berechnungsvorschrift (vgl. Technik *Spezifikation von Berechnungsvorschriften*, Abschnitt 3.5.2.5.2) im Messverfahren der DQ-Kennzahlen gewährleistet grundsätzlich die Möglichkeit, Messwerte verschiedener Kennzahlen zu aggregieren. Dabei sind jedoch u. U. unterschiedliche Einschränkungen der geprüften Daten (vgl. Aktivität I.2) zu beachten.
- *Akzeptanz.* Die durch Interviews oder Workshops in Aktivität I.3 angeregte Diskussion über mögliche DQ-Messungen erleichtert die Einführung der Kennzahlen und fördert die Akzeptanz ihrer Nutzung.

- *Beeinflussbarkeit*. Datendefekte als grundlegende Struktur der DQ-Kennzahlen liefern als Ergebnis des Spezifikationsprozesses für jede Kennzahl konkrete Ansatzpunkte für DQM-Massnahmen, um die überwachten Defekte zu beheben bzw. zu verhindern.
- *Geschäftsbezug*. Die Identifikation geschäftskritischer Datendefekte (vgl. Aktivität I.3) und die Dokumentation der Kausalketten (vgl. Aktivität II.2) gewährleistet, dass jede Validierungsregel und jede DQ-Kennzahl mindestens einem Geschäftsproblem zugeordnet werden kann.
- *Messbarkeit*. Die Spezifikation von Validierungsregeln als Messverfahren (vgl. Technik *Spezifikation von Validierungsregeln*, Abschnitt 3.5.2.5.1) gewährleistet, dass Messwerte für die spezifizierten DQ-Kennzahlen ermittelt werden können.
- *Normierung*. Die Verwendung der in Abschnitt 3.5.2.5.2 erläuterten Berechnungsvorschriften gewährleistet die Interpretation von DQ-Messungen auf einer Verhältnisskala und somit die Vergleichbarkeit von Messwerten.

3.5.2.5 Datenqualitätskennzahlen spezifizieren

Ziel von Aktivität II.2 ist die Spezifikation von DQ-Kennzahlen zur Überwachung der zuvor identifizierten Datendefekte (vgl. Tabelle B3-26). Dazu sind die Datenbehälter, für deren Datenelemente die Messung von Kennzahlenwerten durchgeführt werden soll (Messpunkte), die zu prüfenden Daten, die Zeitpunkte der Messungen (Messfrequenz), eine Messskala, eine Masseinheit und ein Messverfahren zu beschreiben.

Zentrale Aufgabe des Konzern-Datenstewards ist die Spezifikation der Messverfahren, d. h. die Gestaltung von Verfahren die untersuchen, ob die geprüften Daten einen oder mehrere der in Aktivität I.3 identifizierten geschäftskritischen Datendefekte aufweisen oder nicht. Die im Metamodell (vgl. Abbildung B3-8) aufgeführten Attribute *Einschränkungen*, *Messfrequenz* und *Messpunkt* sind zwar ebenfalls für die spätere Implementierung der Kennzahlen entscheidend, können aber oft erst im Verlauf erster Tests (z. B. Prüfung einzelner Datenobjekte durch spezifizierte Validierungsregeln) konkretisiert werden. Zwei Arten von Messverfahren lassen sich unterscheiden.

- *Befragung von Datennutzern*. In Interviews oder Umfragen kann die Qualität von Daten durch Nutzer dieser Daten bewertet werden [Lee et al. 2002, Price et al. 2008]. Ein Interviewleitfaden oder ein Fragebogen ist somit ein mögliches Messverfahren, das insbesondere dann zu wählen ist, wenn die identifizierten Datendefekte technisch schwer fassbaren (*subjektiv* nach PRICE ET AL. [2005]) DQ-Dimensionen wie *Objektivität*, *Vertrauen* oder *Verständlichkeit* zuzuordnen sind.

- *Überprüfung durch Regeln.* Die Qualität eines Datenobjekts kann durch die Überprüfung bestimmter Eigenschaften der Datenelemente oder -attribute gemessen werden [Loshin 2001, Olson 2003, Fan et al. 2008]. Weist ein Datenobjekt alle geforderten Eigenschaften auf, ist seine Qualität einwandfrei. Eigenschaften von Daten (z. B. letzter Änderungszeitpunkt eines Datenelements, Vollständigkeit eines Datenelements) können als Regeln formuliert und dadurch automatisch überprüft werden. Solche Validierungsregeln zur Prüfung von für ein Datenobjekt (in einem bestimmten Kontext) geforderten Eigenschaften sind somit ein mögliches Messverfahren.

<p>Aktivität II.2 Ziel ist die Spezifikation mindestens einer DQ-Kennzahl pro zuvor identifizierter Kausalkette. Dazu sind die Datenbehälter, für deren Datenelemente die Messung von Kennzahlenwerten durchgeführt werden soll (Messpunkte), die zu prüfenden Daten, die Zeitpunkte der Messungen (Messfrequenz), eine Messskala, eine Masseinheit und ein Messverfahren zu beschreiben. Als Messverfahren schlägt die Methode Validierungsregeln vor und unterstützt die Gestaltung von Regeln und Berechnungsvorschriften durch die Beschreibung von zwei Techniken.</p>		
<p>Ergebnisdokument Liste mit Spezifikationen von DQ-Kennzahlen</p>	<p>Rollen</p> <ul style="list-style-type: none"> • Fachliche Datenstewards • Konzern-Datensteward • Technische Datenstewards 	<p>Techniken</p> <ul style="list-style-type: none"> • Datenqualitätssimulation • Interview/Workshop • Spezifikation von Berechnungsvorschriften • Spezifikation von Validierungsregeln
<p>Checkliste</p> <ul style="list-style-type: none"> <input type="checkbox"/> Jede DQ-Kennzahl ist vollständig spezifiziert, d. h. die zu prüfenden Daten, ein Messpunkt, eine Messskala, eine Masseinheit, eine Messfrequenz und ein Messverfahren sind jeweils beschrieben. <input type="checkbox"/> Die Spezifikation der Messverfahren berücksichtigt alle zuvor dokumentierten Kausalketten, soweit keine Begründung, warum die Kausalkette nicht berücksichtigt wird, angegeben ist. <input type="checkbox"/> Alle spezifizierten Messverfahren können an den spezifizierten Messpunkten implementiert werden. 		

Tabelle B3-8: Methodenfragment zu Aktivität II.2.

Grundsätzlich sind beide Verfahren geeignet, um Daten auf mögliche Defekte zu prüfen. Da Interviews oder Umfragen jedoch bei jeder Messung (z. B. der Befragung) Ressourcen binden und daher nicht häufig durchgeführt werden können, fokussiert die vorgestellte Methode die Datenüberprüfung mit Validierungsregeln als Messverfahren. Zur Vorbereitung von Interviews und Befragungen enthalten die Anhänge E.6.1, E.6.2 und E.6.3 Leitfragen [Lee et al. 2002, Price et al. 2008].

Die Technik *Spezifikation von Validierungsregeln* unterstützt den Konzern-Datensteward bei der Formulierung von Validierungsregeln. Soweit die in Aktivität I.3 gesammelten Informationen nicht ausreichen, kann er ausserdem die Technik *Interview/Workshop* nutzen, um z. B. in Interviews mit Data Stewards die zu prüfenden Daten exakter einzuschränken, die Verfügbarkeit der Datenelemente und Metadaten (z. B. Zeitpunkt der letzten Änderung eines Attributs) an den möglichen Messpunkten zu prüfen oder Datenmodelle für exakte Bezeichnungen von Attributen zu bekommen.

Da die Anwendungsart der Technik *Interview/Workshop* stark von dem Ergebnis von Aktivität I.3 abhängt, bietet die Methode hier keine weitere Unterstützung in Form von Leitfragen oder spezifischen Dokumentationsvorlagen. Der Konzern-Datensteward kann sich jedoch an den Dokumentationsvorlagen in Anhang E.7.5 orientieren und prüfen, welche Informationen ihm bereits vorliegen und welche nicht.

Um Aussagen über Datenqualität mit Geschäftsbezug (z. B. „die Qualität der für Kundenanschriften benötigten Daten liegt bei nur 70%“) machen zu können, sollte der Konzern-Datensteward bestimmte Validierungsregeln bündeln und die Prüfungen der einzelnen Regeln zu einem Messwert zusammenführen (vgl. Technik *Spezifikation von Berechnungsvorschriften*, Abschnitt 3.5.2.5.2).

Aus der Analyse der identifizierten Kausalketten haben sich bei der Konsumgüter AG 32 Validierungsregeln für 7 Problembereiche (z. B. Gefahrgutindikatoren, Stücklistenqualität, Zollinformationen) ergeben. Für jeden Problembereich aggregiert eine Berechnungsvorschrift (gleiches Berechnungsschema für alle Kennzahlen) die Prüfergebnisse der jeweils zugeordneten Validierungsregeln zu einem Messwert.

Um das Kennzahlensystem nicht zu komplex zu gestalten (d. h. nicht noch weitere Berechnungsebenen einzuführen), ist es sinnvoll, genau eine Berechnungsvorschrift pro Datenqualitätskennzahl zu formulieren und für alle Berechnungsvorschriften das gleiche Schema zu verwenden (vgl. Abschnitt 3.5.2.5.2). So verwenden zudem alle Kennzahlen die gleiche Skala und Masseinheit. Zur Unterstützung des Konzern-Datenstewards stellt Anhang E.7.5 Dokumentationsvorlagen für die Spezifikation von Validierungsregeln und die Aggregation der Regeln durch eine Berechnungsvorschrift bereit.

3.5.2.5.1 Technik Spezifikation von Validierungsregeln

Die Technik unterstützt den Konzern-Datensteward bei der Spezifikation von Validierungsregeln als Messverfahren für DQ-Kennzahlen und nutzt die in Aktivität I.3 erzeugten Ergebnisdokumente. Ziel ist die Spezifikation einer oder mehrerer Validierungsregeln für jeden in Aktivität I.3 identifizierten geschäftskritischen Datendefekt.

Ein Datendefekt ist eine Eigenschaft eines Attributs oder eines Datenelements, die nicht den für den zugehörigen Datenobjekttyp spezifizierten (kontextspezifischen) Anforderungen entspricht. Ein fehlendes Datenelement eines Attributs zur Repräsentation des Vornamens eines Kunden ist beispielsweise ein Datendefekt, wenn in einer Prozessaktivität Kunden personalisierte Anschreiben bekommen sollen. Bei der Rechnungserstellung ist der Vorname nicht notwendig (anders als wieder beim Rechnungsversand) und daher das Fehlen des Datenelements kein Datendefekt.

Validierungsregeln können Datendefekte aufdecken, indem sie die Datenobjekte (z. B. Kundendaten) sukzessive überprüfen. Konkret überprüft eine Validierungsregel, ob ein

Attribut oder ein Datenelement eine bestimmte Eigenschaft hat oder nicht. Die Regeln dienen also der Validierung von Daten – wird in einem Datenobjekt kein durch die Validierungsregeln repräsentierter Datendefekt gefunden, gilt das Datenobjekt als korrekt (vgl. Differenzierung in Abschnitt 3.5.2.5.2). Basierend auf einer Literaturrecherche (vgl. Abschnitt 3.3) und Projekterfahrung (vgl. Beispiele) können die folgenden Regelschemata unterschieden werden. Abbildung B3-15 zeigt Beispielregeln für die drei Regelschemata in einer Pseudocode-Notation.

- *Wertprüfung*. Die Regel prüft, ob ein Datenelement vorhanden ist oder ob ein Datenelement im geforderten Definitionsbereich (z. B. vordefinierte Werte, Zahlenbereiche) liegt.
- *Wertevergleich*. Die Regel prüft, ob die Datenelemente von zwei oder mehr Attributen gleich sind oder ihr Abstand innerhalb einer bestimmten Toleranz liegt.
- *Metadatenprüfung und -vergleich*. Die Regel ermittelt Metadaten eines Attributs und führt eine Wertprüfung oder einen Wertevergleich dieser Metadaten durch.

Kennzahlen	Genauigkeit	Konsistenz	Vollständigkeit	Anzahl Regeln
Geschäftspartner	VR03, VR06	VR07	VR01, VR02, VR04, VR05	7 Regeln
Vertrag	VR08	—	—	1 Regeln
Vertragskonto	VR10	VR09, VR11	—	3 Regeln
Neukunde (Einzug)	VR03, VR06	VR07	VR01, VR02, VR04, VR05	7 Regeln
Rechnungserstellung	VR08, VR10	VR07	—	3 Regeln
Briefkampagne	VR03, VR06	—	VR01, VR02, VR04, VR05	6 Regeln
Zahlung	—	VR11	—	1 Regeln
Änderung (Kundeninfo)	VR03, VR06	VR07	VR01, VR02, VR04, VR05	7 Regeln
Rechnungsversand	VR03, VR06	VR09	VR01, VR02, VR04, VR05	7 Regeln
Versand Kundenzeitschrift	VR03, VR06	VR07	VR01, VR02, VR04, VR05	7 Regeln
DQ-Index	VR03, VR06, VR08, VR10	VR07, VR09, VR11	VR01, VR02, VR04, VR05	11 Regeln

Legende:

VR01: Anrede Rolle Person (Vollständigkeit)	VR04: Name (Vollständigkeit)	VR08: Abrechnungssperrgrund (Genauigkeit)
VR02: Anrede Rolle Organisation (Vollständigkeit)	VR05: Anschrift (Vollständigkeit)	VR09: Rechnungs- / Mahnepfänger (Konsistenz)
VR03: Anrede (Genauigkeit)	VR06: Anschrift (Genauigkeit, Dubletten)	VR10: Rechnungsformular (Genauigkeit)
	VR07: Sprache (Konsistenz)	VR11: Eingangs- / Ausgangszahlweg (Konsistenz)

Abbildung B3-14: Zuordnung von Validierungsregeln (VR) zu DQ-Kennzahlen

Die Anhänge E.6.4, E.6.5 und E.6.6 enthalten verschiedene Beispiele von Validierungsregeln und können den Konzern-Datensteward bei der Spezifikation unterstützen. Anhang E.7.5 stellt ausserdem eine Dokumentationsvorlage bereit, die ein einfaches Schema zur Dokumentation von zu prüfenden und zu vergleichenden Attributen enthält (vgl. Abbildung B3-14).

Der Konzern-Datensteward kann diese Vorlage in Interviews (vgl. Abschnitt 3.5.2.1.2) mit den für den jeweils fokussierten Datenobjekttyp zuständigen Data Stewards nutzen, um diskutierte Datendefekte direkt in einer Struktur zu dokumentieren, die der Prüf- und Vergleichslogik der Validierungsregeln ähnelt. Eine weitere Dokumentationsvorlage aus Anhang E.7.5 unterstützt die Zuordnung von Validierungsregeln zu DQ-Kennzahlen, die zwar bereits durch die in Aktivität I.3 identifizierten Kausalketten implizit vorgegeben ist, nach der Spezifikation der Validierungsregeln jedoch noch explizit dokumentiert werden sollte.

3.5.2.5.2 Technik Spezifikation von Berechnungsvorschriften

Bei der Prüfung von Datenobjekten durch Validierungsregeln verknüpft eine Berechnungsvorschrift die Ergebnisse einzelner Validierungsregeln zu einem Messwert. Wie oben bereits erwähnt ist es sinnvoll, genau eine Berechnungsvorschrift pro Kennzahl zu verwenden. In diesem Fall berechnet die Berechnungsvorschrift einer DQ-Kennzahl bei einer Messung den Messwert der Kennzahl aus den Ergebnissen der für die Messung verwendeten Validierungsregeln.

```
for each data object d in data base

    // Beispiele für Wertprüfungen
    if d.Attribute == NULL           : return 0
    if d.Attribute < 42              : return 0
    if d.Attribute != 'foo'         : return 0
    if !List.Contains(d.Attribute)  : return 0

    // Beispiele für Wertvergleiche
    if d.Attribut1 != d.Attribute2  : return 0
    if d.Attribut1 < d.Attribute2   : return 0
    if d.Attribut1 + 42 != d.Attribute2 : return 0

    // Beispiele für Metadatenprüfung und -vergleich
    if d.Attribute.Type != 'numeric' : return 0
    if d.Attribute.LastChange > Now - 2 weeks : return 0
    if d.Attribute.NumberOfChages > 42      : return 0
    if d.Attribut1.Type != d.Attribute2.Type : return 0

    return 1 // Datenobjekt ist korrekt

end for
```

Abbildung B3-15: Regelschemata mit Beispielen

Um die Akzeptanz der neu spezifizierten DQ-Kennzahlen zu gewährleisten, sollte die Berechnungsvorschrift grundsätzlich leicht verständlich sein. Je nach späterer Anwendung (z. B. Nutzung in Zielvereinbarungen) sollten Messwerte der Kennzahlen aber auch Effekte durchgeführter Massnahmen sichtbar machen und über längere Zeiträume vergleichbar sein. Um das Verständnis der Kennzahlen zu gewährleisten, sollten klare Aussagen wie „ein Datenobjekt ist dann korrekt, wenn es keine bekannten Defekte hat“ möglich sein.

Die folgenden Vorschläge für Berechnungsvorschriften gehen davon aus, dass einer Datenqualitätskennzahl m eine bestimmte Anzahl an Validierungsregeln $r \in R|_m$ zugeordnet ist, wobei die einzelnen Teilmengen $R|_{m_1}, R|_{m_2}, \dots, R|_{m_n}$ der Menge aller Regeln R nicht disjunkt sein müssen (Mehrfachzuordnungen können sinnvoll sein: Vollständige Adressdaten sind beispielsweise sowohl für den Rechnungsversand (Kennzahl 1), als auch für bestimmte Marketing-Initiativen (Kennzahl 2) erforderlich). Ausserdem wird das folgende Regelschema vorausgesetzt (vgl. Abbildung B3-15).

$$r(d) = \begin{cases} 0, & \text{wenn } d \text{ den durch } r \text{ repräsentierten Defekt hat.} \\ 1, & \text{wenn } d \text{ den durch } r \text{ repräsentierten Defekt nicht hat.} \end{cases} \quad (4-1)$$

Eine einfache Möglichkeit zur Berechnung eines Werts der Kennzahl m zur Messung der Qualität der Datenobjekte D (z. B. die Kundendaten eines Unternehmens) ist das Verhältnis der Anzahl defekter Datenobjekte $|D|_{\text{defekt}}$ zur Anzahl aller Datenobjekte $|D|$. Subtrahiert man diesen Wert von 1, erhält man den Anteil korrekter Datenobjekte, also einen Wert, der als *Datenqualität* $m(D)$ mit $0 \leq m(D) \leq 1$ auf einer Verhältnisskala interpretiert werden kann. Sind beispielsweise 58 von 100 geprüften Kundendatenobjekten defekt, wäre mit dieser Berechnungsvorschrift die Aussage „die Datenqualität der Kundendaten ist 42%“ möglich.

$$m(D) = 1 - \frac{|D|_{\text{defekt}}}{|D|} \quad (4-2)$$

Entscheidend bei dieser einfachen Berechnungsvorschrift ist, zu definieren, unter welchen Umständen ein Datenobjekt korrekt ist. Eine einfache und gut verständliche Definition ist, dass ein Datenobjekt aus Sicht einer DQ-Kennzahl m genau dann korrekt ist, wenn es keinen bekannten Datendefekt hat, also wenn keine Validierungsregel $r \in R|_m$ bei der Validierung des Datenobjekts d einen Defekt findet.

$$D|_{\text{korrekt}} = \left\{ d \in D \mid r(d) = 1 \forall r \in R|_m \right\} \quad (4-3)$$

Die Strom AG hat entschieden, ein Datenobjekt nur dann als korrekt zu definieren, wenn es keinen bekannten Datendefekt hat. Insbesondere die Klarheit und damit verbundene einfach Kommunizierbarkeit dieser Definition hat zu der Entscheidung geführt. Um Effekte von DQM-Massnahmen, die aufgrund der restriktiven Definition nicht in DQ-Messungen sichtbar werden, trotzdem nachweisen zu können, sollen auch Messwerte der Validierungsregeln, und nicht nur die aggregierten Messwerte der Kennzahlen, gespeichert und visualisiert werden.

Die Gefahr der geschilderten restriktiven Definition ist, dass Effekte von DQM-Massnahmen (z. B. Bereinigung von Dubletten) nicht im Verlauf der Messwerte einer

DQ-Kennzahl sichtbar werden, da bisher nicht behobene Defekte den Effekt überlagern. Ein Schwellwert s für den Anteil tolerierter Datendefekte in der Definition korrekter Datenobjekte kann die Sensibilität der DQ-Kennzahl erhöhen.

$$D|_{\text{korrekt}} = \left\{ d \in D \mid \frac{\sum_{i=1}^{|R|_m} r_i(d)}{|R|_m} \leq s \right\} \quad (4-4)$$

Ein weiteres Problem der bisher beschriebenen Berechnungsvorschriften ist, dass alle einer DQ-Kennzahl m zugeordneten Validierungsregeln für alle Datenobjekte des Datenobjekttyps D geprüft werden. Zwar ist eine allgemeine Einschränkung des Datenobjekttyps möglich (z. B. *nur aktive Datenobjekte* oder *nur Datenobjekte von in Europa verkauften Produkten*), eine Differenzierung der zu prüfenden Datenobjekte (d. h. eine Zuordnung bestimmter Datenobjekte $d_j \in D|_{r_i}$ zu bestimmten Validierungsregeln $r_i \in R|_m$) ist aber nicht möglich. Die folgende Berechnungsvorschrift enthält eine solche Differenzierung, jedoch auf Kosten einer komplexeren und damit schwerer zu kommunizierenden Berechnungsvorschrift.

$$m(D) = 1 - \frac{\sum_{i=1}^{|R|_m} \left(\frac{\sum_{j=1}^{|D|_{r_i}} r_i(d_j)}{|D|_{r_i}} \right)}{|R|_m} \quad (4-5)$$

Die bei der Konsumgüter AG identifizierten Datendefekte betreffen meist sehr spezifische Probleme, die nur bei wenigen Datenobjekten überhaupt vorkommen können. So sind beispielsweise fehlende Gefahrgutkennzeichen als Geschäftsproblem identifiziert worden, jedoch nur bei speziell für Marketingzwecke produzierten Kleinserien. Zu spät verfügbare Zolldaten (z. B. Produktionsland) führen ebenfalls zu Geschäftsproblemen, aber nur bei Datenobjekten zu Produkten, die über bestimmte Landesgrenzen hinweg transportiert werden. Da die Prüfung der spezifizierten Validierungsregeln auf allen Datenobjekten jedoch zu sehr geringen Fehlerquoten (>99% Datenqualität) führen würde, hat sich die Konsumgüter AG entschieden, pro Validierungsregel die zu prüfenden Datenobjekte möglichst exakt zu spezifizieren.

Gewichtungsfaktoren $\omega_i \geq 0$ ermöglichen zusätzlich die Steuerung des Einflusses einzelner Regeln r_i auf den berechneten Messwert (z. B. könnte eine Regel zur Prüfung von Gefahrgutkennzeichen höher gewichtet werden als eine Regel zur Prüfung der korrekten Produktklassifizierung). Hierbei ist jedoch wichtig, stets alle Gewichtungsfaktoren untereinander zu vergleichen und die beabsichtigten Effekte (z. B. „ r_1 ist doppelt so wichtig wie r_2 “) für spätere Änderungen zu dokumentieren.

$$m(D) = 1 - \frac{\sum_{i=1}^{|R_m|} \left(\frac{\sum_{j=1}^{|D_{r_i}|} \omega_i r_i(d_j)}{|D_{r_i}|} \right)}{\sum_{i=1}^{|R_m|} \omega_i} \quad (4-6)$$

3.5.2.5.3 Anwendung der Technik Datenqualitätssimulation in Aktivität II.2

Bei der Auswahl und Konfiguration der Berechnungsvorschrift kann die Simulation verschiedener Berechnungsmöglichkeiten insbesondere die Auswahl von Gewichtungsfaktoren und Schwellwerten unterstützen.

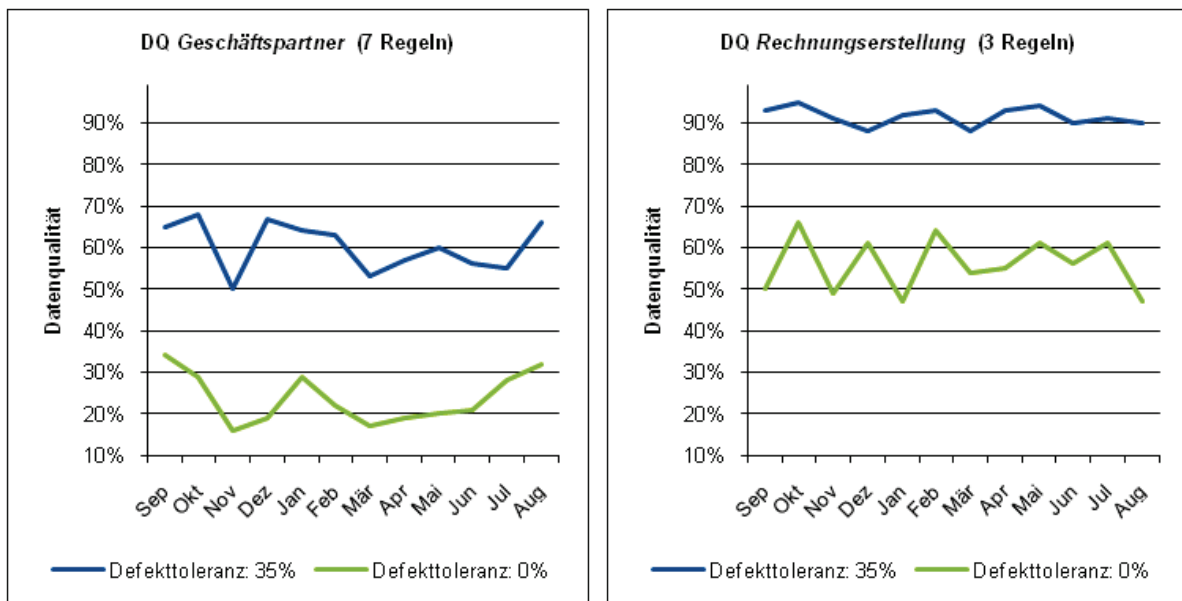


Abbildung B3-16: DQ-Messung mit unterschiedlichen Berechnungsvorschriften

Die Strom AG hat DQ-Messungen mit unterschiedlichen Berechnungsvorschriften simuliert, um die Konfiguration der ausgewählten Formel zu veranschaulichen. Dabei sind das Ergebnis der Überprüfung von 100 Datenobjekten durch 11 Validierungsregeln und die daraus berechneten Messwerte von 10 DQ-Kennzahlen simuliert worden. Die verwendete Berechnungsvorschrift basiert auf den Formeln (4-2) und (4-4). Abbildung B3-16 zeigt simulierte Messwerte für zwei Kennzahlen, jeweils mit einem Schwellwert von 0% (grün) und 35% (blau).

3.5.2.6 Erfüllung der Anforderungen prüfen

Aktivität III.1 prüft, ob die in Aktivität II.1 definierten Anforderungen von den in Aktivität II.2 spezifizierten DQ-Kennzahlen erfüllt werden (vgl. Tabelle B3-27). Falls hoch priorisierte Anforderungen nicht erfüllt sind, müssen die Anforderungen, ihre Priorisierung und die Kennzahlenspezifikationen überprüft und angepasst werden (d. h. Rücksprung zur Aktivität II.1). Wie in Abschnitt 3.5.2.4 für Aktivität II.1 beschrieben, sollte der Konzern-Datensteward die Anforderungserfüllung gemeinsam mit

den Data Stewards in einem Workshop (vgl. Abschnitt 3.5.2.1.2) prüfen. Die Dokumentationsvorlage aus Anhang E.7.6 unterstützt diese Aktivität.

Aktivität III.1: Erfüllung der Anforderungen prüfen Ziel ist zu prüfen, ob die in Aktivität II.1 definierten Anforderungen erfüllt sind. Falls hoch priorisierte Anforderungen nicht erfüllt sind, müssen Anforderungen, Priorisierung und die Spezifikation der DQ-Kennzahlen überprüft werden (d. h. Rücksprung zu Aktivität II.1).		
Ergebnisdokument Liste mit bewerteten Anforderungen	Rollen <ul style="list-style-type: none"> • Fachliche Datenstewards • Konzern-Datensteward • Technische Datenstewards 	Techniken <ul style="list-style-type: none"> • Interview/Workshop
Checkliste <ul style="list-style-type: none"> <input type="checkbox"/> Alle hoch priorisierten Anforderungen sind erfüllt. <input type="checkbox"/> Für alle nicht erfüllten Anforderungen, ist pro Kennzahl, die die Anforderung nicht erfüllt, eine Begründung angegeben. 		

Tabelle B3-9: Methodenfragment zu Aktivität III.1

3.5.2.7 Spezifikation dokumentieren

Aktivität III.2 ist die letzte Aktivität des Vorgehensmodells und liefert somit das Ergebnis einer Anwendung der Methode. Ziel der Aktivität ist die Dokumentation der Spezifikationen der DQ-Kennzahlen als implementierungsfähiges Fachkonzept, also in Form eines Dokuments, auf dessen Basis die spezifizierten Messverfahren und Skalen implementiert werden können (vgl. Tabelle B3-28).

Aktivität III.2: Spezifikation dokumentieren Ziel ist die Dokumentation der Spezifikationen der DQ-Kennzahlen als implementierungsfähiges Fachkonzept, also als Dokument, auf dessen Basis die spezifizierten Messverfahren und Skalen implementiert werden können.		
Ergebnisdokument Fachkonzept zur Implementierung von DQ-Kennzahlen	Rollen <ul style="list-style-type: none"> • Auftraggeber • Konzern-Datensteward 	Techniken <ul style="list-style-type: none"> • Dokumentation • Datenqualitätssimulation
Checkliste <ul style="list-style-type: none"> <input type="checkbox"/> Alle spezifizierten DQ-Kennzahlen sind dokumentiert, d. h. insbesondere Messverfahren und Skala für jede Kennzahl. <input type="checkbox"/> Für jede DQ-Kennzahl ist eine Kausalkette dokumentiert, also die Auswirkung der durch sie überwachten Datendefekte. 		

Tabelle B3-10: Methodenfragment zu Aktivität III.2

Aktivität III.2 (und damit die gesamte Methode) schliesst mit der Freigabe der dokumentierten Kennzahlenspezifikation (ggf. nach Anpassungen) durch den Auftraggeber.

3.5.2.7.1 Technik Dokumentation

Da Unternehmen meist mit ihrer IT-Abteilung oder externen IT-Dienstleistern abgestimmte Vorlagen für Fachkonzepte nutzen, enthält Anhang E.7.7 nur eine generische Gliederung für ein Fachkonzept mit Hinweisen zur Verwendung der im Verlauf der Methodenanwendung erzeugten Ergebnisdokumente.

Die Nutzung der Technik *Datenqualitätssimulation* ermöglicht es, die identifizierten und spezifizierten Elemente (Kausalketten, Validierungsregeln, Berechnungsvorschriften usw.) zu veranschaulichen. Insb. für die Abstimmung mit dem Auftraggeber ist die Modellierung verschiedener Szenarien und Kausalketten sinnvoll.

Die Telefon AG hat sechs Geschäftsprobleme (Fehlerfälle der Verfügbarkeitsprüfung) identifiziert und neun Validierungsregeln als Messverfahren einer DQ-Kennzahl spezifiziert. Auf der Basis von Fehlerprotokollen kann die Auftrittswahrscheinlichkeit der einzelnen Defekte geschätzt werden. Abbildung B3-17, Abbildung B3-18 und Abbildung B3-19 zeigen eine Simulation der Datendefekte, die dadurch verursachten Geschäftsprobleme und den Verlauf der Messwerte der DQ-Kennzahl. Die Validierungsregeln zur Überwachung der Datendefekte sind dabei paarweise disjunkt formuliert. D. h. wenn ein Datendefekt auftritt, dann verursacht er sicher einen Fehlerfall der Verfügbarkeitsprüfung, und keine Inkonsistenz wird von mehr als einem Datendefekt überwacht.

	De-fekt 1	De-fekt 2	De-fekt 3	De-fekt 4	De-fekt 5	De-fekt 6	De-fekt 7	De-fekt 8	De-fekt 9
Risiko	953, 61	734, 50	1146, 127	1028, 37	428, 35	643, 71	495, 84	761, 129	369, 62
Wirkung	Fehler 1	Fehler 1	Fehler 2	Fehler 3	Fehler 3	Fehler 4	Fehler 5	Fehler 5	Fehler 6

Tabelle B3-11: Parameter (Mittelwert, Varianz) normalverteilter Defekthäufigkeiten

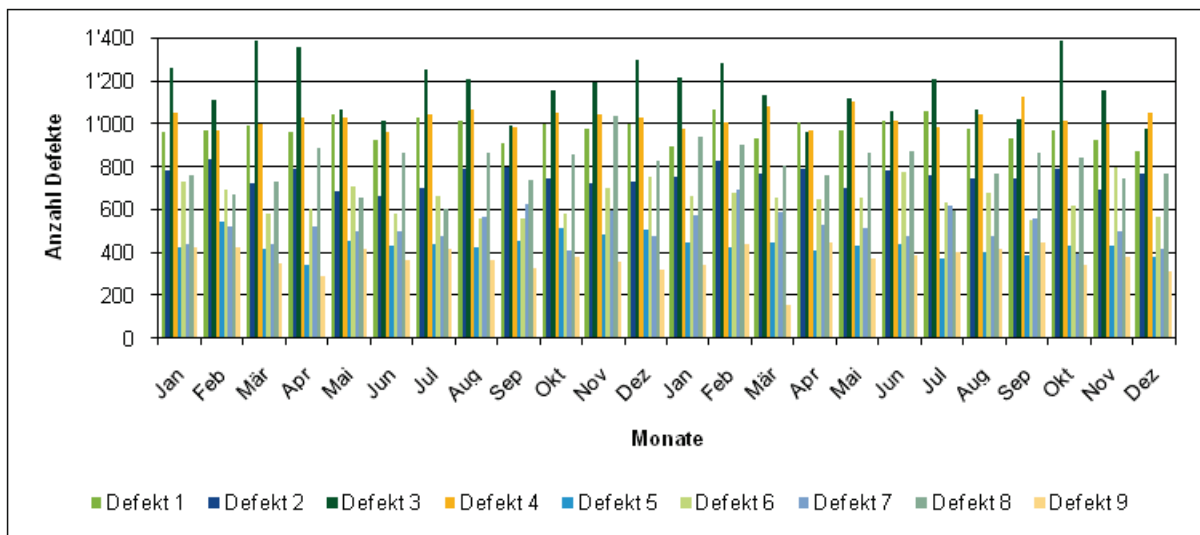


Abbildung B3-17: Simulation von Datendefekten

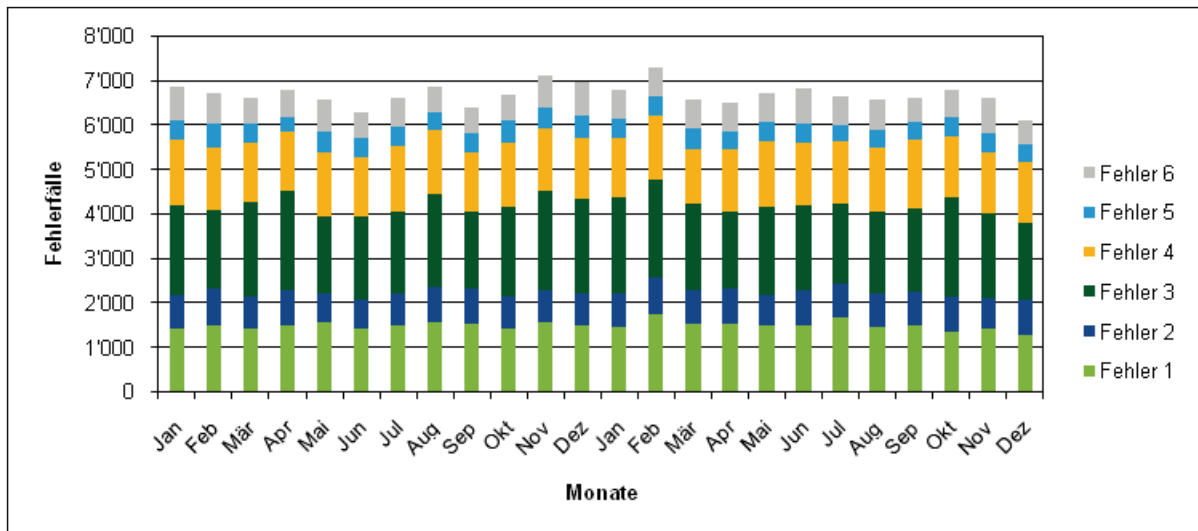


Abbildung B3-18: Durch Datendefekte verursachte Geschäftsprobleme

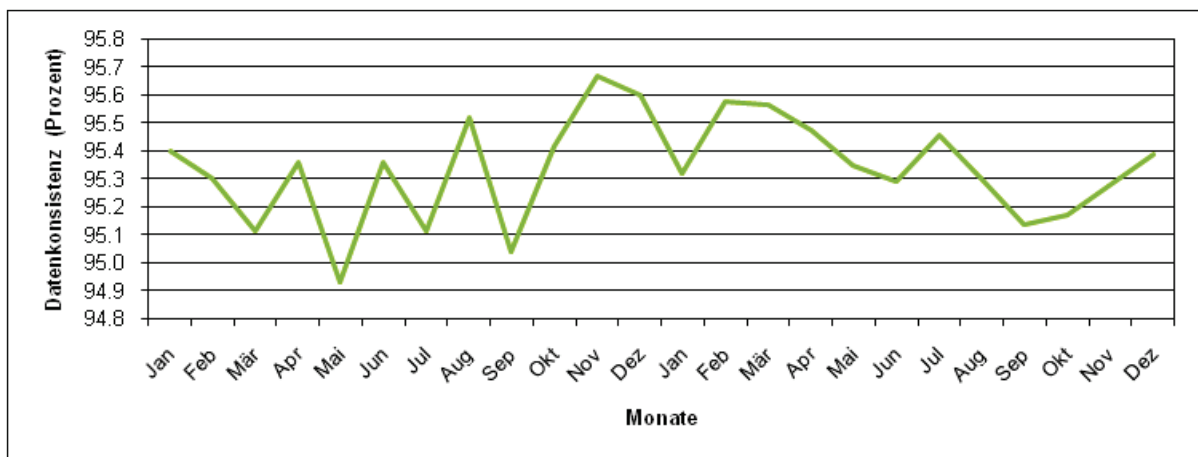


Abbildung B3-19: Überwachung von Datendefekten mit einer DQ-Kennzahl

3.6 Evaluation der Methode

Am 9. Februar 2010 haben elf Fachexperten aus zehn Unternehmen im Rahmen eines zweitägigen Workshops des CC CDQ in einem 90-minütigen Fokusgruppeninterview die Methode diskutiert und ihren Nutzen bewertet. Anhang E.8 zeigt die Teilnehmer (s. Tabelle E-21) und die Leitfragen (s. Tabelle E-22) des Fokusgruppeninterviews. Die Leitfragen zur Nutzenbewertung orientieren sich an Grundsätzen ordnungsmässiger Modellierung [Schütte/Rotthowe 1998, S. 245-249] und an Evaluationskriterien für Referenzmodelle [Frank 2007, S. 123-136].

Alle Fachexperten bestätigten die Notwendigkeit von DQ-Messungen und auch die Notwendigkeit methodischer Unterstützung bei der Gestaltung unternehmensspezifischer DQ-Kennzahlen (vgl. Tabelle B3-30). Die Ausrichtung von DQ-Kennzahlen an strategischen Unternehmenszielen wurde ebenfalls als wichtiges Gestaltungsziel bestätigt.

Relevanz-Perspektive (relevance perspective)
<ul style="list-style-type: none"> • DQ-Messungen sind wichtig, um reaktives DQM durch die präventive Identifikation kritischer Datendefekte zu ersetzen. • Die Methode hilft, die Komplexität der Spezifikation von DQ-Kennzahlen zur Überwachung geschäftskritischer Datendefekte zu reduzieren. • Die Methode adressiert mit der Spezifikation geschäftsorientierter DQ-Kennzahlen eine wichtige Aufgabe, berücksichtigt aber nicht ausreichend dynamische Prozessveränderungen, die identifizierte Kausalketten „altern“ lassen. • Die Berücksichtigung von Kausalketten zwischen Datendefekten und der Leistung von Geschäftsprozessen ist ein Erfolgsfaktor für die Implementierung wirksamer DQ-Kennzahlen.

Tabelle B3-12: Bewertungen der Fachexperten für die Relevanz-Perspektive

Aufgrund der hohen Relevanz von DQ-Messungen wurde der Bedarf an einer sorgfältigen Kosten-Nutzen-Analyse für die Spezifikation und Implementierung von DQ-Kennzahlen von einigen Fachexperten als gering eingestuft (vgl. Tabelle B3-31). Insbesondere Befragte aus Unternehmen, die in stark regulierten Märkten tätig sind (z. B. pharmazeutische und chemische Industrie), beschrieben in der Diskussion Datendefekte, die unmittelbar zu Verstößen gegen gesetzliche Vorgaben führen können (z. B. die Deklaration von Schadstoffanteilen in Pflanzenschutzprodukten, oder Haltbarkeitsangaben). Für die Spezifikation von DQ-Kennzahlen zur Überwachung solcher Datendefekte ist nach Meinung der Fachexperten eine Wirtschaftlichkeitsanalyse überflüssig, da die Unternehmen zur Überwachung gesetzlich verpflichtet sind. Andere Fachexperten haben die Notwendigkeit von Kosten-Nutzenanalysen bestätigt und auf die Möglichkeit verwiesen, die Spezifikation von DQ-Kennzahlen in andere Projekte einzubetten. Ein Unternehmen der chemischen Industrie hat Validierungsregeln bspw. im Verlauf der Harmonisierung von Einkaufsprozessen spezifiziert [Ebner et al. 2011].

Ökonomische Perspektive (economic perspective)
<ul style="list-style-type: none"> • Eine Kosten-Nutzen-Analyse ist nur für weniger kritische Datendefekte notwendig. Die Einhaltung gesetzlicher Vorgaben muss bspw. überwacht werden. • Validierungsregeln zur Überwachung von Datendefekten können ein Nebenprodukt bereits durchgeführter oder geplanter Projekte sein (z. B. Harmonisierung von Geschäftsprozessen, Implementierung eines neuen ERP-Systems). Die Identifikation solcher Projekte kann den für die Spezifikation und Implementierung von DQ-Kennzahlen veranschlagten Aufwand reduzieren.

Tabelle B3-13: Bewertungen der Fachexperten für die ökonomische Perspektive

Die Komplexität der Methode wurde von einigen Fachexperten kritisch bewertet, da sie hohen Erklärungsbedarf bei der Durchführung der einzelnen Aktivitäten befürchten (vgl. Tabelle B3-32). Überflüssige oder redundante Elemente konnten allerdings in einer Diskussion der Komplexität nicht identifiziert werden. Andere Fachexperten haben hingegen die für die Anwendung der Methode notwendige Kommunikation zwischen verschiedenen Rollen im Unternehmen positiv bewertet, da sie das grundsätzliche Bewusstsein für Datenqualität stärkt.

Einführungs-Perspektive (deployment perspective)

- Die Überwachung von Arbeitsergebnissen einzelner Mitarbeiter, also von Datenqualität als Ergebnis von Dateneingabe- und Datenpflegeprozessen, ist immer heikel. Um dennoch die Unterstützung der beteiligten Mitarbeiter zu bekommen, muss das Ziel von DQ-Messungen in Interviews und Workshops verständlich erklärt und veranschaulicht werden.
- Der Prozess zur Identifikation geschäftskritischer Datendefekte stärkt das Bewusstsein und die Aufmerksamkeit für Datenqualität und hilft, die Bedeutung von Datenqualität für den Erfolg des Unternehmens zu erläutern.
- Die Methode wäre mit weniger Komponenten und mehr Beispielen besser verständlich.
- Um das Problembewusstsein für Datendefekte zu erhöhen, sollten die Ergebnisse von DQ-Messungen an die Mitarbeiter berichtet werden, die Daten eingeben.

Tabelle B3-14: Bewertungen der Fachexperten für die Einführungs-Perspektive

Die Konstruktions-Perspektive (engl. engineering perspective) wurde ebenfalls kontrovers diskutiert, wobei die Fachexperten keine konkreten Änderungsbedarfe identifizieren konnten (vgl. Tabelle B3-33). Als konkrete Ergänzung wurde eine Datenkunden / -Lieferanten-Perspektive vorgeschlagen, um die Identifikation von Kausalketten zu unterstützen. In Workshops zur Identifikation von Kausalketten sollten Repräsentanten beider Sichten Ursache und Wirkung von Datendefekten diskutieren, um neben Kausalketten auch präventive Massnahmen identifizieren zu können. Einige Fachexperten haben ausserdem die Bereitstellung von Validierungsregeln für allgemeine Anforderungen (z. B. zur Überwachung gesetzlicher Vorgaben) angeregt, um den Spezifikationsprozess zu beschleunigen.

Konstruktions-Perspektive (engineering perspective)

- Geschäftsprobleme sind ein guter Startpunkt für die Identifikation kritischer Datendefekte.
- Es gibt zu viele Rollen, vier Rollen sollten ausreichen.
- Eine Sammlung allgemeiner DQ-Kennzahlen, z. B. zur Überwachung der Einhaltung gesetzlicher Vorgaben, wäre wünschenswert.
- Die Rollen sollten eine Datenkunden- und eine Datenlieferanten-Perspektive berücksichtigen.
- Die Methode sollte ausserdem die Definition von Berichtswegen für Ergebnisse von DQ-Messungen berücksichtigen.
- Monokausale Beziehungen zwischen einem Geschäftsproblem und einem Datendefekt (der nur eine DQ-Dimension betrifft) könnten nicht ausreichen, um reale Probleme zu modellieren. Geschäftskritische Kausalketten können komplexer sein.

Tabelle B3-15: Bewertungen der Fachexperten für die Konstruktions-Perspektive

Die Fachexperten haben die erkenntnistheoretische Perspektive (engl. epistemological perspective) der Methode (vgl. Anhang E.8, Tabelle E-22) nicht diskutiert, da die Qualität des Gestaltungsprozesses ihre Gesamtbewertung der Methode nicht beeinflusst. Nützlicher als die Erläuterung von Gestaltungsentscheidungen seien nach Meinung der Fokusgruppe weitere Beispiele für DQ-Kennzahlen und ihre Implementierung.

Insgesamt hat die Fokusgruppe den Bedarf geschäftsorientierter DQ-Kennzahlen und den Bedarf methodischer Unterstützung bei deren Spezifikation bestätigt. Die Methode selbst wurde zwar als komplex und erklärungsbedürftig, aber dennoch als anwendbar und nützlich bewertet. Erweiterungsbedarf wurde insbesondere für die Sammlung ex-

emplarischer Kausalketten (vgl. Anhang E.4.1) und Validierungsregeln (vgl. Anhang E.6.6) identifiziert.

3.7 Schlussfolgerungen und weiterer Forschungsbedarf

Der Bericht beschreibt eine Methode zur Identifikation geschäftskritischer Datendefekte und zur Spezifikation von DQ-Kennzahlen für die Überwachung der identifizierten Datendefekte. Als Messverfahren schlägt die Methode Validierungsregeln vor, mit denen ein Messsystem Eigenschaften von Datenelementen überprüfen und Datendefekte erkennen kann. Die durch die Anwendung der Methode spezifizierten DQ-Kennzahlen ermöglichen eine objektive Bewertung der Qualität ausgewählter Datenelemente und somit vergleichbare DQ-Messungen.

Die Methode ist ein komplexes Artefakt, dessen Anwendung Aufwand erzeugt. Daher ist im Einzelfall abzuwägen, welche Methodenelemente tatsächlich Nutzen stiften. So ist es beispielsweise nicht zwingend erforderlich, bei jeder Methodenanwendung alle Ergebnisdokumente zu erzeugen oder Anforderungen an Datenqualitätskennzahlen zu definieren und ihre Erfüllung zu prüfen. Die Methodenelemente sind vielmehr als strukturierte Sammlung von Handlungsanweisungen und Techniken zu verstehen, die der Konzern-Datensteward an unternehmensspezifische Anforderungen anpassen kann. Gleichwohl gibt die Methode eine sinnvolle Strukturierung der Elemente vor, die nur begründet geändert werden sollte.

Weiterer Forschungsbedarf besteht insbesondere im Ausbau der Sammlung von Kausalketten und Validierungsregeln. Die Anwendung der Methode in weiteren Fällen liefert dazu notwendige Ergebnisse und dient gleichzeitig der Evaluation der Methode selbst. Die Spezifikation von Validierungsregeln für allgemeine (d. h. nicht unternehmensspezifische) Datenqualitätsanforderungen (z. B. für gesetzliche Vorgaben) ermöglicht dabei die Nutzung gleicher DQ-Kennzahlen in verschiedenen Unternehmen. So können DQ-Messungen von unterschiedlicher Unternehmen verglichen und erfolgreiche DQM-Strukturen und -Massnahmen identifiziert werden. Zudem wird der Aufwand zur Spezifikation von DQ-Kennzahlen für einzelne Unternehmen reduziert.

4 Product data quality in supply chains: the case of Beiersdorf

Titel	Product data quality in supply chains: the case of Beiersdorf ¹⁹
Autoren	Kai M. Hüner; Andreas Schierning; Boris Otto; Hubert Österle Universität St. Gallen, Institut für Wirtschaftsinformatik Müller-Friedberg-Strasse 8, 9000 St. Gallen, Schweiz {kai.huener boris.otto hubert.oesterle}@unisg.ch Beiersdorf AG, Supply Chain Data Process Management Unnastraße 48, 20245 Hamburg, Deutschland andreas.schierning@beiersdorf.com
Publikationsorgan	Electronic Markets 21(2), S. 141-154
Status	Veröffentlicht

Tabelle B4-1: Bibliographische Angaben zum Beitrag „Product data quality in supply chains: the case of Beiersdorf“

Abstract

A number of business requirements (e.g. compliance with regulatory and legal provisions, diffusion of global standards, supply chain integration) are forcing consumer goods manufacturers to increase their efforts to provide product data (e.g. product identifiers, dimensions) at business-to-business interfaces timely and accurately. The quality of such data is a critical success factor for efficient and effective cross-company collaboration. If compliance relevant data (e.g. dangerous-goods indicators) is missing or false, consumer goods manufacturers risk being fined and see their company's image damaged. Or if logistics data (e.g. product dimensions, gross weight) is inaccurate or provided not in time, business with key account trading partners is endangered. To be able to manage the risk of business critical data defects, companies must be able to a) identify such data defects, and b) specify and use metrics that allow to monitor the data's quality. As scientific research on both these issues has come up with only few results so far, this case study explores the process of identifying business critical product data defects at German consumer goods manufacturing company Beiersdorf AG. Despite advanced data quality management structures such defects still occur and can result in complaints, service level impairment and avoidable costs. The case study analyzes product data use and maintenance in Beiersdorf's ecosystem, identifies typical product data defects, and proposes a set of data quality metrics for monitoring those defects.

4.1 Introduction

A number of business requirements (e.g. compliance with regulatory and legal provisions, the ongoing consolidation process in the retail industry, diffusion of global stan-

¹⁹ © Institut für Wirtschaftsinformatik Universität St. Gallen 2011

dards) are forcing consumer goods manufacturers to increase their efforts to provide product data (e.g. identifiers, dimensions) at business-to-business interfaces timely and accurately. The following examples are supposed to illustrate the situation.

Large retailers such as Walmart (approx. 422 billion USD revenue in 2010), Carrefour (approx. 90 billion EUR revenue in 2010) or Metro (approx. 67 billion EUR revenue in 2010) in recent years have put up strict requirements on cross-company processes, which consumer goods manufacturers are expected to meet. Carrefour, for example, demands that logistics related product data be provided as early as eight months before a product's launch in order to ensure high-quality intra-company supply chain planning. In order to adequately respond to this challenge, consumer goods manufacturers in general need to enhance their innovation and marketing processes by an integrated product data management process allowing to capture relevant data at an early stage and provide this data timely and in good quality.

Global Standards One (GS1), a leading standardization organization, has specified the Global Trade Item Number (GTIN), allowing to unambiguously identify trade items all over the world [GS1 2010, pp. 22-77]. Based on this numerical system, external service providers offer data pools for multi lateral synchronization of product data [Nakatani et al. 2006, pp. 970-977]. Consumer goods manufacturers making their product data available in such a pool are required to label every traded item so that each item and each logistic unit (e.g. shrink, packaging unit, pallet) is unambiguously identifiable by a unique GTIN.

The present paper primarily aims at a) identifying typical business critical product data defects occurring in a consumer goods manufacturer's ecosystem and b) proposing a set of data quality metrics for monitoring those defects. As product data defects and their business impact have been rarely investigated in scientific studies, case study research constitutes an appropriate research methodology. The paper is structured in five sections and continues with a brief overview of the state of the art regarding data quality measuring and cross-company master data management, followed by a methodological justification of the presented case study. The next section describes the initial situation at Beiersdorf and reports on how business critical defects were identified and how data quality metrics were specified. Identified defects and specified metrics are presented as well. Finally, the paper concludes with a short summary of the results and a brief discussion of implications for practice and research.

4.2 Data quality in cross-company supply chains

4.2.1 Data quality management

Various studies conceptualize data quality by data quality dimensions (e.g. accuracy, completeness, objectivity) on the basis of empirical research [Wang/Strong 1996], on-

tological and semiotic conclusions [Wand/Wang 1996, Price/Shanks 2005] or practitioners' experiences [Redman 1996, pp. 245-266, English 1999, pp. 87-118]. These studies have established the definition of data quality as the data's 'fitness for use', i.e. whether data is of good or poor quality depends on the context it is used in and the user it is used by. This context dependent data quality definition provides the basis for the specification of business-oriented data quality metrics: Data is of good quality if it meets business requirements (e.g. requirements of business partners or internal processes). Of course, requirements regarding one data object may vary from one process to another.

Data quality management comprises activities for improvement of data quality [Batini/Scannapieco 2006, pp. 69-71]. Going beyond mere reactive action (e.g. identification and correction of data defects) [Shankaranarayanan/Cai 2006, pp. 303-304], data quality management works as a preventive concept, characterized by a continuous cycle consisting of activities to define, measure, analyze and improve data quality [Wang et al. 1998, English 1999, pp. 69-81, Eppler/Helfert 2004]. Preventive data quality management includes the design and deployment of appropriate management structures such as data governance [Weber et al. 2009a, Khatri/Brown 2010] or the specification and implementation of data quality metrics [Heinrich et al. 2009]. An overview of the most relevant approaches for data quality management is given by Batini et al. [2009b].

4.2.2 Data quality measurement

Apart from the conceptualization of data quality by means of various data quality dimensions the measurement of data quality (i.e. of data quality dimensions) has been a central issue of many scientific studies. In this context, data quality metrics serve to operationalize data quality dimensions: They specify the data to be measured, a measuring point, a measuring technique, and a measuring scale [Batini/Scannapieco 2006, p. 19]. Some studies provide calculation instructions for single data quality dimensions, such as timeliness [Heinrich et al. 2007, Heinrich et al. 2009], accuracy [Gorla/Krishnan 2002], or completeness [Cai/Ziad 2003]. Other studies present procedures and techniques for measuring data quality by means of interviews and surveys [Lee et al. 2002, Nicolaou/McKnight 2006, Price et al. 2008] or by means of validation rules [Hipp et al. 2007, Fan et al. 2008]. For the process of identifying data defects and specifying data quality metrics in a specific context procedure models and analysis techniques have been proposed as well [Batini et al. 2007, Caballero et al. 2008, Gelinias/Dull 2009, Heinrich et al. 2009]. For the design of business-oriented data quality metrics (i.e. metrics for monitoring business critical data defects) causal relations between data defects, business operations problems, and strategic business goals should be analyzed [Otto et al. 2009a]. The objective is to identify those process activities a) the outcome of which is critical in achieving strategic business goals, and b) the out-

come of which is strongly dependent on the quality of the data used. Examples of such causal relations are presented in following sections.

4.2.3 Data sharing in supply chains

Studies that investigate cross-company data exchange distinguish between exchange of transaction data (for example, wholesalers and retailers reporting sales figures to the manufacturer to facilitate demand planning) and exchange of master data [Legner/Schemm 2008, p. 126]. While the former has been investigated in numerous scientific studies [Lee et al. 1997, Cachon/Lairiviere 2000, Chen et al. 2000, Lau et al. 2002, Kulp et al. 2004, Sheu et al. 2006], the latter has somewhat been neglected. Although it is common sense that electronic data exchange in supply chains [Beck/Weitzel 2005] and electronic marketplaces [Wang et al. 2006] hold benefits for most companies, and that correct interpretation of exchanged data is business critical [Madnick 1995, Goh et al. 1999, Vermeer 2000, Zhao 2007], business impacts of the quality of exchanged master data (both intra-company and cross-company) have not been thoroughly examined so far. Therefore, this paper investigates business problems that are caused by product master data defects. For monitoring those defects, the paper proposes data quality metrics that have been designed at Beiersdorf.

In general, master data specifies business objects, i.e. those essential business entities a company's business activities are based on. Such entities are, for example, business partners (customers, suppliers), products and the components and materials of which they are comprised, or employees [Smith/McKeen 2008, pp. 64-65]. Basically, master data can be differentiated by three concepts: master data class, master data attribute, and master data object [Loshin 2008, pp. 5-8]. A master data object represents a concrete business object (a product manufactured in a certain plant at a certain point in time, for example), and it specifies selected characteristics of this business object (color, features, or price, for example) by means of attributes. The attributes selected for representation of a specific class of business objects (customers or products, for example) constitute a master data class (which is usually specified by a data model).

4.3 Research methodology

The underlying research method of the presented work is participative exploratory case research [Scholz/Tietje 2002, pp. 11-12, Yin 2002, pp. 3-5]. The scientific discipline of information systems research often uses case study research to deal with organizational issues in which the boundaries between the real-world phenomenon itself (here: product data quality at the interfaces of Beiersdorf and their business partners) and its environment (here: the ecosystem of Beiersdorf's supply chain) cannot clearly be identified [Yin 2002, p. 13]. The present case study aims at

- describing a method based process of identifying business critical data defects,

- analyzing product data use and maintenance in a consumer goods manufacturer's ecosystem in order to identify typical product data defects, and
- proposing a set of data quality metrics for monitoring product data.

Conducting research this way is reasonable if a phenomenon has not been investigated to a large extent so far, or if a phenomenon has not been fully understood theoretically, or if during a research process questions to be tackled by future research are raised [Walsham 1995, van der Blonk 2003].

The data for this case study was collected over the period of four months (from February 2010 to June 2010, see Table B4-2) from varied sources including various company documents, observations made from attending project meetings and seven semi-structured interviews. This multi-method approach can help to generate data that is often rich in detail and rigor [Yin 2002, pp. 83-101]. As the authors of this paper have contributed methodology and concepts (i.e. the procedure of identifying data defects and specifying data quality metrics) to the observed project, presented results might contain biased interpretation as it is due to participative case studies [Baskerville 1997, p. 29]. However, during the course of interviews (i.e. the process of data collection), the authors have not been involved in discussions or the process of identifying data defects. Interview results (i.e. issue descriptions, ratings) were recorded in writing and were reviewed and approved by interviewees.

Date	Organizational unit	Meeting type (and topic)	Number of participants	Duration (hours)
February 15, 2010	Data Process Management	Project meeting (project planning)	3	3
February 26, 2010	Data Process Management	Project meeting (attribute selection and grouping, interview preparation)	3	6
March 3, 2010	Quality Management (Distribution)	On-site interview in person	1	2
March 17, 2010	Distribution Center (Eastern Europe)	Invited group interview in person	3	4
March 24, 2010	Third Party Manufacturer (mass aerosol products)	On-site group interview in person	2	3
March 30, 2010	Marketing and Sales	On-site group interview in person	2	3
March 31, 2010	Distribution Center (Central Europe)	On-site group interview in person	2	3
April 1, 2010	Plant (complex care products, <i>Eucerin</i>)	Invited group interview in person	2	3
April 7, 2010	Plant (mass products, <i>Nivea</i>)	On-site interview in person	4	4
April 26, 2010	Data Process Management	Project meeting (interview analysis)	3	6
May 10,	Data Process Management	Project meeting (metrics spe-	3	6

Date	Organizational unit	Meeting type (and topic)	Number of participants	Duration (hours)
2010	ment	cification)		
June 16, 2010	Data Process Management, Beiersdorf Shared Services, External Software Provider	Project meeting (planning of metrics implementation)	6	4

Table B4-2: Project meetings and interviews

4.4 The Case of Beiersdorf

4.4.1 Current situation

4.4.1.1 Company profile

Beiersdorf, with headquarters in Hamburg, Germany, is a global company having 150 subsidiaries and employing over 19,000 people all over the world [Beiersdorf 2010]. The company is divided in two business segments. The Consumer business segment (approx. 5.3 billion EUR revenue in 2010) offers products for the skin and beauty care market as well as medical products. *Nivea*, one of the oldest and most popular body care brands in the world, comprises a set of products contributing the biggest portion to the company's revenue. Other well-known products in this field are *8x4* and *Labello* in the mass distribution market, and *Juvena* and *La Prairie* in the exclusive-products market. Medical products offered comprise plasters, adhesive tape, sore protection, and bandages, distributed under brands such as *Eucerin* and *Hansaplast*. The *tesa* business segment (approx. 873 million EUR revenue in 2009) comprises self-adhesive system and product solutions both for the industry and for the consumer market.

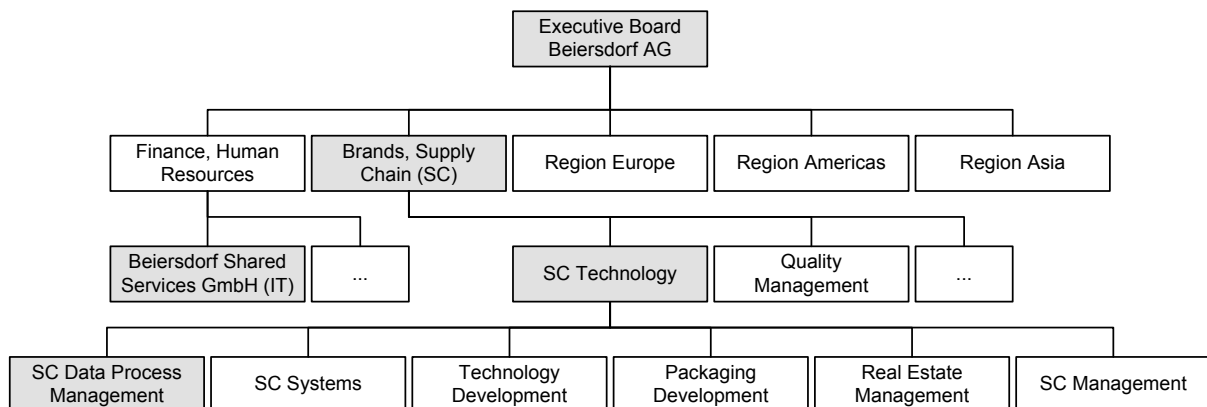


Figure B4-1: Organizational structure of Beiersdorf and reporting line of Data Process Management

Beiersdorf's organizational structure is divided into two functional and three regional areas of responsibility (see Figure B4-1). Information technology is managed by Beiersdorf Shared Services, a shared service organization and subsidiary of Beiersdorf. The organizational unit Supply Chain Data Process Management, located in the corpo-

rate supply chain division, is responsible for organization of company-wide product master data management.

4.4.1.2 Ecosystem

As the focus of the case study is on cross-company data exchange, Figure B4- shows a segment of Beiersdorf's ecosystem in which corporate functions form only one of eleven actors. The illustration differentiates between intra-company data flows (flows 1, 2, 8, and 12), data flows between Beiersdorf and external parties (flows 3, 4, 5, 6, 7, 9, 10, 11, 13, and 14) and flows between non-Beiersdorf entities (flows 15 and 16). Application systems involved in the exchange of data are not shown in the figure.

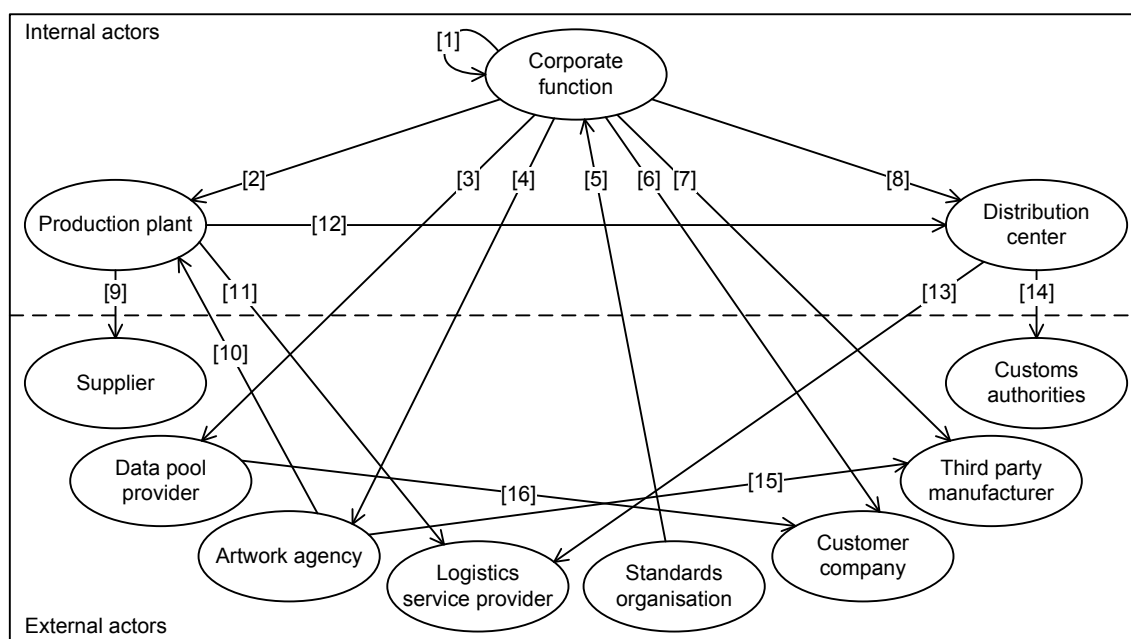


Figure B4-2: Product data exchange in the ecosystem of Beiersdorf

In the focus of the case study are the following actors and data flows from Beiersdorf's ecosystem:

- *Corporate function.* Central company functions (e.g. demand planning, product marketing, product development, packaging design) use and maintain product data (see [1] in Figure B4-; for example, a product recipe is created by the product development department and used by the demand planning function) and make product data available (for example, GTINs or net weights included in an artwork specification [4]) for external partners).
- *Production plant, supplier, third party manufacturer, artwork agency.* Globally distributed production plants and external production partners use, for example, item lists or GTINs [2, 7] as well as country specific artworks [10, 15], and order raw materials (positions of bills of material [9]) from suppliers.

- *Distribution center, logistics service provider, custom authorities.* For storage and transportation of goods, distribution centers use logistics data provided by corporate functions [8] and modified or added by production plants [12]. External service providers also use logistics data as well as environment, health and safety data [11, 13]. Customs relevant data [14] need to be made available to customs authorities.
- *Data pool provider.* Beiersdorf makes GTINs [3] available in a data pool to be used by their customer companies [16].
- *Customer company.* Beiersdorf provides logistics data to customer companies in order to support their planning processes ([6], particularly data on packaging dimensions).
- *Standards organization.* Beiersdorf requests data from standardization organizations for definition of new GTINs [5].

4.4.1.3 Data management organization

Tasks to be done by the Data Process Management unit (see Figure B4-1) comprise typical duties of a Chief Data Steward [Weber et al. 2009a, pp. 10-13], such as strategic development of master data management or further development of the central master data system. System support is provided by Beiersdorf Shared Services.

The head of Supply Chain (see Figure B4-1) represents executive sponsorship for master data management at Beiersdorf. Counterparts of the Data Process Management are people responsible for master data management in the different corporate or local functional departments, who take over a coordinating function for all master data related issues. This role is fulfilled by various central or local organizational units. On a corporate level, one person from the marketing department is chosen to be responsible for each product line (e.g. *Nivea Sun*, *Nivea Body*). And on the level of subsidiaries usually one Business Data Steward from the material management function is appointed per country.

Responsibility for the creation of the master data object of a specific product depends on the market the product is supposed to be sold. Product data representing products marketed only in one country is created by the local subsidiaries therein, whereas product data of internationally distributed products is created by brands on a corporate level. Who is then responsible for further maintenance depends on where the data was created. For the central company functions approx. 200 people are involved in the process of maintaining global data attributes, whereas on the local level usually the product management function manages master data [Schemm 2008, pp. 230-232].

4.4.1.4 Data management systems

For the management of global product data, such as identifiers, logistics data, and product hierarchies, Beiersdorf is using a central product lifecycle management (PLM) system, which was implemented in 2004. The PLM system at regular intervals (i.e. every three hours) provides new or changed product data to four regional enterprise resource planning (ERP) systems and a number of other global application systems (e.g. a decision support system (BW), a planning system (APO), a procurement system (EBP)). As the data is directly committed into the receiving systems, a consistent database is always ensured. The systems are operated by Beiersdorf Shared Services. Figure B4- illustrates the flows of master data within Beiersdorf’s application landscape. The application architecture depicted is typical for a global company, comprising both global applications supporting processes that affect several organizational units and local applications supporting processes within discrete organizational units [Lehmann 2003].

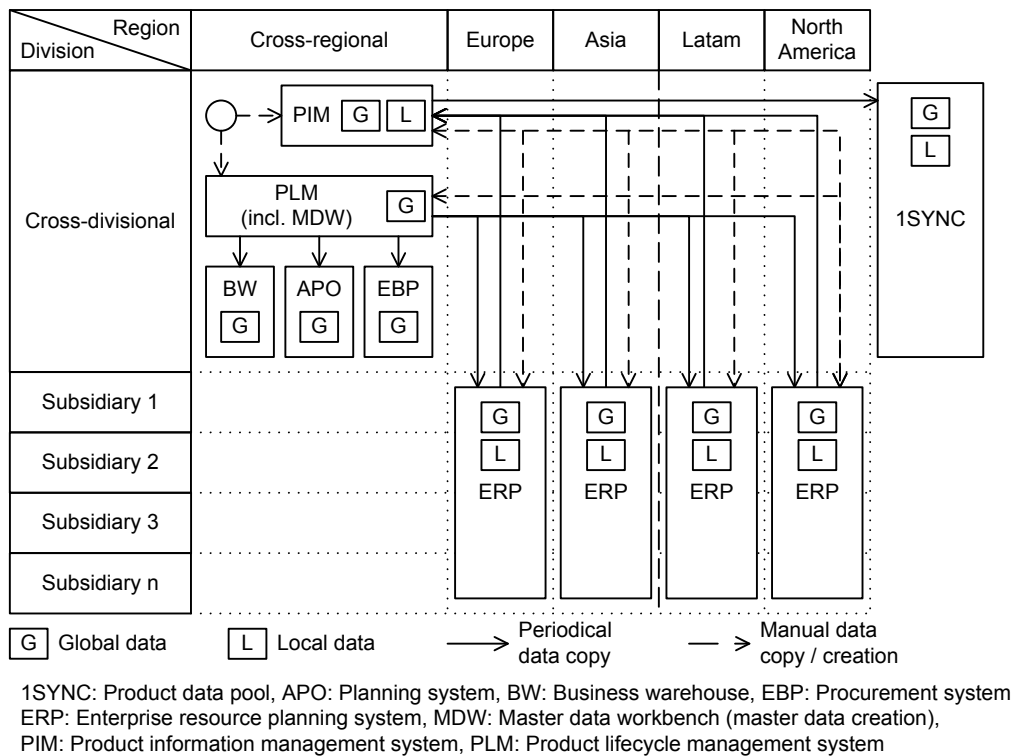


Figure B4-3: Master data flows within the application landscape of Beiersdorf

Being part of the PLM system, the Master Data Workbench (MDW) provides functionality for master data creation, thereby ensuring that master data is captured by the PLM system right at the moment it is being created. The users of the system (about 150) work with input masks specifically designed to match with the product innovation process and the product data maintenance process. The whole process of master data gathering and creation is done by means of one single application system. Fast and accurate data processing is ensured, since there is no media breakage and a number of PLM functions (e.g. allocation of unique identifiers, check routines) can already

be used in the process of data creation. After the release of a product master data record, the PLM system provides product data to the regional ERP systems.

Product data distribution to wholesalers and retailers is controlled by means of a central product information management (PIM) system. The PIM system is provided with global and local master data from the regional ERP systems. The PIM system also controls the data transfer to 1Sync, Beiersdorf's data pool.

4.4.2 Data quality measurement

4.4.2.1 Project objectives

Basically, the given organizational and technical master data management structures at Beiersdorf have been allowed for smooth cross-company supply chain processes. Regularly occurring business operations problems that counteract strategic business goals (e.g. compliance with legal and regulatory provisions, high service level) or that cause high costs have not regularly been reported. Nevertheless, there have been critical voices on the quality of product data, particularly with regard to their cross-company use. For example, some distribution centers have been complaining about the accuracy of weights of newly launched products (i.e. of related logistic units). Such defects of logistics data can result in additional costs due to repackaging and restocking if tolerance range for pallet weights is exceeded.

Noticing these complaints, together with the growing awareness that product data quality is of particular importance at the interfaces with external partners (e.g. customer companies, data pool providers, logistics service providers), Beiersdorf's central Data Process Management unit initiated a project aiming at a) the identification of business critical data defects, and b) the specification of data quality metrics for monitoring these defects. The project plan comprised the following phases:

- Phase I: *Scoping*. Identification of data attributes and data quality dimensions deemed business critical.
- Phase II: *Interviews*. Identification of business operations problems caused by defects in the attributes focused (interviews with representatives from various actors of Beiersdorf's ecosystem).
- Phase III: *Analysis*. Consolidation of interview results and identification of critical data defects.
- Phase IV: *Specification*. Specification of data quality metrics for monitoring critical data defects.

4.4.2.2 Identification of critical data defects

As the product data model of Beiersdorf's PLM system comprises over 800 attributes, in Phase I the project team grouped these attributes in data clusters and made a pre-selection of 20 clusters (comprising about 100 attributes) to be examined in the interviews. Then four data quality dimensions were selected, namely *Accuracy* (i.e. data values correct?), *Completeness* (i.e. data values existent?), *Consistency* (i.e. same data values for an attribute processed by different systems?), and *Timeliness* (i.e. data values always timely available?). The team selected data clusters (e.g. bill of material, GTIN) and data defect types (e.g. ambiguous, incomplete, out-dated, wrong value) that were mentioned in complaints about product data quality. For example, GTIN inconsistencies could occur due to data changes in regional ERP systems (see Figure B4-3). The dimensions were only used to structure the interviews, i.e. the team prepared several questions for each dimension that have been discussed for each selected attribute (e.g. "Are there any wrong or inaccurate data?", "Are there any out-dated data?", "Are there any inconsistent data regarding data in the PLM system?").

In Phase II, interview partners were selected and interviews were conducted (see Table B4-2). Thereby, important parts of the supply chain as well as the complexity of the product portfolio (by taking into account products for the mass market, complex products, and both self-produced and externally produced products) could be covered well by seven interviews. Moreover, all interviewees have been with Beiersdorf for years (some of them for more than 20 years) and are well versed in the company's data, processes and systems. Interview participants were invited by the head of Supply Chain Technology (see Figure B4-1) in order to highlight the interviews' significance. All interviewees supported the collaborative data defect identification by providing and discussing issues from business processes they are involved in. Complaints regarding other actors of the ecosystem could not be observed. Even for issues where required data is not provided timely in some cases (see [10] in Figure B4-2), interviewees of the impaired party (production plant in the example) strived for the identification of mutual process improvements.

The interviews were based on an interview guide containing the grouped attributes for each cluster plus questions for each data quality dimension. Other questions referred to the interviewees' assessment of the current data quality in general, their confidence in the correctness of data, and other issues or problems not covered by the interview guide (e.g. further attributes). Each business operations problem identified was described, the data attribute and the defect type causing the problem were documented, and the frequency and business impact (i.e. costs, service level impairment, risk of complaints) of the problems were assessed using a scale ranging from 1 to 5. Given examples on how to interpret the respective rating scales ensured an approximately equidistant interpretation of the scale levels and a comparable use of the scales in different interviews. Table B4-3 summarizes the assessment results of all interviews and

shows data attributes and corresponding issues that were mentioned during the interviews. For reasons of confidentiality Table B4-3 does not distinguish between particular interviews and issues and only provides means of frequency and impact ratings.

Data attribute	Data cluster	Issues	Frequency	Impact
Gross weight	Logistics data	8	2.47	0.99
Country artworks	Artworks, technical drawings	7	2.60	0.53
Dangerous goods indicator	Dangerous goods indicator	5	2.33	0.97
GTIN (packaging unit)	GTIN	5	1.40	1.80
Measurement	Logistics data	5	2.42	0.55
Status	Bill of material	5	3.25	0.44
Temperature conditions	Temperature conditions	5	1.75	2.25
GTIN (piece level)	GTIN	4	1.25	1.00
GTIN (shrink level)	GTIN	3	2.00	1.55
Packing data	Bill of material	3	4.67	0.67
Product formula	Bill of material	3	2.67	0.33
Technical drawings	Artworks, technical drawings	3	4.50	0.67
Format of product marking	Format of product marking	2	4.50	1.67
Gross weight (pallet level)	Logistics data	2	3.00	0.33
Material cluster	Material data (basic data)	2	1.00	0.84
Material description	Material data (basic data)	2	2.00	0.50
Material number	Material data (basic data)	2	3.00	0.33
Stacking plan	Logistics data	2	1.50	2.33
Maximum storage period	Storage Data	1	3.00	0.55
Sub process cluster	Material data (basic data)	1	1.00	0.67

Table B4-3: Assessment results of interviews

In Phase III the interview findings were consolidated in order to be able to identify those data clusters that are critical from the perspective of the entire ecosystem (see Table B4-3). As a result, the following seven issues could be identified.

- *Missing temperature conditions for transportation.* There were cases where, due to extreme outdoor temperatures, products had been delivered frozen, leading to customer complaints and costs for return, quality inspection, and new delivery. The configuration of the PLM system was lacking product specific information on ambient temperatures recommended for transport.
- *Wrong format of product marking (expiry date).* Like many other companies, Beiersdorf has had in some cases problems using the correct format for indicating the expiry date on product packages. These formats differ from country to country. As there is no complete central documentation containing all valid formats for being looked up, provisions on the correct format must be often researched.

- *Missing dangerous goods indicator.* Beiersdorf has always made sure that dangerous goods (e.g. deodorant sprays) are marked and labeled accordingly, and so missing or wrong dangerous goods indicators have not been reported so far. However, products for special marketing campaigns (e.g. displays offering a bag with a shampoo and a deodorant spray) mostly are not packed in the plants where their components are produced but in the distribution centers. The need for dangerous goods indication then has to be identified by researching through the combined products' bills of materials. In several interviews this manual process was described as laborious and bearing risks, even if no concrete cases of failure were reported.
- *Missing or wrong GTIN.* There were cases where GTINs for logistic units (e.g. shrink, packaging unit, pallet) were missing, wrong, not unique (for example, product GTINs used also for shrinks), or not consistent with the data transmitted to the GS1 data pool, potentially leading to problems in product distribution and lowering the service level.
- *Inaccurate or not timely available logistics data.* Calculated product weight (for example, volume multiplied by density) has not always been replaced in the PLM system by actual values determined after production according to final adaptations. In case of products with a transparent package (for example, shampoo filled in a clear bottle) sometimes a little bit more content than indicated on the label is filled in up to the bottles cap (otherwise the bottle looks somewhat empty). Although such changes in content usually is within the tolerance range of 20 per cent as demanded by GS1, logistics processes might be affected by pallet weights being too high. Besides, information on product dimensions requested by wholesalers and retailers has not always been available in time.
- *Bill of material not timely released.* There were cases where bills of material had not been available in time or had been changed several times after the actual release date, leading to delays in the production process, increased effort in production plants, and potentially lower service level.
- *Documents not timely available.* There were cases where artworks and technical drawings had not been available in time, leading also to delays in the production process, increased effort in production plants, and potentially lower service level.

The identification of these seven problems constituted the starting point of the specification of data quality metrics, with two modifications being made. First, the problem of missing temperature conditions for transportation was not considered to be a data quality issue, as there was simply not enough information available for defining suitable thresholds. Second, in the course of an interim presentation, customs relevant data

could be identified as another critical cluster. The data quality metrics were then extended by adding validation rules on timeliness for the attributes *commodity code* and *country of origin*. Table B4-4 lists the data defects (i.e. affected data attribute(s) and data quality dimension) deemed critical and assigns them to the data flows within Beiersdorf's ecosystem as shown in Figure B4-. The dimension *Change Frequency* was added in the analysis phase, as several issues referred to attributes (mostly entries and status values of bills of material) that were modified too often after the data's actual release.

Data flow	Data cluster / attribute	Data quality dimension	Business impact
[2]	Bill of material	Timeliness, Change Frequency	<ul style="list-style-type: none"> • Additional costs due to new production, if modifications of the bill of material make existing products unusable. • Service level impaired due to delays in production.
	Format of product marking	Accuracy, Completeness, Change Frequency	<ul style="list-style-type: none"> • Additional costs due to extra work (e.g. gathering information, communication). • Additional costs due to rework, if defect is not identified before start of production. • Risk of being fined by regulatory authorities if defect is not detected before shipment of goods.
[3]	GTIN	Consistency	<ul style="list-style-type: none"> • Service level impaired if inconsistencies are not detected before shipment of goods. • Risk of being fined by regulatory authorities if defect is not detected before shipment of goods.
	Logistics data	Accuracy	<ul style="list-style-type: none"> • Service level impaired if inconsistencies are not detected until shipment of goods. • Risk of being fined by regulatory authorities if defect is not detected before shipment of goods.
[4]	GTIN	Accuracy, Consistency	<ul style="list-style-type: none"> • Additional costs due to need for new artwork if defect is not detected before production or distribution.
[6]	Dangerous goods indicator	Completeness	<ul style="list-style-type: none"> • Risk of being fined by regulatory authorities if defect is not detected before shipment of goods.
	Logistics data	Timeliness	<ul style="list-style-type: none"> • Costs due to lost revenues.
[7]	Bill of material	Timeliness, Change Frequency	<ul style="list-style-type: none"> • Additional costs due to new production or rework, if modifications of the bill of material make existing products unusable. • Service level impaired due to delays in production.
[8]	Dangerous goods indicator	Completeness	<ul style="list-style-type: none"> • Additional costs due to extra work (e.g. gathering information, communication).
	GTIN	Accuracy, Completeness, Change Frequency	<ul style="list-style-type: none"> • Additional costs due to extra work (e.g. gathering information, communication).

Data flow	Data cluster / attribute	Data quality dimension	Business impact
		cy	<ul style="list-style-type: none"> Risk of being fined by regulatory authorities if defect is not detected before shipment of goods.
	Logistics data	Accuracy, Completeness, Change Frequency	<ul style="list-style-type: none"> Additional costs due to extra work (e.g. gathering information, communication), particularly regarding GTINs for logistic units.
[10]	Artworks, technical drawings	Timeliness	<ul style="list-style-type: none"> Service level impaired due to delays in production.
[11]	Dangerous goods indicator	Completeness	<ul style="list-style-type: none"> Risk of being fined by regulatory authorities if defect is not detected by logistics service provider.
[12]	GTIN	Accuracy, Consistency, Completeness	<ul style="list-style-type: none"> Additional costs due to extra work (e.g. gathering information, communication), particularly regarding GTINs for logistic units.
	Logistics data	Accuracy, Change Frequency	<ul style="list-style-type: none"> Additional costs due to extra work (e.g. gathering information, communication), particularly regarding gross weight.
[13]	Dangerous goods indicator	Completeness	<ul style="list-style-type: none"> Risk of being fined by regulatory authorities if defect is not detected by logistics service provider.
	Logistics data	Accuracy	<ul style="list-style-type: none"> Additional costs due to repackaging and restocking if tolerance range is exceeded.
[14]	Customs relevant data	Timeliness	<ul style="list-style-type: none"> Additional costs due to extra work (e.g. gathering information, communication). Service level impaired due to delays in production.
[15]	Artworks, technical drawings	Timeliness	<ul style="list-style-type: none"> Service level impaired due to delays in production.

Table B4-4: Critical data defects and business impact

4.4.2.3 Specification of data quality metrics

The aim of Phase IV was to specify data quality metrics allowing to monitor the data defects identified in the course of the interviews. Measured values of data quality metrics were normalized to the interval $[0;1]$ supposed to be interpreted as follows:

- Best value (1): No data object validated contains a critical defect.
- Worst value (0): Each data object validated contains at least one critical defect.

The data clusters identified to be the cause of critical business problems constitute the structure of the metric system (i.e. one metric for each cluster). In order to be able to aggregate and compare measured values of metrics at a later stage, a calculation formula was determined that is applicable to all metrics (see below). For the same reason, a uniform structure for all validation rules was specified. A validation rule r checks if a data object d (representing the set of data attributes representing one product) shows

all the characteristics defined by the rule (output value: 0) or if at least one criterion is not fulfilled (output value: 1).

$$r(d) = \begin{cases} 1, & \text{if the checked data object } d \text{ does not meet all criteria defined by } r \\ 0, & \text{if the checked data object } d \text{ meets all criteria defined by } r \end{cases}$$

The value of a data quality metric m is calculated by applying all $|R|_m$ validation rules specified for m (R : set of all rules defined for a data class D) to those $|D|_{r_i}$ data objects $d_j \in D$ which are defined to be checked by a specific rule r_i . For each validation rule r_i the weighting factor ω_i determines the rule's impact on the measured value of the metric m .

$$m(D) = 1 - \frac{\sum_{i=1}^{|R|_m} \left(\frac{\sum_{j=1}^{|D|_{r_i}} \omega_i r_i(d_j)}{|D|_{r_i}} \right)}{\sum_{i=1}^{|R|_m} \omega_i}, \text{ with } \omega_i \geq 0$$

The formula's complexity results from the intention to be able to specify for each rule r_i as precisely as possible the subset $|D|_{r_i}$ of D containing those $|D|_{r_i}$ data objects in which defects can occur at all with respect to the data defect represented by r_i . Table F-1 (see Appendix F) shows the seven specified product data quality metrics and 31 validation rules. The rules are implemented in a reporting tool that periodically analyses all product data objects and calculates the metric values. The metrics are to be evaluated monthly.

4.4.3 Findings and discussion

Few complaints about product data quality in general and the result of conducted interviews indicate that Beiersdorf is dealing with a relatively small number of business critical product data defects which certainly is due to the good organization and technical support of the company's master data management. Inconsistent intra-company product data, duplicate data, or wrong positions in bills of material can largely be avoided by measures such as centralized assignment of unique identifiers, integration of product data maintenance processes already in the product innovation process (see MDW in Figure B4-3), or clear specification of roles and responsibilities for product data maintenance. Another indicator for high maturity of Beiersdorf's master data management organization is the commitment of all interviewees regarding the data defect identification and metric specification process. All interviewees understood and supported the project's objectives, provided valuable input and offered additional support for answering further questions and for reviewing results.

Nevertheless, critical data defects do occur at Beiersdorf, particularly with regard to data exchanged with other businesses. Although the pre-selection of data clusters by the Data Process Management team (cf. Phase I of the project) led to a focus of product data that is used by multiple partners of the ecosystem, the interviewees did not raise any locally used attributes. Only customs relevant data was added which is used by multiple partners as well. Fortunately, many of the defects (e.g. missing dangerous-goods indicator, wrong indication of pallet weights) are detected in the course of manual inspections done by experienced personnel. Well-defined cross-departmental or cross-divisional data management processes (e.g. maintenance of GTINs, design of artworks) have not been sufficiently established yet. Continuous monitoring of the data defects identified with the help of the data quality metrics specified will show if planned measures (e.g. new workflows, revised GTIN allocation process, redesigned process for artwork production) are able to sustainably improve product data quality.

To realize the specified data quality metrics system, Beiersdorf has been conducting a project for implementing specified metrics that comprises four phases: In a first phase, validation rules for measuring logistics data quality (see Table B4-4, VR24 – VR30) and a reporting system have been implemented. The second phase comprises validation rules for measuring GTIN's and bill of material's change frequency, consistency and timeliness (VR01 – VR03, VR20 – VR 23, VR28 – VR30). In contrast to the first phase, validation rules in Phase 2 require access to time related metadata (i.e. change protocols) and to further systems (e.g. VR19). In Phase 3, validation rules for monitoring GS1 compliance (e.g. VR VR09 – VR18) will be implemented. And Phase 4 will provide remaining validation rules such as VR05 – VR07 for monitoring timely availability of documents and customs data (VR04), for monitoring consistent dangerous goods indicators (VR08) and for product marking format (VR31, VR32).

4.5 Implications for practice and research

The case study at Beiersdorf describes the process of identifying business critical product data defects and proposes a set of seven data quality metrics that allow to monitor such defects. To measure these metrics, 32 validation rules are specified using only commonly used (i.e. not specific to Beiersdorf) product data attributes. They may be used also by other companies as elements of their own metrics systems. Thus, an obvious implication for practice is to use and evaluate the specified metrics as elements of their own data quality metrics systems and to implement the validation rules within their measuring and reporting tools.

As is the case with all case studies, whether the results (here: product data defects, data quality metrics) can be generalized needs to be found out by conducting more case studies with similar companies and by applying proven theories to get a better understanding of the interrelations identified in this case study. An important implication for

research therefore is to motivate scientists to evaluate and adapt the metrics and validation rules specified for Beiersdorf, and to investigate organizational (e.g. degree of collaboration) and technical (e.g. complexity of data models) factors impacting the quality of shared product data.

Another implication for research is due to potential dependencies between the impact of data defects and the time of their detection regarding the data lifecycle. Wrong logistics data (e.g. a wrong pallet weight) may cause only little additional costs (for reweighing and repackaging) when the defect is detected at a distribution center. When goods arrive at the customer's site, however, this data defect may bring about high additional costs (for return of the goods and lowered service level). This means that not just the type of data defect but also the time a data defect is detected seem to have an effect on its business impact. Modeling and monitoring a data lifecycle perspective of a company's ecosystem comprising all create, read, update, deactivate and archive activities would facilitate model-based analysis of data flows and data maintenance processes and help identify weak points in data maintenance processes (e.g. missing responsibilities, cycles, delays).

5 A cybernetic view on data quality management

Titel	A Cybernetic View on Data Quality Management
Autoren	Boris Otto; Kai M. Hüner; Hubert Österle Universität St. Gallen, Institut für Wirtschaftsinformatik Müller-Friedberg-Strasse 8, 9000 St. Gallen, Schweiz {boris.otto kai.huener hubert.oesterle}@unisg.ch
Publikationsorgan	Leidner, D. E., Elam, J. J. (Hrsg.): Proceedings of the 16th Americas Conference on Information Systems, Lima, Peru, 2010
Status	Veröffentlicht

Tabelle B5-1: Bibliographische Angaben zum Beitrag „A cybernetic view on data quality management“

Abstract

Corporate data of poor quality can have a negative impact on the performance of business processes and thereby the success of companies. In order to be able to work with data of good quality, data quality requirements must clearly be defined. In doing so, one has to take into account that both the provision of high-quality data and the damage caused by low-quality data brings about considerable costs. As each company's database is a dynamic system, the paper proposes a cybernetic view on data quality management (DQM). First, the principles of a closed-loop control system are transferred to the field of DQM. After that a meta-model is developed that accounts for the central relations between data quality, business process performance, and related costs. The meta-model then constitutes the basis of a simulation technique which aims at the explication of assumptions (e.g. on the effect of improving a data architecture) and the support of DQM decision processes.

5.1 Introduction

High-quality data is a prerequisite for companies to meet the requirements posed by numerous strategic business drivers, such as global process harmonization, integrated customer management, global spend analysis, or compliance with legal provisions and laws. Companies use data when performing their everyday business operations, i.e. to produce goods, to render services, or to do research and development. The quality of the data used affects business processes (e.g. with regard to lead time) as well as the outcome of business processes (e.g. customer satisfaction, product quality). In small-batch production, for example, fast availability of construction drawings at different steps of the manufacturing process as well as completeness and correctness of parts lists are crucial prerequisites for short process lead times. And in the services industry (in a call center, for example) fast availability of all data about each customer (e.g. previous orders and contracts, complaints) is mandatory for short processing times and high customer satisfaction. In order to be able to work with data of good quality, data

quality requirements must clearly be defined. In doing so, one has to take into account that both the provision of high-quality data and the damage caused by low-quality data brings about considerable costs.

However, studies about the relation between the effort required to achieve a certain degree of data quality and the costs caused by this effort are missing. Lee et al. [Lee et al. 2006] do propose the use of real options [Black/Scholes 1973] for cost-benefit analysis in the field of data quality management (DQM) as well as simulation as a technique to determine the value of such options [Amram/Kulatilaka 1998]. However, these reflections have not been further specified so far.

As each company's database is a dynamic system, the paper proposes a cybernetic view on DQM. On the basis of a meta-model a simulation technique is developed, which does not intend to allow for automatic control of DQM initiatives or forecasting of the development of data quality over time, but which aims at the explication of assumptions (e.g. expected effect of a data clearing measure) and the support of decision processes in corporate DQM. Furthermore, the simulation technique supports the specification process of business oriented data quality metrics to identify data defects that are considered to have a mission critical impact [Otto et al. 2009a]. When searching for causal relations (i.e. a specific data defect causes a specific business problem which jeopardizes a specific strategic company goal) underlying such metrics, scenarios may be created and simulated to verify the plausibility of assumptions made.

The paper starts with a short presentation of related work. Then the underlying research methodology is presented. After that the paper introduces a meta-model as a basis for a DQM simulation technique. To facilitate scenario modeling the paper specifies various impact patterns representing relations between various elements of the model plus a simulation model. Both the meta-model and the simulation technique are then evaluated in a business scenario at an international telecommunications company. Finally the paper concludes with a summary and an outlook on future research to be done.

5.2 State of the art

5.2.1 Measuring data quality

Data quality has been a research topic in numerous empirical [Wang/Strong 1996], pragmatic [Redman 1996, English 1999] and theoretical [Wand/Wang 1996, Price/Shanks 2005] studies, the outcome of which mainly were lists and categories of data quality dimensions. Many of these studies have in common that data quality is seen as something that is mainly determined by the data's *'fitness for use'*, i.e. whether data is of good or poor quality depends on the context it is used in and the user it is used by. While the quality of a certain set of customer data, for example, can be sufficient to conduct a product marketing initiative (having available correct and complete

e-mail addresses), it may not be good enough for duplicate free consolidation with another set of customer data (as an unambiguous key is missing and address data are inconsistent). Apart from the structural conceptualization of data quality by means of various data quality dimensions, the measurement of data quality (i.e. of data quality dimensions) has been a central issue of many scientific studies [Naumann/Rolker 2000, Burgess et al. 2004, Gebauer et al. 2005, Gustavsson 2006, Heinrich et al. 2009]. In this respect, the impact of data defects on the operations of companies needs to be taken into account when data quality metrics are being designed [Batini et al. 2007, Caballero et al. 2008, Otto et al. 2009a].

5.2.2 Costs and benefits of data quality (management)

A number of studies has dealt with the impact data defects have on business operations [Joshi/Rai 2000, Fisher/Kingma 2001, Eppler/Helfert 2004], introducing various classifications of costs caused by DQM. In this respect, it is important to note that both the provision of high-quality data and the damage caused by low-quality data brings about considerable costs. For example, a clockwork manufacturer's production process involves both automated and manual sub-processes, the latter comprising activities such as assembly of special pointers for certain product series and brands. If for such a manual process an assembly instruction drawing is not available for some reason (from an information systems perspective: the linking of an electronic document with the parts list of a production order), process lead time gets longer while process performance per time unit (e.g. the number of clockworks produced per hour) gets lower. Also, outdated or false construction drawings can lead to production of scrap parts, leading also to decreased process performance. Potential measures to ensure good availability of data relevant in the production process may include regular verification of assembly and construction drawings or training of constructing engineers to raise awareness about the effects of missing or false drawings.

Obviously, data defects (here: a missing link) can cause business problems (here: missing construction drawings), which in turn lead to increased costs (here: production of less clockworks per time unit). DQM initiatives could help prevent or counteract data defects (here: through manual linking of construction drawings prior to the production process or through implementation of validation systems in preceding processes). So while DQM initiatives can reduce overall costs, they also cause costs for their development and implementation. Therefore, the objective is to determine the minimum of cumulated costs taking into account costs caused by data defects on the one hand and costs of DQM initiatives on the other hand [Eppler/Helfert 2004]. Figure B5-1 shows a cost-benefit perspective on DQM.

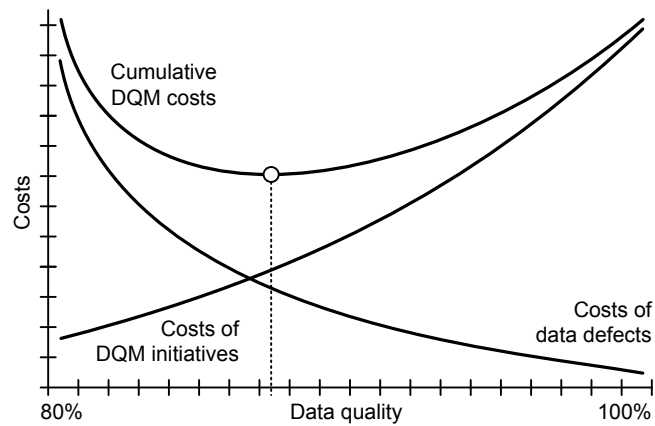


Figure B5-1: Cost-benefit-perspective on DQM

5.3 Research methodology

The meta-model together with the DQM simulation technique is an outcome of design oriented research, following Design Science Research as a methodological paradigm. Design Science Research aims at the design of artifacts (i.e. constructs, models, methods and instantiations) supposed to solve practical problems [Hevner et al. 2004]. In this context, constructs – similar to a language – define concepts for describing a specific problem, while models use constructs for describing concrete problems and possible solutions. The meta-model presented in this paper is a Design Science Research construct defining concepts for the description of simulation models which in turn describe concrete DQM scenarios by formulating assumptions regarding the impact of data defects and DQM initiatives as probabilistic variables.

Taking into account various requirements regarding the design of artifacts [Hevner et al. 2004], the design process for each artifact consists of several phases, namely problem identification, definition of design objectives, artifact design, artifact evaluation, and – if necessary – artifact redesign [Pefferers et al. 2008]. The research context of the design process of the method is constituted by the Competence Center Corporate Data Quality. The Competence Center is part of the Business Engineering research program at the University of St. Gallen. Since 2006 it has been developing – together with renowned partner companies – solutions for support of quality management of corporate data.

5.4 Simulation approach

5.4.1 Data Quality Management as a closed-loop control system

Each company's database can be considered as an open system characterized by the quality of the data changing over time [Orr 1998]. Data quality is affected by both internal factors (e.g. no standard procedures in the use of data, no validation procedures concerning the identification of duplicates in the data collection process) and external

factors (e.g. customer address data affected by customers who have moved to another place). As data quality shows a dynamic behavior over time and also because there has to be a balancing between the costs for DQM on the one hand and the costs expected to occur from poor data quality on the other hand, a systems theory approach can be applied. Figure B5-2 shows the interrelations described from a systems theory perspective [Lunze 2008].

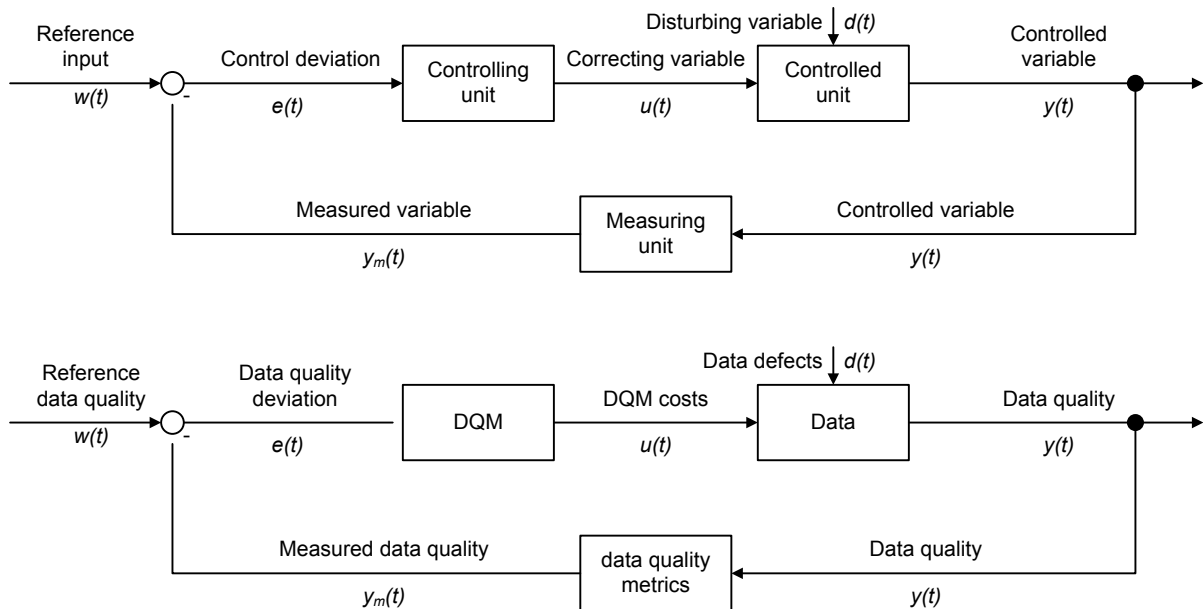


Figure B5-2: Block diagram of generic closed-loop control system (above) and adaptation of system elements for DQM (below)

If DQM is considered that way (i.e. as a closed-loop control system, see Table B5-1 for further explanation), the organizational unit responsible for DQM in a company must decide what measures for improving data quality need to be applied. To do so, expected costs of DQM initiatives planned, expected effects of these measures, expected risks of the occurrence of data defects, and expected costs caused by data defects need to be estimated. To facilitate such estimations, the paper presents a meta-model and a simulation technique, with the simulation technique being capable of calculating the data quality to be expected based on the assumptions specified in the meta-model.

Element	Generic closed-loop control system	Adaptation for DQM
Controlled unit	Dynamic system determined by the correcting variable (which can be manipulated), interfered by the disturbing variable (manipulated), and producing the controlled variable (measured).	Corporate data. The aim is to keep the data quality at a certain level.
Disturbing variable d	Parameter interfering with the controlled unit. Cannot be manipulated.	Data defects (see below).
Controlled variable y	Parameter measuring the output of the Controlled unit. Its value is dependent on the correcting variable and the disturbing Variable.	Data quality.
Measuring	(Dynamic) System measuring the controlled variable or	Data quality metrics

Element	Generic closed-loop control system	Adaptation for DQM
unit	– in case the controlled variable cannot be measured – calculating the controlled variable from other measured variables.	(see below).
Measured variable y_m	Parameter representing the measured output of the controlled unit. Due to potential dynamics of the measuring unit itself, controlled variable and measured variable have to be distinguished.	Measured data quality. Values calculated from data quality metrics.
Reference input variable w	Parameter for controlling the control unit. The closed-loop control system tries to adjust the controlled variable according to this parameter and to compensate the interference of the disturbing variable. The objective is $y(t) = w(t)$ at any time.	Data quality target value.
Control deviation e	Difference between reference input variable and controlled variable / measured variable.	Difference between data quality target value and actual value of data quality.
Controlling unit	System controlling the controlled variable (by means of the reference input variable) so that the Controlled unit works as desired.	DQM comprising preventive and reactive initiatives.
Correcting variable u	Parameter by which the controlling unit tries to compensate the effect of the disturbing variable and match the controlled variable with the reference input variable.	Costs of (preventive and/or reactive) DQM initiatives.

Table B5-1: Elements of generic closed-loop control system adapted for DQM

5.4.2 Simulation objectives

The overall aim of the DQM simulation model proposed is to explicate assumptions regarding various risks and effects (e.g. occurrence of a certain data defect or impact of a certain DQM measure) and support identification of ‘realistic’ causal relations for specific scenarios (e.g. false negative responses of a product availability check service due to data inconsistencies). A simulation model comprises those assumptions (e.g. ‘When entering customer address data into the CRM system, the probability of entering the wrong street name is 60 percent.’). An entire DQM simulation process consists of one or several simulation runs, all of which use the same simulation model (i.e. the same assumptions). A simulation run, in turn, consists of several simulation steps. In each simulation step the value of a certain variable of the simulation model is calculated (e.g. the value of a data quality metric). What is actually generated here are values of random variables, with the distribution of those random variables representing the assumptions made. Figure B5-3 shows the meta-model of the DQM simulation technique presented, i.e. the elements a simulation model may consist of. Denominators and semantics of the meta-model are based on industry standards [IEEE 1998, ISO/IEC 2007], and previous contributions to DQM research [Eppler/Helfert 2004, Caballero et al. 2007].

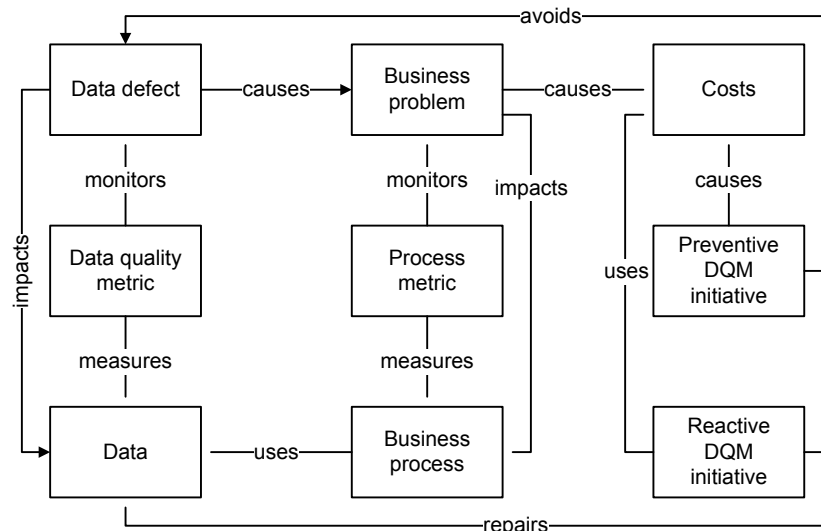


Figure B5-3: Block diagram of generic closed-loop control system (above) and adaptation of system elements for DQM (below)

5.4.2.1 Design objects for DQM simulation models

- **Business problem.** State (e.g. delivery of goods not possible) or incident (e.g. production of scrap parts) leading to decreased process performance and hence to poorer values of process metrics. A business problem poses a risk (in terms of probability of occurrence and intensity of impact, both of which are represented by a random variable) to a business process.
- **Business process.** Sequence of chronologically and typologically linked activities intended to generate a clearly defined output bringing about customer benefit. Transforms a given input (e.g. a certain material) into a desired output (e.g. a watch movement) under consideration of certain rules and by using certain resources (e.g. data). Is controlled and designed by means of process metrics defined as part of an overall business strategy.
- **Costs.** Measure to value the use of resources (e.g. money or material and immaterial means) for a cost objects (e.g. products, projects, human performance). The paper considers costs for DQM initiatives and costs caused by business problems.
- **Data.** Representations of objects (e.g. customers) and relations between objects (e.g. Customer A orders Article X) based on the description of certain object characteristics. A data element (or attribute) is a component of a data model which cannot be further subdivided in logical terms in a given or presupposed context. The paper considers company master data (e.g. customer address data, material data, or parts lists), with a focus not on data models (data classes and data elements) but on values assigned to data elements.

- *Data defect*. Incident leading to poorer values of data quality metrics. A data defect poses a risk (in terms of probability of occurrence and intensity of impact, both of which are represented by a random variable) to data.
- *Data quality metric*. Quantitative measure (cf. Metric, as well denoted by DQ measure [Caballero et al. 2007, ISO/IEC 2007] of the degree to which data possess a given quality attribute (e.g. completeness, timeliness). For a data quality metric, one or more measuring methods need to be provided (i.e. where the measurement is made, what data are included, what measuring device is used, and what scale is used for measuring).
- *Preventive DQM initiative*. Initiative reducing either the probability of occurrence of data defects or the impact of data defects.
- *Process metric*. Quantitative measure (cf. Metric, as well denoted by operational measures) of the degree to which a business process possesses a given quality attribute (e.g. lead time or scrap rate). Results directly from process performance, e.g. as the average lead time for processing of an order, or as the number of order cancelations per month.
- *Reactive DQM initiative*. Initiative for correction of defective data (e.g. through identification and correction of false address data).

5.4.2.2 Relations between design objects of DQM simulation models

- *affects (Business Problem to Business Process)*. A business process can be affected by a business problem. Besides data defects there are lots of other possible causes of business problems.
- *affects (Data Defect to Data Quality Metric)*. A data defect (e.g. missing values for attributes of customer address data) can have a negative impact on the value of a data quality metric (e.g. completeness of customer address data).
- *affects (Data Defect to Data)*. Data can be affected by data defects (e.g. missing or wrong values for customer address attributes).
- *affects (Business Problem to Process Metric)*. A business problem (e.g. delivery of goods to the wrong address) can have a negative impact on the value of a process metric (e.g. provisioning time).
- *causes (Business Problem to Costs)*. A business problem (e.g. delay in process lead time due to delivery of goods to the wrong address) can bring about costs (e.g. lower revenues due to cancellation of orders).
- *causes (Data Defect to Business Problem)*. A data defect (e.g. wrong customer address) can bring about a business problem (e.g. delivery of goods to the wrong address).

- *measures (Data Quality Metric to Data)*. A data quality metric measures the degree to which data possess a given quality attribute (cf. Data Quality Metric).
- *measures (Process Metric to Business Process)*. A process metric measures the degree to which a business process possesses a given quality attribute (cf. Process Metric).
- *prevents (Preventive DQM Initiative to Data Defect)*. A preventive DQM initiative (e.g. implementation of a company wide data architecture) can help prevent a data defect (e.g. inconsistent data transfer).
- *repairs (Reactive DQM Initiative to Data Defect)*. A reactive DQM initiative (e.g. identification and correction of wrong data) corrects a data defect, i.e. there is a certain probability (represented by a random variable) that the initiative positively affects the value of a data quality metric affected by the data defect.

5.4.3 Impact Patterns

5.4.3.1 *Patterns for the Impact of a Data Defect on a Data Quality Metric Measurement*

When a simulation model is created (i.e. when the meta-model presented in this paper is instantiated), it is not only necessary to define values for various variables (e.g. probability of occurrence of a certain data defect) but also to specify the impact for each cause-effect relation (e.g. ‘If data defect A occurs, it must exceed a certain threshold value X to cause business problem B.’). The following sections describe a number of impact patterns that are supposed to facilitate the creation of a simulation model.

A data defect has a certain probability of occurrence and a certain impact. In a simulation step, values are generated for both random variables, i.e. a value (,0’ or ,1’) is generated for probability of occurrence and a value is generated for impact. The scale for measuring the impact is the same scale that measures the data quality metric affected by the data defect. For example, missing values for address data attributes have an impact on a metric for completeness of address data, which is given as a percentage value. The data defect’s impact in this case would have to be indicated as a percentage value as well (e.g. as a reduction of the value of the metric by 2 percent, cf. Figure 3). Three patterns are distinguished here (the following description follows the principle: the higher the value of the data quality metric, the higher the quality of the data):

- *Binary Pattern*. When a data defect occurs in a simulation step, the data quality metric value is reduced by a fixed value, regardless of the data defect’s impact.

- *Threshold Pattern.* When a data defect occurs in a simulation step and when the data defect's impact exceeds a certain threshold value, the data quality metric value is reduced by a fixed value.
- *Linear Correlation Pattern.* When a data defect occurs in a simulation step, the reduction of the data quality metric correlates with the data defect's impact (i.e. the value for the data defect's impact is multiplied with a certain factor and the result is deducted from the value of the data quality metric).

5.4.3.2 Patterns for the Impact of a Business Problem on a Process Metric Measurement

The patterns specifying the impact of business problems on process metrics are the same patterns specifying the impact of data defects on data quality metrics (see above).

5.4.3.3 Patterns for the Impact of a Data Defect on a Business Problem

A data defect can increase the probability of occurrence and/or the impact of a business problem. Four patterns are distinguished here:

- *Binary Pattern.* When a data defect occurs in a simulation step, the probability of occurrence and/or the impact of a business problem is increased by a fixed value, regardless of the data defect's impact.
- *Threshold Pattern.* When a data defect occurs in a simulation step and when the data defect's impact exceeds a certain threshold value, the probability of occurrence and/or the impact of a business problem is increased by a fixed value.
- *Linear Correlation Pattern.* When a data defect occurs in a simulation step, the probability of occurrence and/or the impact of a business problem is increased by a fixed value correlates with the data defect's impact (i.e. the value for the data defect's impact is multiplied with a certain factor and the result is added to respective value).
- *Indirect Pattern.* The probability of occurrence and the impact of a business problem are calculated over the value of a data quality metric (cf. Linear Correlation Pattern). Thus they are only indirectly affected by data defects. If need by the indirect dependency can also be modeled considering a threshold (cf. Threshold Pattern).

5.4.3.4 Patterns for the Impact of a Business Problem on Costs

The patterns specifying the impact of business problems on costs are the same patterns specifying the impact of data defects on business problems (see above), with the costs' impacts simply being their monetary value.

5.4.3.5 Patterns for the Impact of a Preventive DQM Initiative

A preventive DQM initiative (e.g. implementation of a company wide data architecture) can help prevent a data defect by reducing its risk. The effect of a preventive DQM initiative (i.e. the risk reduction given by new parameters for the probability distribution) has a duration (i.e. the simulation procedure uses the lower risk for a given number of simulation steps). Two patterns are distinguished here:

- *Binary Pattern.* The initiative changes the risk parameters (i.e. the parameters of the random variables for probability of occurrence and impact) by a fixed value.
- *Initiative Booster.* The initiative reinforces the impact of another preventive or reactive DQM initiative (cf. Binary Pattern). An example would be the definition and operationalization of a DQM strategy, which indirectly contributes to risk reduction, e.g. by guaranteeing sustained financing for DQM for consistent maintenance of a data architecture.

5.4.4 Simulation procedure

```

dataQ      = initial data quality
processQ   = initial process quality
costs      = 0

FOR step = 0 TO duration
  FOR EACH initiative IN preventiveInitiatives
    IF initiative.IsActive()
      FOR EACH defect IN listOfDataDefects
        initiative.AdjustDataDefect(defect)
      END FOR
      IF initiative.GetStart() == step
        costs += initiative.GetCosts()
      END IF
    END IF
  END FOR
  FOR EACH defect IN listOfDataDefects
    DataQ -= defect.GetImpact()
  END FOR
  FOR EACH initiative IN reactiveInitiatives
    IF initiative.IsActive(step)
      dq += initiative.GetImpact(step)
      costs += initiative.GetCosts()
    END IF
  END FOR
  FOR EACH problem IN listOfBusinessProblems
    FOR EACH defect IN listOfDataDefects
      IF problem.IsCausedBy(defect)
        pq -= problem.GetImpact(defect)
      END IF
    END FOR
    costs += problem.GetCosts()
  END FOR
END FOR

```

Listing B5-1: Procedure of a DQM Simulation Run

Listing B5-1 describes the simulation procedure in a pseudo code notation in order to explain how the values for data quality, process quality and costs are calculated. The DQM simulation technique only considers scenarios involving one data quality metric (e.g. consistency of address data) and one process metric (e.g. provisioning time). Each simulation step calculates

- the impact of preventive DQM initiatives,
- the probability of occurrence and the impact of data defects,
- the impact of reactive DQM initiatives,
- the value of the data quality metric,
- the probability of occurrence and the impact of business problems,
- the value of the process metric,
- the cumulated costs caused by DQM initiatives and business problems.

5.5 Procedure Model for Application

The steps listed in the following constitute a procedure model that can be followed to create a simulation model for simulating scenarios.

- *Simulation Scale Specification.* To specify measuring units for the simulation steps. This value defines a simulation run's temporal interpretation. For example, in a simulation run comprising 40 steps the development of DQ may be simulated over a period of 40 days, 40 weeks, or 40 months.
- *Metric Specification.* To specify measuring units for the data quality metric (e.g. 'percent' for a metric measuring data consistency), the process metric (e.g. 'days' for a metric measuring provisioning time), and the respective initial values (e.g. '8' as average provisioning time without any interference by business problems). If a simulation refers to costs, a currency must be specified as well.
- *Data Defect Specification.* To specify data defects by means of two random variables, probability of occurrence and impact.
- *Business Problem Specification.* To specify business problems by means of two random variables, probability of occurrence and impact.
- *Metric Impact Specification.* To specify the impact of data defects on data quality metrics and the impact of business problems on process metrics, respectively (see impact patterns).
- *Problem Impact Specification.* To specify the impact of data defects on business problems and the impact of business problems on costs (see impact patterns). It must be decided for each business problem whether it is considered as independent from the data defects modeled or whether it is affected by the data defects modeled (and if so, how).

- *Initiative Specification.* To specify preventive and reactive DQM initiatives. For each initiative a period of time (starting point and duration, cf. function *IsActive* in Listing B5-1) and the costs expected must be indicated. The initiative's impact can be described as a random variable too, in order to simulate its probable impact.
- *Simulation.* To simulate a modeled scenario. The simulation procedure calculates the values to be simulated (e.g. metric values or costs). From the results of several simulation runs expected values (e.g. for the data quality metric) and statistical variances can be calculated.

5.6 Example of Application

5.6.1 Business Scenario

To evaluate the meta-model and the simulation technique a simulation model was developed for and applied at an international telecommunications company (employing approx. 250,000 people and having a turnover of 60 billion Euro in 2009). The scenario modeled refers to the process of customers ordering IP-TV. Part of this process is a validation service offered by the company for the customer to see whether IP-TV is available at their residence. The service's response is either positive confirming availability of the product (approx. 80 percent of responses) or negative (six different responses possible; e.g. 'Address unknown' or 'Resources for production of IP-TV used up').

Samples taken from this validation service have revealed that as much as five percent of the negative responses are false, i.e. IP-TV actually would be available at these residences. With an average of 5,000 requests per day concerning product availability and 20 percent of these requests resulting in signed contracts the company has to face lost revenues of approx. 9 million Euro per year, given an estimated value of each IP-TV customer of approx. 500 Euro per year.

As the reason for false negative responses inconsistencies between the data used for the product availability check and the data stored in the source systems (e.g. network infrastructure data, contractual data) could be identified. These inconsistencies were mapped onto rules for validation, which constitute the techniques for measuring six data quality metrics, with each data quality metric referring to the possible causes of one of the six false negative responses. The rules for validation are pairwise disjoint, i.e. a cause of a false negative response validated by one rule is not validated by another rule.

5.6.2 Simulation Model and Application

Based on the measuring of data quality metrics the risk of the occurrence of a specific data defect (i.e. the probability of a validation rule identifying a data defect) can be quantified. Table B5-2 shows the parameters of the pertinent, normally distributed random variable. Also, the table shows which data defect (DD) is used for measuring which data quality metric (indicated as errors, E). Figure B5-4 depicts the simulation (based on the parameters for Sim1 given in Table B5-2) of data defects over a period of 24 months, while Figure B5-5 depicts the simulation of false negative responses over the same period of time.

Data defect	DD1	DD2	DD3	DD4	DD5	DD6	DD7	DD8	DD9
Risk Sim1	953; 61	734; 50	1146; 127	1028; 37	428; 35	643; 71	495; 84	761; 129	369; 62
Risk Sim2	477; 31	367; 25	573; 64	514; 19	214; 18	322; 36	248; 42	381; 65	185; 31
Caused error	E1	E1	E2	E3	E3	E4	E5	E5	E6

Table B5-2: Parameters (i.e. mean, variance) of normal distributed risk of data defects (occurrence probability = 1 for all defects)

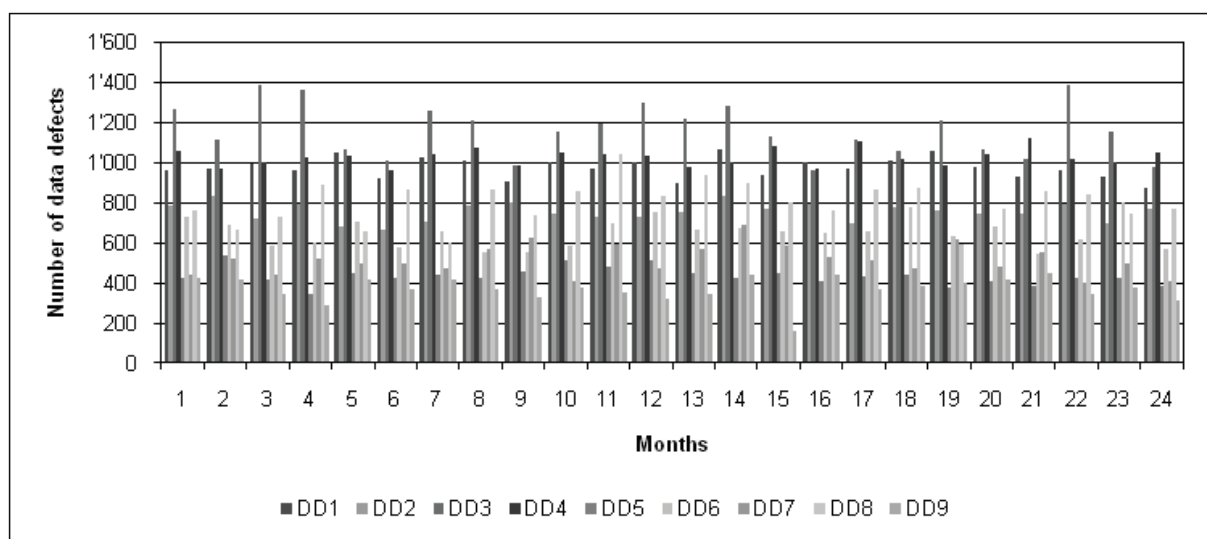


Figure B5-4: Simulation of data defects over a period of 24 months

Due to the already mentioned ratio of product availability requests and signed contracts the monthly value of a false negative response equals 20 percent (= 8.33 Euro) of the monthly value of a customer (= 41.67 Euro). Figure B5-6 depicts the cumulated costs (i.e. the total of lost revenues due to false negative responses of the product availability check) incurred over the entire period of time for Sim1. Furthermore, Figure B5-6 shows the result of a simulation run (Sim2) calculating the effect of a preventive DQM initiative (i.e. an improved data architecture). The architecture's design and implementation cost 500,000 Euro (projected on the first month of the simulation) but it also avoids 50 percent of the risk of data defects (see Table B5-2, second row). Therefore, the cumulated costs are lower for the preventive DQM scenario.

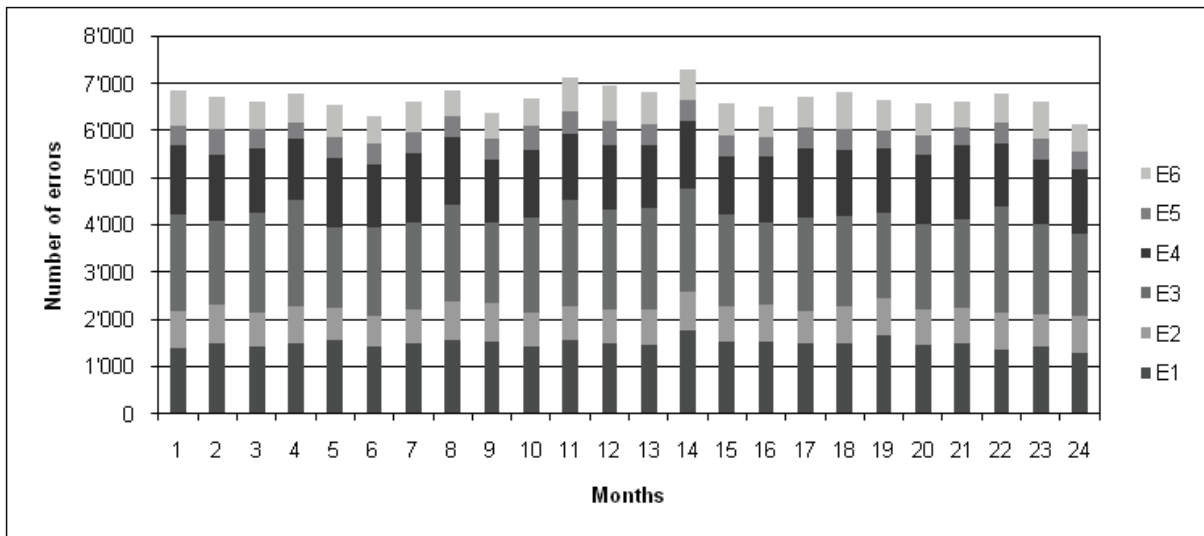


Figure B5-5: Simulation of false negative responses over a period of 24 months

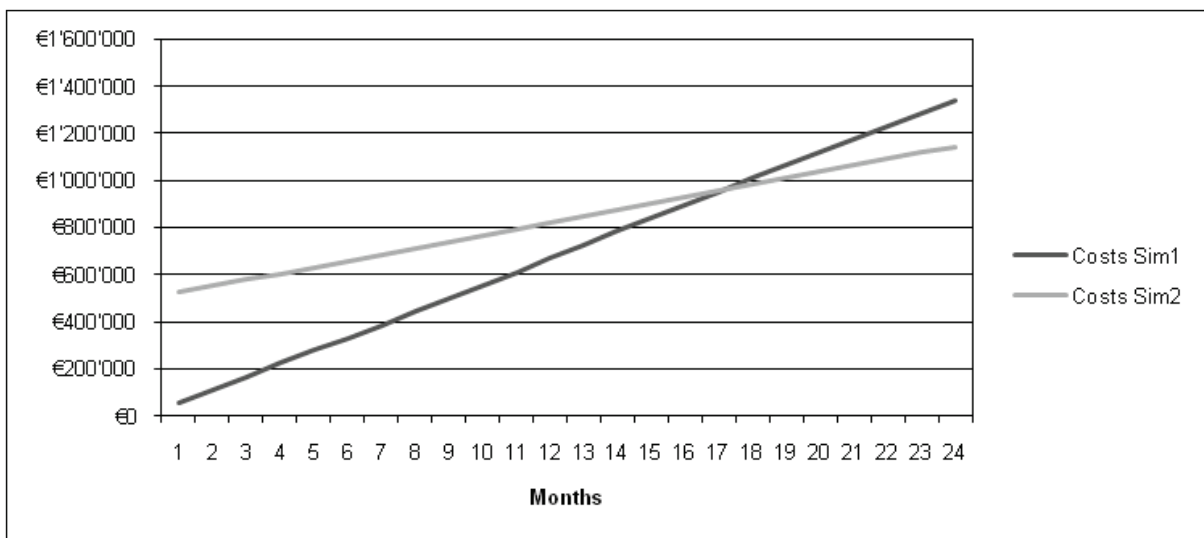


Figure B5-6: Cumulated costs due to false negative responses, on the right with data defect risk reduction by preventive DQM

5.7 Discussion and Future Work

The paper presents a cybernetic view on data quality management (DQM), with data quality being the controlled variable and costs of DQM initiatives being the correcting variable in a closed-loop control system. Also, the paper introduces a meta-model for description of DQM scenarios by means of a simulation model supposed to help an organizational unit responsible for DQM select appropriate measures. To facilitate scenario modeling the paper specifies various impact patterns representing relations between various elements of the model plus a simulation model. Both the meta-model and the simulation technique are then evaluated in a business scenario at an international telecommunications company.

When using the simulation model the main challenge is to identify and quantify causal relations. While this process can be delineated by a procedure model, in real-world

cases the information required (e.g. frequency of the occurrence of a specific data defect) has to be collected from various sources within a company (e.g. by means of interviews and expert assessments) and mapped onto the elements of the model (e.g. the parameters for distribution of a data defect). This requires both availability and willingness to cooperate on the part of subject matter experts and sufficient knowledge about the simulation model.

Need for further research is seen in further validating appropriate simulation models and in comparing the simulated development of data quality with real effects. Although a simulation model can take into consideration only selected aspects of reality, the aim is to describe real developments at least by approximation. To be able to do so, real measuring values of data quality metrics and process metrics must be available over a substantial period of time, and the effect of DQM initiatives on such measuring values needs to be traceable. Furthermore, research should focus on further specification of the meta-model, particularly with regard to specifying different types of costs and benefits as well as different cost bearers and beneficiaries. Through such further specification DQM simulations would not just be able to explicate the basic benefit of DQM initiatives, but it would be possible also to identify those company departments likely to expect the biggest benefit from a specific DQM measure, which could then bear a bigger proportion of the costs. Finally, further research refers to the use of the simulation model for a detailed determination of the behavior of DQM as a system. Simulation results and knowledge about companies' individual preferences would allow to select the controlling unit most appropriate (e.g. PID control) including suitable parameters.

6 Towards a maturity model for corporate data quality management

Titel	Towards a Maturity Model for Corporate Data Quality Management ²⁰
Autoren	Kai M. Hüner; Martin Ofner; Boris Otto Universität St. Gallen, Institut für Wirtschaftsinformatik Müller-Friedberg-Strasse 8, 9000 St. Gallen, Schweiz {kai.huener martin.ofner boris.otto}@unisg.ch
Publikationsorgan	Shin, D. (Hrsg.): Proceedings of the 2009 ACM Symposium on Applied Computing, Honolulu, Hawaii, USA, 2009, S. 231-238
Status	Veröffentlicht

Tabelle B6-1: Bibliographische Angaben zum Beitrag „Towards a maturity model for corporate data quality management“

Abstract

High-quality corporate data is a prerequisite for world-wide business process harmonization, global spend analysis, integrated service management, and compliance with regulatory and legal requirements. Corporate Data Quality Management (CDQM) describes the quality oriented organization and control of a company's key data assets such as material, customer, and vendor data. With regard to the aforementioned business drivers, companies demand an instrument to assess the progress and performance of their CDQM initiative. This paper proposes a reference model for CDQM maturity assessment. The model is intended to be used for supporting the build process of CDQM. A case study shows how the model has been successfully implemented in a real-world scenario.

6.1 Introduction

Today's large scale firms are facing market and regulatory driven challenges regarding the management of corporate data [Newman/Logan 2006]. High-quality corporate data is a key prerequisite for a number of strategic business objectives, like global supply chain management [Tellkamp et al. 2004], customer relationship management [Reid/Catterall 2005], decision-making and business intelligence [Shankaranarayanan et al. 2003], or compliance with regulatory and legal requirements [Friedman 2006]. When recognizing the need for comprehensive data management, companies ask for guidelines telling them how to achieve sustainable solutions. Efforts are not just about introducing software tools, but rather about designing a corporate data architecture, achieving consistent semantic data definitions [Shankaranarayanan et al. 2003, p. 113, Berson/Dubov 2007], and establishing CDQM as a corporate function [Wende 2007]. The key challenge is to transform a company's Data Quality Management (DQM)

²⁰ © Association for Computing Machinery (ACM) 2009

from reactive to proactive behavior [Wang/Strong 1996, p. 109, Wang et al. 1998, p. 99, Lee et al. 2006, p. 109].

A number of models have been developed for assessing the maturity of business processes (e.g. CMMI [SEI 2006b], BPMM model [Rosemann et al. 2004]) and business excellence of organizations (e.g. EFQM Excellence Model [EFQM 2003b]). Apart from these general maturity models, there are specific approaches for assessing data management maturity that are proposed by both practitioners [English 1999, pp. 427-437, Bitterer 2007, IBM 2007, DataFlux 2008] and researchers [Lee et al. 2002, Ryu et al. 2006]. Yet, while scientific approaches mostly focus on single aspects of data quality management (e.g. Ryu et al. [2006] focus on information systems), practitioners' approaches usually lack sufficient depth and objectivity (cf. Section 6.2.3). Despite the broad discussion of data quality in the scientific community [Wang/Strong 1996, Pipino et al. 2002, Olson 2003, Batini/Scannapieco 2006], a comprehensive and well-founded approach for assessing a company's maturity regarding DQM is missing.

The paper at hand proposes a reference model for maturity assessment in the context of Corporate Data Quality (CDQ, cf. Section 6.2.1). We refer to the maturity model as CDQ MM. The CDQ MM is intended to be used during the time CDQM is being established in an organization. The CDQ MM is developed as a design research artifact of type model (cf. Section 6.3.2 and Section 6.3.3). Usually, maturity models are used to assess process performance in regular operations (i.e. reducing process lead times, increasing process quality). However, the CDQ MM focuses on the implementation phase of a comprehensive CDQM project. In the process of CDQM establishment, CDQM components (e.g. well-defined data accountabilities, unique data definitions, or the DQM process itself) are being designed and implemented. In this respect, the CDQ MM is intended to be used as a management tool that supports a company's transformation process from reactive to proactive DQM.

Section 6.2 gives a definition of CDQ and CDQM and provides some background information on maturity models that are considered for the presented approach. Furthermore, we analyze the models in comparison to the given definitions of CDQM. Section 6.3 is about the research methodology our approach is based on. The CDQ MM itself is presented in Section 6.4, where we describe the model's requirements, construction, and evaluation. Section 6.5 summarizes our approach and gives an outlook on future work.

6.2 Background

6.2.1 Corporate data quality management

Despite of the important distinction between data and information [Boisot/Canals 2004], we do not focus on delimiting syntactical and semantic aspects of data. For the

paper at hand, we use the terms ‘data quality’ and ‘information quality’ synonymously and define data quality as ‘fitness for use’ [Wang/Strong 1996, p. 6, Redman 2001, p. 73f] where the ‘fitness’ depends on the data user’s perception. The term demands the ability of data to satisfy the requirements of intended use in a specific situation.

We are emphasizing the corporate scope of data quality (Corporate Data Quality, CDQ). The challenge of being able to ensure good data quality is particularly salient in decentralized organizations acting on a global level. Such companies possess a diversified portfolio of data storing and processing systems due to a history of mergers and acquisitions, deviant requirements of business units, and different regulations among countries. Data quality problems often occur when it comes to gathering information across business functions or organizational boundaries from several distributed systems [Batini/Scannapieco 2006, p. 69]. Therefore, our focus is on multi-business and multi-national organizations looking for a corporate approach to manage data quality.

A data quality activity is any process performed directly on data in order to improve data quality [English 1999, p. 285ff]. Besides these reactive activities, there are attempts towards data defect prevention [English 1999, p. 212, Eppler/Helfert 2004, Batini/Scannapieco 2006, p. 89]. We refer to the whole set of activities intended to improve data quality (both reactive and proactive) as Corporate Data Quality Management (CDQM). The concept of a framework for CDQM was proposed by Otto et al. [2007]. This framework incorporates a large body of scientific findings in the context of data quality [Wang/Strong 1996, Wang 1998, English 1999, Redman 2001, Lee et al. 2006]. We regard this framework as a given design artifact (cf. Section 6.3) and use it as a domain representation for CDQM. As the CDQ MM focuses on the implementation phase of a comprehensive CDQM project, we have to consider both governance (P1 – P2) and execution (P4 – P6) perspectives. The following listing shows the six CDQ practices defined by the CDQ framework [Otto et al. 2007, pp. 921-924]:

- *P1: Develop CDQ strategy.* The CDQ strategy includes the strategic objectives which are pursued by data quality management, i.e. how CDQM is aligned with the company’s strategic business goals and its overall functional scope.
- *P2: Design CDQ organization.* CDQM requires the definition of clear roles and responsibilities across divisional boundaries.
- *P3: Design CDQ IS architecture.* This practice is supposed to ensure consistent understanding of data quality across the enterprise.
- *P4: Communicate and control CDQ strategy.* Since strategic realignment of CDQ management requires a change in the behavior of employees, this practice should make intensive use of measurements to accompany organizational change management.

- *P5: Execute and monitor CDQ processes.* Effective CDQM requires compliance with standards, policies, and procedures. Compliance is monitored according to previously defined metrics and performance indicators and reported to stakeholders.
- *P6: Operate and maintain CDQ IS architecture.* This practice deals with operation and maintenance of the IS architecture for data quality management. It includes the systems for storage and distribution of data (such as enterprise resource planning systems) as well as the systems for management and improvement of data quality.

6.2.2 Maturity models

In the 1930s, Shewhart [1931] began work in process improvement. His principles were further refined by Deming [1986], Crosby [1979], and Juran [1988]. Humphrey and others extended Sheward's findings and applied them to software. This adaptation is referred to as Capability Maturity Model (CMM). It provided a description of the basic principles and concepts on which many of today's maturity models are based.

The principle idea of a maturity model is to briefly describe the typical behavior (activities) exhibited by an organization at a number of levels of maturity [Fraser et al. 2002, p. 244]. For each activity, it provides a description of the activity as it might be performed at each defined maturity level [Fraser et al. 2002, p. 246]. In general, maturity models are designed to assess the maturity of a selected domain [de Bruin et al. 2005]. Besides assessment criteria, a maturity model provides guidelines how to reach the next, higher maturity level, as the descriptions of higher maturity levels can be seen as best practice guidance [Fraser et al. 2002, p. 244].

In general, maturity assessment is conducted by assessors [EFQM 2003b, SEI 2006b, p. 68]. However, maturity assessment can also be made by an external auditor or by self-assessment [Fraser et al. 2002, p. 247]. As the CDQ MM is intended to be used as a management tool supporting a company's transformation process, we focus on internally performed self-assessment.

6.2.3 Analysis of existing maturity models

As already stated in Section 6.1, there are multiple approaches for both researchers and practitioners to develop maturity models. In this section, we will provide a brief description of some models and we will discuss whether or not the models cover the domain of CDQ. We will then analyze the domain fit along the six CDQ practices described in Section 6.2.1 (referred to P1 – P6).

Since 2002, the Software Engineering Institute (SEI) of the Carnegie Mellon University has been promoting the Capability Maturity Model Integration (CMMI). CMMI

uses six levels for maturity assessment regarding business processes of organizations [SEI 2006b, p. 73ff]. Also provided is an integrated set of guidelines for improvement of processes. CMMI covers 22 domains (e.g. project planning, or risk management) and provides guidelines for improvement by defining goals and practices [SEI 2006b, p. 29ff]. Furthermore, well-defined requirements and proceedings for CMMI assessment are provided [SEI 2006a]. Although CMMI does not consider any data management specific objectives, it is the de-facto standard of maturity models and therefore has been selected here for evaluation.

EFQM (European Foundation for Quality Management) has defined a framework for organizational management systems, the so-called EFQM Excellence Model [EFQM 2003b]. There are essentially two parts to the EFQM Excellence Model: five enabler criteria and four result criteria, with the enablers underpinning the results. While enabler criteria are concerned with how an organization undertakes key activities, result criteria are concerned with what results are achieved. Against these criteria, an organization's level of excellence can be assessed. For assessment, a generic methodology (RADAR Logic [EFQM 2003a]) is provided which assigns scores to each criterion. Finally, the resulting scores are weighted to gain an overall score representing the degree of excellence. Like the CMMI, the EFQM model does not address any data quality specific aspects, but as the RADAR logic and its scoring process is an interesting approach to assess maturity, the EFQM model will be included in our analysis.

Lee et al. [Lee et al. 2002] have proposed a methodology called AIM quality (AIMQ), which forms a basis for information quality assessment and benchmarking. AIMQ consists of a data quality model, a questionnaire to measure data quality, and techniques for analysis for interpreting the measures. The questionnaire consists of 65 items grouped by data quality dimensions (e.g. 'Accessibility: This information is easily retrievable'). Each item is measured on a 0 to 10 scale. The different analysis techniques are then used to compute weighted average scores of the items. The results of the analysis are useful for determining the area best suited for data quality improvement. AIMQ has been successfully applied in various organizational settings (i.e. financial, healthcare, and manufacturing industries), where it helped to identify data quality problems and monitor improvements over time.

The DQM Maturity Model (DQMMM) focuses on data structure quality and management [Ryu et al. 2006]. The maturity model can be applied to business fields to assess the level of data quality management. The DQMMM introduces four maturity levels for data structure quality as well as the requirements for each level. Furthermore, guidelines for reaching the next, higher maturity level are provided. The levels and its descriptions are based on the findings of several case studies conducted by the authors.

The Information Quality Management Maturity (IQMM) is based on Crosby's maturity model [Crosby 1979] and uses five levels to assess data quality [English 1999, pp.

429-437]. The purpose of the model is to understand the essential ingredients required to implement a data quality management function and to make it an effective business management tool. The author also gives simple guidelines to make a self-assessment.

Gartner's Data Quality Maturity Model (GDQMM) identifies five levels of data quality management success [Bitterer 2007]. The GDQMM is a tool to assess an organization's current state of DQM awareness and effectiveness. The model helps to evaluate gaps in applied DQM practices. Similar consultant models [IBM 2007, DataFlux 2008] will not be described here. Instead, we chose GDQMM as a representative for all these models, which all are CMM based and provide a framework for prioritizing actions, a starting point, a common language, and a method to measure progress.

Table B6-2 illustrates the results of our analysis regarding the models' coverage of the CDQM domain. We have analyzed the domain fit along the six CDQ practices described in Section 6.2.1 (referred to P1– P6). The Harvey Balls are used to indicate the degree to which each of the maturity models covers the particular practice. An empty ball (○) indicates low coverage; a fully filled ball (●) indicates high coverage. CMMI and EFQM cover the six CDQ practices poorly, which was expected. As mentioned before, these models do not consider any data management specific objectives. However, both models make general statements about the topics addressed in P1, P2, P4, and P5.

	P1	P2	P3	P4	P5	P6
CMMI	◐	◐	○	◐	◐	○
EFQM	◐	◐	○	◐	◐	○
AIMQ	○	○	○	◐	○	○
DQMMM	○	○	○	◐	○	●
IQMM	◐	◐	◐	◐	◐	◐
GDQMM	◐	◐	◐	◐	◐	◐

Table B6-2: Domain fit assessment results

AIQM focuses on the measurement of data quality only (P4). It is a profound tool to assess and benchmark data quality in organizations as well as to identify data quality problems. The model does not cover the other CDQ practices though. DQMMM can be seen as a complete realization of P6, since it discusses in great detail the build process of an information architecture, which finally ensures data quality. This build process is accompanied and controlled by measurements and self-assessments (P4). IQMM and GDQMM address all six CDQ practices. However, neither practice is discussed in great depth and detail, nor neither model provides specific guidelines for improvement.

As a bottom line, we can sum up our findings concerning existing data quality models as follows: either the model is focusing on certain aspects of single CDQ practices, or

the model covers almost all CDQ practices but fails to go into detail or provide guidelines for improvement.

6.3 Research Methodology

6.3.1 Research context

The context of this work is formed by a collaborative research program aiming at the advancement of the scientific body of knowledge and the development of artifacts of practical benefits in the domain of corporate data quality management. Table B6-3 shows key data of the participating companies. All of them are large corporations, headquartered in Europe, with international operations. The majority of companies seeks to establish a CDQM function for the first time (marked as ‘new’ in Table B6-3).

Sector	Revenue	Staff	CDQM scope	CDQM status
Chemicals	5.8 bn. EUR	17,800	corporate	existing
Automotive	99.4 bn. EUR	272,382	line of business	new
Public Sector	4.0 bn. EUR	36,058	corporate	new
Telecommunications	62.5 bn. EUR	244,000	line of business	new
Utilities	68.7 bn. EUR	87,815	line of business	new
Consumer Products	3.7 bn. EUR	n/a	subsidiary	existing
Automotive	12.6 bn. EUR	57,372	corporate	new
Electrical Engineering	46.3 bn. EUR	271,000	corporate	new

Table B6-3: Research partners (data from 2007)

Focus group interviews were conducted in order to evaluate the adaptability of the CDQ framework and to analyze practical demands on a CDQ MM in general. The focus group method is an approach to collect information on a certain topic through group interaction [Cavana et al. 2001]. Participants in both focus groups were subject matter experts from the research program partners, mainly project and line managers from IT and data management functions. The homogeneity of the group allowed for deep insight into the domain [Cavana et al. 2001].

Evaluation of the research results of the collaborative program is taking place in action research projects within the partner companies. Action research is a participatory approach combining the advancement of scientific knowledge and the development of solutions to practical problems [Hult/Lennung 1980, Baskerville/Wood-Harper 1996]. Action research studies on the establishment of CDQM that is based on the CDQ framework are carried out with partners listed in Table B6-3. An action research study on the evaluation of the CDQ MM is conducted with the partner given in the last line. Section 6.4.3 documents the results of the first action research cycle, which covers diagnosing, action planning, and action taking as well as first evaluation results.

6.3.2 Design research

The paper at hand adopts design research as a research framework, which is recommended for scientific studying in the field of Information Technology (IT) when artificial, human-made phenomena, such as organizational and/or information systems, are examined [March/Smith 1995, p. 253]. IT artifacts capable of solving research problems in the information system realm constitute the centerpiece of design research as they are both developed and evaluated [Hevner et al. 2004, p. 78, Gregor 2006, p. 62]. The essential elements of the design research approach are summarized in seven guidelines propagated by Hevner et al. [Hevner et al. 2004, pp. 82-90]. The following listing shows how we have adopted these guidelines to our work.

- *G1: Design as an Artifact.* We develop the CDQ MM as a reusable reference model [vom Brocke 2007, p. 50ff]. Models are recognized as viable artifacts in information system research [Hevner et al. 2004, p. 78].
- *G2: Problem Relevance.* We are requested to develop a scientifically founded approach for CDQM maturity assessments in the context of a collaborative research program in cooperation with a number of large enterprises (cf. Section 6.3.1).
- *G3: Design Evaluation.* The collaborative research program (cf. Section 6.3.1) forms the context for the evaluation of the CDQ MM. Multiple cycles of action research and focus group interviews are used for evaluation purposes. Section 6.4.3 illustrates the evaluation process in the case of one research program partner. Due to the multiple facets of reference model evaluation (cf. Section 6.3.3), this test for adaptability is just seen as a first step in a long-term evaluation process.
- *G4: Research Contributions.* The scientific contribution of our research is to provide an approach to assess the maturity of data quality improvement initiatives. Further research is recommended to extend, evaluate, or refine our approach.
- *G5: Research Rigor.* The design of the artifact is based on established research methods (i.e. conceptual modeling, reference modeling (cf. Section 6.3.3), design research) and systematically enhances existing concepts (existing maturity models, CDQ framework).
- *G6: Design as a Search Process.* The research results have been derived iteratively within the research program setting outlined above. Further iteration cycles will be conducted with the adoption of the CDQ MM in other companies, leading to a refinement of the model and prompting more significant evaluation results.

- *G7: Communication of Research.* The insights gained in the course of our research as well as the concepts elaborated are being disseminated via publications and the collaborative research program.

6.3.3 Reference modeling

As stated for design research guideline G1, the CDQ MM is designed as a reusable reference model. A reference model can be defined in terms of a formal language with a formal grammar as a precise description of the language. It provides a set of constructs and rules, with the rules defining which of the possible combinations of constructs constitute valid statements in the language. A modeling method provides procedures by which a grammar can be used. The modeling process produces scripts which are statements in the language that is defined by the grammar. A script is a representation of a real-world domain (e.g. CDQM of a firm) using a particular grammar. In these terms, a reference model is a script representing a class of domains (e.g. CDQ). It can be used as a blueprint ('to-be' situation) for information system development. To actually use reference models in real-world scenarios, they must be adapted to company requirements [Fettke/Loos 2003b, p. 80].

Reference models are descriptive and prescriptive at the same time. On the one hand, they are intended to provide appropriate descriptions of a domain. On the other hand, they aim at delivering blueprints for a distinctively good design of information systems and related organizational settings [Frank 2007, p. 119]. For evaluation of reference models, several evaluation dimensions have been identified [Fettke/Loos 2003b, p. 87f, Moody 2005, p. 263f, Frank 2007, p. 123ff]. For the work at hand, we consider case studies and action research as appropriate evaluation techniques for the CDQ MM (cf. Section 6.4.3).

Vom Brocke [2007] describes a framework for identifying reference model relations in the context of reuse. This framework differentiates five dimensions of reuse. Table B6-4 illustrates these dimensions and shows how the CDQ MM is classified within the framework. The CDQ MM is designed as a reference model that is intended to be reused in real-world scenarios. Due to the adoption of concepts of existing maturity models (i.e. CMMI and EFQM Excellence Model), the CDQ MM is also designed with reference models in the broadest sense.

In the design process, CMMI (i.e. the differentiation of areas, goals, and practices) and the EFQM Excellence Model (i.e. the differentiation of assessment dimensions and the calculation of maturity levels) are reused in terms of an analogy (cf. Section 6.4.2). Furthermore, the CDQ framework is reused as a domain reference model by means of an aggregation (cf. Section 6.4.2). In the application process, the CDQ framework has to be configured in order to meet company specific needs (not all framework components have to be considered at the beginning of a CDQM build process). If the CDQ

framework is considered to be incomplete, extension of the framework for the purpose of specialization is explicitly allowed. In that case, one has to consider the CDQ framework meta-model (cf. Figure B6-1) as an interface. The dimension of instantiation is considered, if an already existing assessment structure (e.g. already performed EFQM audits, or an existing demand management approach) is reused as an assessment model (cf. Section 6.4.3).

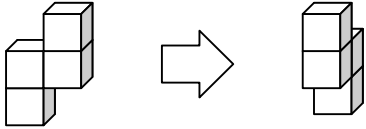
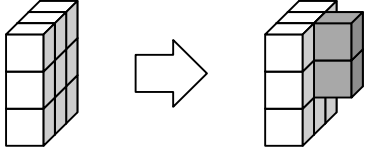
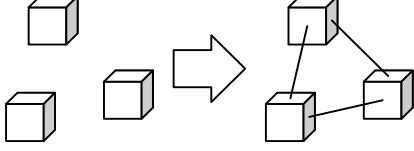
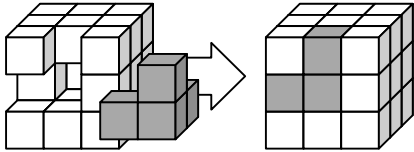
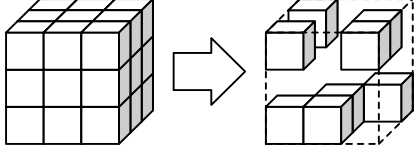
Result dimension	CDQ MM application
<p>Aggregation</p>  <p>Combine existing models</p>	<ul style="list-style-type: none"> The CDQ Framework is adopted into the CDQ MM.
<p>Analogy</p>  <p>Transduced a new model from an existing model</p>	<ul style="list-style-type: none"> There are similar aspects between the CDQ MM and the CMMI and EFQM Model. CMMI and EFQM Excellence Model serve for orientation.
<p>Configuration</p>  <p>Select relevant entities</p>	<ul style="list-style-type: none"> The CDQ Framework is configured for implementation. Not all elements of the framework are always needed.
<p>Instantiation</p>  <p>Integrate existing models into place holders</p>	<ul style="list-style-type: none"> The assessment model (cf. Section 6.4.2) serves as a place holder. Existing company structures can be used to define maturity levels.
<p>Specialization</p>  <p>Derive a special model from a generic model</p>	<ul style="list-style-type: none"> If the CDQ Framework is considered to be incomplete, it is necessary to add some aspects to cover company specific needs.

Table B6-4: Reference modeling reuse dimensions

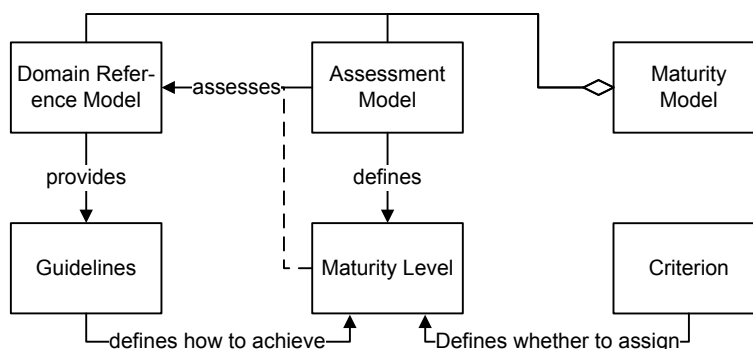


Figure B6-1: Components of a maturity model

6.4 CDQ maturity model

6.4.1 Model requirements

In Section 6.2.3, we have analyzed whether or not existing maturity models cover the domain of CDQM. Besides this generic requirement, we have identified further requirements on a CDQ MM during several focus group interviews (cf. Section 6.3.1). The result is a set of six requirements on a maturity model in the context of CDQM. This set of requirements is regarded as neither complete nor universally valid. It is a result in the given research context and it is used as a design guideline in the construction process described in Section 6.4.2.

- *R1: Specify guidelines how to improve.* The CDQ MM is supposed to be used as a management tool supporting the build process of CDQM. So the maturity model must not only assess a company's CDQM maturity level but it also has to provide guidelines (i.e. best practices) how to reach the next, higher level of maturity.
- *R2: Maximize objectivity during assessments.* Fuzziness of assessments can be reduced by subdividing the domain into smaller entities. In order to ensure standardized assessment procedures (cf. R5), a domain structure with small, independently assessable entities is needed. However, it has to be noted that a maturity model usually contains a portion of fuzziness [Fraser et al. 2002, p. 247, Farrukh et al. 2003, p. 1500].
- *R3: Enable Separating assessments already at an early stage.* CDQ is a complex, multifaceted construct. However, companies that began to establish CDQM have to monitor progress already at an early stage, e.g. to argue investments or for effective resource management. For that reason, the CDQ MM has to be applicable also for small and initial initiatives.
- *R4: Enable assessments along multiple dimensions.* During discussions about the suitability of the CMMI and the EFQM Excellence model, we identified the

need for distinct dimensions for CDQM maturity assessments. Three dimensions will be specified in Section 6.4.2.

- *R5: Provide founded assessment methodology.* The maturity model has to provide a methodology (i.e. guidelines and tools) how to conduct maturity assessments in order to avoid ‘finger in the wind’ ratings.
- *R6: Enable adaptable self-assessments.* It is hardly realistic to establish new quality assessments in a large-scale firm. Mostly, there are already existing approaches, both standardized and customized. As the proposed version of the CDQ MM is intended to be used for self-assessments, it has to provide sufficient flexibility to adopt existing quality management structures.

Table B6-5 depicts whether or not the selected models are able to fulfill these requirements. The differentiation of areas, goals, and practices of CMMI meets the requirements of CDQM maturity assessments (R1, R2). Its structure allows focusing on both general areas (e.g. organizational issues) and specific activities (e.g. definition of target values for data quality metrics). Furthermore, it reflects well the idea of guidance by defining practices that lead to tangible results. However, CMM based maturity models focus on the improvement of already defined and more or less established processes [SEI 2006b, p. 4]. When using these models for assessing a company’s CDQM maturity, one has to assign level 0 (i.e. incomplete) to all companies that do not have a performable CDQM process in place [SEI 2006b, p. 33]. For that reason, CMM based models are considered to be inappropriate for CDQM maturity assessments (cf. R3).

	R1	R2	R3	R4	R5	R6
CMMI	●	●	◐	○	●	◐
EFQM	◐	◐	●	●	●	●
AIMQ	○	◐	●	●	◐	◐
DQMMM	◐	◐	◐	○	◐	◐
IQMM	◐	◐	◐	○	◐	◐
GDQMM	◐	◐	◐	○	◐	◐

Table B6-5: Domain fit assessment results

CMMI, DQMM, IQMM, and GQMM are considered to be incapable of differentiating multiple maturity dimensions (cf. R4), as they only focus on assessing process performance. This shortcoming of most CMM based maturity models is also identified by Rosemann and de Bruin [2005]. The EFQM Excellence Model meets this requirement as business excellence is assessed along nine criteria.

CMMI and EFQM provide well defined tools and methodologies (SCAMPI and the RADAR logic) to be able to make self-assessments (cf. R2, R5, R6). EFQM has some advantages as it provides weightings of the assessed criteria. Therefore, it is more cus-

tomizable to a company's specific needs. In terms of objectivity, CMMI's SCAMPI methodology is more rigorous and offers less fuzziness.

The result of the analysis recommends a CDQ MM which combines the strengths and advantages of both the EFQM and CMMI model. While EFQM covers four requirements (R3– R6), CMMI refers to R1 and R2. Therefore we propose a CDQ MM based on a modified EFQM model. The domain context is provided by the contents of the CDQ framework (cf. Section 6.2.2). Although the other models explicitly address the domain of data quality, their structures and assessment methodologies are considered not flexible enough to fulfill the requirements identified.

6.4.2 Model construction

Analyzing the requirements from Section 6.4.1, we can differentiate three requirements regarding domain structuring (R1, R2, and R3) and three requirements regarding maturity assessment (R4, R5, and R6). Due to this categorization, we decided to describe the CDQ MM by two complementary components: a domain reference model and an assessment model. Even though this separation is somewhat artificial (existing maturity models do not draw a strict cutting line between these aspects), our design allows constructing exchangeable components and enables easier reuse of already existing components.

The domain reference model provides a structured view on a specific domain (e.g. CDQ). In terms of Fraser et al. [2002, p. 246], it describes the activities performed by a firm regarding the focused domain. The assessment model defines maturity levels used during the assessment process to which the components described by the domain reference model are assigned (e.g. activity A is on level 3). In terms of Fraser et al. [2002, p. 246], the assessment model provides a description of each activity as it might be performed at each defined maturity level. During the assessment process, these descriptions are regarded as criteria that indicate which level to assign. Figure B6-1 shows the components as well as the relations between them in form of a meta-model. This meta-model does not claim completeness. Instead, it serves as an instrument to provide well-defined terms and dependencies for the construction process of the CDQ MM.

We regard the CDQ framework (cf. Section 6.2.1) as a given design artifact and adopt it as domain reference model for the CDQ MM. However, in comparison to ratings for R3 (cf. Table B6-5), the single-layered structure of the originally proposed framework with only six entities (cf. Section 6.2.1) is not appropriate to build a suitable domain reference model for the CDQ MM. Due to this drawback, we designed a three-tier meta-model for the framework, which is transduced from the CMMI structure (cf. Figure B6-2).

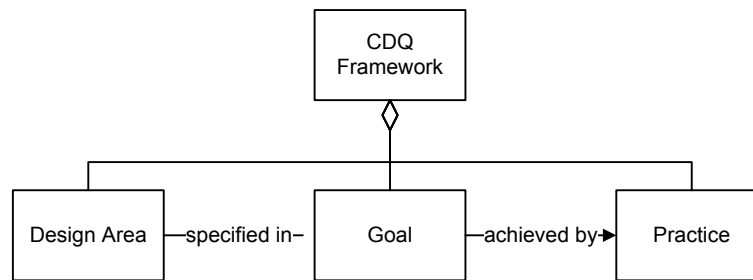


Figure B6-2: CDQ framework meta-model

The following listing defines the three components of the revised CDQ framework following the CMMI specification [SEI 2006b]. The framework consists of design areas which are specified in goals that in turn are achieved by practices.

- A design area is a cluster of related best practices in a certain area, which when implemented collectively meet a set of goals considered important for achieving significant improvement in that area [SEI 2006b, p. iv].
- A goal describes characteristics that must be present to satisfy a process area [SEI 2006b, p. 19].
- A practice is the description of an activity that is considered important in achieving the associated goal. The practices describe the activities that are expected to result in achievement of the goals of a process area [SEI 2006b, p. 20].

The new framework structure explicitly considers R1 by providing (best) practices for the achievement of goals. Besides the still existing top-level structure provided by six process areas (i.e. revised entities of the original CDQ framework), the current version of the new framework provides 52 practices and 17 goals that enable assessments on a very detailed level (cf. R3). Some goals and practices will be discussed in the context of a real-world implementation in Section 6.4.3.

For the assessment model, we consider the CDQ framework as a domain reference model. However, regarding Figure B6-1, any reference model for the domain of CDQ might be used as long as the model provides guidelines how to get to the next, higher level and maturity levels can be assigned to subcomponents of the model.

During the design process of the assessment model, we conducted focus group interviews in order to find out about the applicability of existing assessment models of both the CMMI and the EFQM Excellence Model. As discussed in Section 6.2.3, these models do not meet the requirements for CDQM maturity assessments. So we decided to construct a meta-model for the assessment model that covers requirements R4, R5, and R6. Due to sufficient ratings for the EFQM Excellence Model regarding R5 and R6, we reuse the EFQM approach for assessments and just add a concept for specifying maturity dimensions (cf. Figure B6-3).

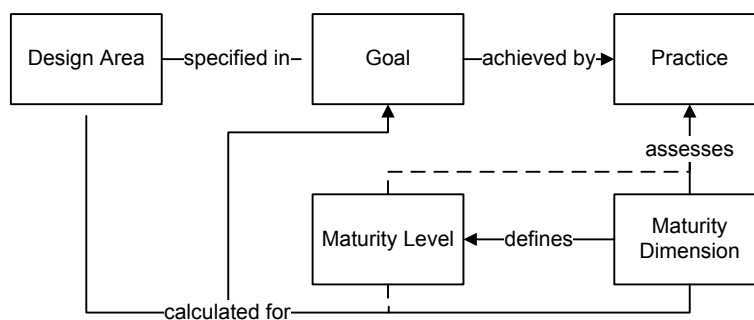


Figure B6-3: CDQ MM assessment meta-model

Figure B6-3 shows the meta-model of the CDQ MM assessment model (based on the meta-model of the CDQ framework). CDQM maturity is assessed along maturity dimensions. Each dimension defines an independent set of maturity levels, where the pair consisting of a set of maturity levels and a dimension is unique for the whole CDQ framework. We assume a number of three to six maturity levels for each set and a logic progression through stages [Fraser et al. 2002, p. 246, de Bruin et al. 2005]. Levels are assigned to practices of the CDQ framework. Maturity levels of goals and design areas are calculated based on ratings of the subordinate level using customizable calculation functions. For the calculated levels, weightings are assigned to all components of the framework concerning individual priorities of a company. In this context, a weighting of zero is similar to the configuration design principle for reference models (cf. Table B6-4). The following listing specifies three maturity dimensions that have been identified by focus group interviews in the already described research context (cf. Section 6.3.1).

- *Propagation of CDQM.* A long-term initiative like the establishment of CDQM in large-scale firms is not carried out in a ‘big bang’ strategy for the entire enterprise at once but has to be introduced step-wise, e.g. per line of business, per country etc. After a proof of concept and some initial success, the initiative is propagated. The dimension propagation of CDQM refers to the percentage of parts of a company covered by CDQM.
- *Corporate data classes.* Similar to the propagation dimension, this dimension refers to the percentage of corporate data classes covered (e.g. material data, vendor master data, chart of accounts).
- *Progress of practices and goal achievement.* Most of the practices proposed by the CDQ framework are time consuming and need further project management efforts. For example, the build process of a data governance model includes conduction of several workshops [Wende 2007]. For that reason, it might be useful to track the progress of some practices along a generic project management structure.

The presented set of dimensions is explicitly not considered to be complete but extensible in terms of instantiation (cf. reuse dimension Instantiation, Table B6-4). Also, the

suggested dimensions can be configured and specialized in order to fit company specific needs. Especially the progress of practices dimension is intended to be instantiated by company specific project or demand management structures.

6.4.3 Model discussion and evaluation

Fraser et al. [2002, p. 247] point out that maturity models have to be evaluated by qualitative approaches such as interviews or beta-testing, as they are generated using experience based principles. To evaluate our approach, we conducted a first implementation at the electrical engineering company mentioned in Table B6-3. In 2006, the company started a corporate master data management (CMDM) initiative aiming at the improvement of the quality and use of corporate master data. Corporate master data is defined as data used in more than one division. It comprises vendor, material, and customer master data, amongst others.

The company has institutionalized the CMDM by issuing a corporate policy defining the overall scope of the initiative (in terms of corporate master data classes such as material master data, vendor master data etc.) and a data governance model [Wende 2007] including the necessary tasks and the roles to carry out the tasks. Roles include master data owners, master data offers, a working group, and a steering committee for CMDM. A master data owner is assigned for each corporate data class. They are responsible for all functional and organizational aspects related to a specific corporate master data class, such as providing a unique definition and description of the corporate data class, developing a functional data model, defining creation and maintenance processes etc. Apart from that, the data governance model instructs master data owners to follow a standardized procedure model ensuring consistent design, handling, and usage of corporate master data. For every corporate master data class, the data governance model defines the following lifecycle stages (cf. reuse dimension Instantiation, Table B6-4):

- Business demands for corporate master data class identified.
- Concept for corporate master data class created (including both a functional and a technical perspective) and approved.
- Resulting projects realized and concept implemented.
- Regular operation.

The organizational unit responsible for coordinating the activities decided to introduce a maturity model to assess and continuously report on the progress of the CMDM initiative. Also, the decision was made to make a self-assessment carried out by every master data owner for their respective corporate data class. In a first step, an evaluation of design areas of the CDQ MM took place, resulting in a principle fit to the requirements. Two adaptations, however, had to be made:

- The company's overall strategy for data management covers all corporate data classes. Therefore, master data owners have to describe the strategy for their class but are not asked to assess the maturity state for their specific class. For that reason, design area Develop CDQ strategy (P1) is not considered (cf. reuse dimension Configuration, Table B6-4).
- Since master data owners only are responsible for functional aspects, design area Operate and maintain CDQ IS architecture (P6) is not used. Instead, another area, Corporate Master Data Usage, is introduced in order to be able to report also on the use of the corporate data provided (cf. reuse dimension Specialization, Table B6-4).

In a second step, the assessment model was developed according to the meta-model proposed in Section 6.4.2. The company put the focus on the identification of appropriate maturity levels. One requirement was to reflect the corporate policy as much as possible since the policy had already been communicated and accepted within the organization. Four maturity levels reflect the lifecycle stages already mentioned above that are specified in the corporate policy: (1) demands identified, (2) concept developed and approved, (3) projects realized and concept implemented, and (4) regular operations. Moreover, in order to depict different roll-out strategies, corporate master data owners specify the propagation progress for each pair of design area and lifecycle stage. For this dimension, a proprietary set of three dimensions is defined (i.e. (1) not completed, (2) partially completed, and (3) fully completed). This graduation applies e.g. when the concept for material master data is implemented country by country (cf. propagation dimension, Section 6.4.2). Besides these two maturity dimensions (i.e. progress of CMDM initiatives and propagation progress), goals of the CDQ framework are used to detail the transition from one maturity level to another. A next, higher maturity level may only be reached if all goals for the complete scope of the corporate master data class are achieved.

6.5 Conclusion and future work

The implementation shows that the CDQ MM can be applied in a real-world scenario, but has to be refined and configured to meet specific requirements. Further implementations will have to show whether these requirements are company specific or can be included in the generic concept. Furthermore, real-world implementations will have to be examined for identifying generic weightings for particular goals and practices or dependencies between several goals (i.e. preconditions in order to develop a staged improvement path). The presented case does not provide clues in this context, as practices were not considered at all, and goals were considered only to a small extent. Moreover, further research will have to focus on developing generic maturity levels in order to enable benchmarking.

7 The effect of using a semantic wiki for metadata management: A controlled experiment

Titel	The Effect of Using a Semantic Wiki for Metadata Management: A Controlled Experiment ²¹
Autoren	Kai M. Hüner; Boris Otto Universität St. Gallen, Institut für Wirtschaftsinformatik Müller-Friedberg-Strasse 8, 9000 St. Gallen, Schweiz {kai.huener boris.otto}@unisg.ch
Publikationsorgan	Sprague, R. H. (Hrsg.): Proceedings of the 42nd Hawaii International Conference on System Sciences, Waikoloa, Hawaii, USA, 2009
Status	Veröffentlicht

Tabelle B7-1: Bibliographische Angaben zum Beitrag „The effect of using a semantic wiki for metadata management: A controlled experiment“

Abstract

A coherent and consistent understanding of corporate data is an important factor for effective management of diversified companies and implies a need for company-wide unambiguous data definitions. Inspired by the success of Wikipedia, wiki software has become a broadly discussed alternative for corporate metadata management. However, in contrast to the performance and sustainability of wikis in general, benefits of using semantic wikis have not been investigated sufficiently.

The paper at hand presents results of a controlled experiment that investigates effects of using a semantic wiki for metadata management in comparison to a classical wiki. Considering threats to validity, the analysis (i.e. 74 subjects using both a classical and a semantic wiki) shows that the semantic wiki is superior to the classical variant regarding information retrieval tasks. At the same time, the results indicate that more effort is needed to build up the semantically annotated wiki content in the semantic wiki.

7.1 Introduction

7.1.1 Corporate metadata management

Effective management of diversified companies implies coherent and consistent understanding of corporate data. Logistics processes harmonized on a global level, global procurement strategies, integrated service management, and compliance with regulatory and legal requirements needs company-wide unambiguous data definitions.

Being able to work with high-quality data is pivotal to enable service-oriented business applications [Newman/Logan 2006, p. 4], help increase the validity of strategic deci-

²¹ Institute of Electrical and Electronics Engineers (IEEE) 2009

sions [Shankaranarayanan et al. 2000, p. 23ff, Muljadi et al. 2006, Paoli/Loregian 2006], and allow for high regulatory compliance [Friedman 2006, p. 5]. Metadata of high quality are the basis for retrieving, managing, evaluating, and using appropriate information through precise queries, as they increase user confidence in data and improve decision-making quality [Marco 2000, p. 18ff]. In this respect, metadata management refers to assigning standard (textual) definitions to data, and to maintenance of such data in a centralized dictionary or metadata repository [Dyché/Levy 2006, p. 165], such as a (semantic) wiki.

7.1.2 Business scenario

In cooperation with a chemical company, we are conducting a Design Research project [March/Smith 1995, Hevner et al. 2004] for developing a central metadata repository. The company is running a global process harmonization project and is planning to establish a central single point of truth for master data. In this project, the company needs harmonized and comprehensible data definitions in order to provide guidance for users (i.e. users of SAP systems maintaining materials, vendor, and customer master data) that work in newly harmonized business processes. The repository is intended to be used world-wide by about 6.000 users for support purposes.

A central metadata repository is already in place providing information on company-specific SAP data fields and transactions. This repository is supposed to be used by SAP users for clarifying the usage of particular SAP data fields, forms, and transactions. However, acceptance and usability of this tool are low. Some problems of the current solution were identified in focus group interviews [Cavana et al. 2001, pp. 153-158]:

- information retrieval is too time consuming,
- search and navigation functionality is poor,
- information is not up to date,
- metadata is not integrated with business processes.

Considering the interview results, the company decided to use a wiki for a new version of the metadata repository. This decision was based on the assumption that the wiki could increase the willingness of employees to maintain support information and to exchange experiences. These expected effects are supposed to improve timeliness of information and increase the number of users. However, the evaluation of this assumption is part of ongoing research. The paper at hand discusses whether or not to use a semantic The prior decision to use wiki based software for a new version of the metadata repository is considered as a given fact.

Focus group interviews about design alternatives [Hevner et al. 2004, p. 88f] for wiki based software results in considerations whether or not to use a semantic wiki instead of a classical wiki (cf. Section 7.2.2). Due to uncertainty about the impact of functio-

nality that is only provided by a semantic wiki, we decided to conduct a controlled experiment to investigate positive or negative effects of the use of a semantic wiki. This experiment was not intended to evaluate the above mentioned wiki-based metadata repository, but to indicate whether or not a semantic wiki is able to overcome supposed drawbacks of a common wiki in general. The intended contribution is to empirically confirm supposed effects in order to provide decision support for wiki development projects. An empirical evaluation of the mentioned Design Research project and a comparison with the findings presented in this paper is part of future work.

Considering the interview results mentioned above, we decided to focus on information maintenance and information retrieval. Research to date has tended to focus on effects of using classical wikis in corporate context [Majchrzak et al. 2006, Muljadi et al. 2006, Paoli/Loregian 2006, White et al. 2008, Yang et al. 2008]. So far, however, there has been little empirical research on effects of using semantic wikis for metadata management.

The paper at hand presents the results of a controlled experiment on the effect of using a semantic wiki for metadata management. In particular, we investigated how enhanced wiki functionality affects 1) time needed to accomplish certain tasks and 2) subjects' assessments regarding effort and usability. In Section 7.2, we outline our understanding of metadata and give some background information on (semantic) wiki technology used for the experiment. Section 7.3 explains the rationale behind the chosen experiment design and gives information on the experiment's implementation and progression. In Section 7.4, we present the main results of the statistical analysis and we discuss possible threats to validity of the results. Section 7.5 provides a summary and discusses future work.

7.2 Background

7.2.1 Definition of metadata

Metadata can be defined in general as data that describe the meaning and the properties of other data demarcating them from other data types, such as transaction and master data [Burnett et al. 1999, p. 1213]. More precisely, we use the term to determine important characteristics that need to be known for both database and application engineering [English 1999, p. 482], or for general, semantically unambiguous understanding of data within an organization. Metadata, accordingly, facilitate identification, detection, use, and management of data as they allow an organization to better understand its data sources and definitions. Tozer [1999, p. xix] summarizes their function by 'seeing metadata as the means by which the structure and behavior of data is recorded, controlled, and published across an organization'. The most comprehensive definition is provided by Marco [2000, p. 5], who describes metadata as 'all physical data and knowledge from inside and outside an organization, including informa-

tion about the physical data, technical and business processes, rules and constraints of the data, and structures of the data used by a corporation’.

From a data management perspective, metadata is any information that supports effective use of data, including information that can facilitate data management (e.g. data authentication, data sequence, data type, or key field indicators), data access and data analysis (e.g. format for data mining, visualization) [Burnett et al. 1999, p. 1212].

7.2.2 Wikis and semantic wikis

Wiki is a piece of server software that allows users to freely create and edit web page content using any web browser. Wiki supports hyperlinks and has simple text syntax (wiki markup or wikitext) for creating new pages and cross-links between internal wiki pages on the fly [Cunningham 1995]. Further typical features are version control (history management) as well as search and navigation functionality. It has to be mentioned that the term wiki does not refer to specific software or technology but to the concept of web based, collaborative collection, organization, and provision of information.

Wikis have become popular tools for collaboration on the web, and many online communities employ wikis to exchange knowledge. For a majority of wikis, public or not, primary goals are to organize and to share the knowledge collected [Krötzsch et al. 2007, p. 251].

One of the best known examples of a wiki is the multilingual, free encyclopedia Wikipedia²², which every user can edit. Due to rather primitive features for searching information provided by the Wikipedia software, Krötzsch et al. [2005] proposed a more structured approach and started an open-source software project to develop an enhanced wiki engine (i.e. an extension of the wiki engine MediaWiki²³ that is used for Wikipedia). This software, called Semantic MediaWiki²⁴ (SMW), can be classified as a semantic wiki. Both MediaWiki and SMW are mentioned here as they are used for our experiment (i.e. the classical wikis are provided by installations of MediaWiki, while the semantic wikis are provided by installations of MediaWiki and the SMW extension).

Unlike classical wikis, semantic wikis provide a structured representation of their content, mostly encoded in RDF based languages [Souzis 2005, Paoli/Loregian 2006, p. 88]. Based on this additional information structure, enhanced search and navigation functionality is provided. As there is no comprehensive classification of semantic wiki functionality, we just provide information on SMW as a representative of semantic wikis. Please note that there are more semantically enhanced wiki engines (e.g. Mak-

²² <http://www.wikipedia.org>

²³ <http://www.mediawiki.org>

²⁴ <http://www.semantic-mediawiki.org>

naWiki [Nixon/Simperl 2006], MILK wiki [Paoli/Loregian 2006] and IkeWiki [Schaffert 2006]), providing further features, e.g., for collaborative ontology editing or document publishing. However, such features are not in the focus of the business scenario mentioned in Section 7.1.2). For more information, Buffa et al. [2008] present a state-of-the-art of semantic wikis. A case study on the management of corporate manufacturing metadata is presented by Muljadi et al. [2006, p. 271ff].

Figure B7-1 shows a page with information about Gorgonzola cheese that was used in the experiment (i.e. in the semantic wikis). SMW uses additional markup to annotate the content of a wiki page [Krötzsch et al. 2007, p. 254f]. In the classical wiki, a link from the Gorgonzola page to the page about Piedmont looks like this: `[[Piedmont]]` (MediaWiki markup). With SMW, one can add a meaning to this link by adding some additional markup: `[[Production Region::Piedmont]]`. The enhanced wiki engine parses the annotated wikitext and builds a data structure based on the additional information (i.e. a directed graph with a labeled relation (Production Region) from Gorgonzola to Piedmont).



Figure B7-1: Semantically annotated wiki page

Figure B7-2 shows an SMW special page that is available for all semantically annotated pages. It provides information already available on the content page (i.e. in the Facts about box, cf. Figure B7-1) in a condensed representation. This extracted information is intended to increase the wiki's usability regarding navigation and information retrieval.

		Has Spiciness mild
		Country Of Origin Italy + ⓘ
Pizza Rosa + ⓘ	Gorgonzola	Has Name From
Has Topping	Topping	Gorgonzola + ⓘ
		Production Region
		Piedmont + ⓘ, Lombardy + ⓘ
		Categories Cheese
		Topping + ⓘ

Figure B7-2: Enhanced navigation enabled by semantically annotated wiki content

7.3 Research methodology

7.3.1 Research objective

We conducted a controlled experiment with an AB/BA cross-over design. Background information on this design is given in Section 7.3.3. Regarding the underlying Design Research project mentioned in Section 7.1.2, we identified the following two research questions:

- How is a wiki's usability affected by functionality which is enabled by semantically annotated wiki content?
- What additional effort is needed to semantically annotate wiki content?

7.3.2 Scenario

As already mentioned, we focus on real-world metadata management scenarios where users are familiar with the underlying data model. One example of such a scenario is company-specific support for an SAP system, where the wiki contains information on how to use specific SAP data fields (e.g. the sales tax ID as part of supplier master data). For that reason, we tried to minimize the effort necessary to understand given tasks by using a simple, plain scenario. We chose an ontology about pizzas, which is also used for education and training purposes in the context of ontology editing software [Jupp et al. 2007]. More precisely, we used a subset of this ontology: We only selected information about pizzas, their toppings, and the spiciness of pizza toppings. We mapped this information to an appropriate wiki representation. The wikis contain pages for several pizzas (e.g. Pizza Capricciosa) and pizza toppings (e.g. Gorgonzola cheese, cf. Figure B7-1). Relations (e.g. between pizzas and their toppings) are determined by the page content (classical wiki) and by semantic annotations (semantic wiki), respectively.

7.3.3 Experiment design

Regarding our research questions, we needed to take measurements on comparable subjects under different conditions: a classical wiki on the one hand, and a semantic wiki on the other hand. To control the treatment of subjects, we designed a controlled experiment [Christensen 2001, p. 181ff] with the following parameters (more detailed information on the tasks T_{Find} , T_{Collect} and T_{Link} is given in Section 7.3.4):

- *Independent variables:* For two tasks (T_{Find} , T_{Collect}), we controlled the wiki's functionality. Subjects had to accomplish the same tasks with both a classical and a semantic wiki. For another task (T_{Link}), we controlled the scope of the task (adding classical links vs. semantic annotations).
- *Dependent variables:* We measured the subjects' assessments regarding effort and usability in trying to accomplish the given tasks by taking ratings on an interval scale (1–9). Furthermore, we measured the time needed for each task to be accomplished. We consider these variables to serve as proper indicators of the shortcomings identified with regard to the metadata repository currently in use (cf. interview results mentioned in Section 7.1.2).

Given three tasks and three dependent variables, we can formulate nine null hypotheses for our controlled experiment:

- No difference in time needed to accomplish T_{Find} , T_{Collect} , and T_{Link} .
- No difference in effort ratings for T_{Find} , T_{Collect} , and T_{Link} .
- No difference in usability ratings for T_{Find} , T_{Collect} , and T_{Link} .

To achieve higher statistical power [Lipsey 1990, p. 80] and to ensure internal validity [Senn 2002, p. 35ff], we decided to use a repeated-measures design for our controlled experiment. In a repeated-measures design, subjects receive two or more treatments (modification of controlled variables). Even with high subject variability, there will be almost always a positive correlation [Keren 1993, p. 181].

A shortcoming of repeated-measures design is the occurrence of carry-over effects [Greenwald 1976, p. 318]. Carry-over effects can occur in two ways: practice effects and sequence effects. In our context, practice effects can occur because subjects might not have used a wiki in the past. Therefore, it is plausible that in the second run subjects will exhibit better performance than in the first run because they had some practice with a wiki, regardless of the wiki type. When confronted with a sequence effect, the effect of the first treatment persists and thus contaminates the measurements on the second treatment. Results of our analysis to identify carry-over effects are given in Section 7.4.2.

To err on the conservative side, we used a counterbalanced repeated-measures design (i.e. an AB/BA cross-over design [Senn 2002, pp. 46-50]). With counterbalancing, all possible orders of treatments are considered and subject groups are randomly assigned to one of the treatment sequences. The effect of counterbalancing is to spread the unwanted variance arising from interference through practice or sequence effects [Christensen 2001, pp. 212-228]. Furthermore, an explicit test for carry-over effects is conducted during our analysis to check if any such effects persisted despite the counterbalancing.

7.3.4 Experiment implementation

Participants were members of an applied research program as well as employees and consultants that are involved in data quality management. All subjects could be classified as interested in the context of data quality. We had a set of 279 potential participants from more than 80 different companies, which were randomly assigned to two groups (GA, GB). Random assignment can be regarded as a method of experimental control because it is assumed that over large numbers of subjects uncontrolled factors are distributed evenly over the treatment conditions [Winer et al. 1991, pp. 261-262].

The experiment was conducted in April 2008. Each subject received an email containing a short introduction, the URLs of two wikis (i.e. a classical and a semantic wiki) and a link to an online survey. As we used an address repository that was built up during the last two years, the address data was not free of errors. Since 31 emails could not be delivered, the size of the sample was reduced to 248 (128 in GA and 120 in GB). Due to some technical problems during the first two days, 13 data sets could not be considered in the analysis. Finally, we received 74 analyzable results (39 in GA and 35 in GB), which corresponds to a response rate of 29%. It has to be noted that there are smaller sample sizes in some cases as not all subjects accomplished each task. For some statistical calculations, it is also necessary to consider only those subjects that have accomplished a task in both the classical and the semantic wiki.

For each participant, we prepared two wikis: a classical wiki and a semantic wiki. We used the wiki engine MediaWiki for both wikis. For the semantic wiki, we also installed the SMW extension (cf. Section 7.2.2). Copies of the wikis used (i.e. the classical wiki²⁵ and the semantic wiki²⁶) are available online. Furthermore, we asked all participants to assess the wikis' usability and the effort needed to accomplish certain tasks. (cf. dependent variables in Section 7.3.3). Besides the ratings, we implemented a logging extension for the wikis that saves timestamps for several wiki actions (i.e. view, edit, and save actions). With these timestamps, we could estimate the time necessary to accomplish the given tasks. The following subsections describe the three tasks each subject had to accomplish in both the classical and the semantic wiki.

²⁵ <http://wiki.iwi.unisg.ch/experiment/ClassicalWiki>

²⁶ <http://wiki.iwi.unisg.ch/experiment/SemanticWiki>

7.3.4.1 ‘Find Content’ Task (T_{Find})

We asked the subjects to go to five wiki pages. For this task, the classical wikis only provide indexed search functionality and links between wiki pages. Besides search and link functionality, the semantic wikis also provide a factbox that summarizes semantically annotated data of a page (cf. Figure B7-1) and further navigation support mentioned in Section 7.2.2 (cf. Figure B7-2). In order to simulate corporate scenarios, we asked to visit semantically related pages (e.g. a specific pizza and afterwards a pizza with a similar topping).

7.3.4.2 ‘Collect Information’ Task ($T_{Collect}$)

We asked the subjects to collect all wiki pages about pizzas that have at least one hot topping (referred to as hot pizzas). The spiciness information was only given in the page text of a pizza topping. In the classical wikis, the subjects had to read all pizza topping pages for identifying hot pizzas. In the semantic wikis, we additionally provided a GUI based interface for intuitive assembly of ask queries.

7.3.4.3 ‘Link Information’ Task (T_{Link})

We asked the subjects to add missing links into a specific wiki page. The text without links was already included. In the semantic wikis, subjects also were asked to semantically annotate the added links.

7.4 Results and analysis

7.4.1 Statistical background

For this paper, we have visualized some of the most important results by means of box plots [Tukey 1977, p. 39ff]. The box of a certain plot contains 50% of the data points. The lower / upper border of the box marks the 25% / 75% quantile. The lower / upper t-bar marks the most extreme data point which is not more than 1.5 times the length of the box away from the lower / upper edge of the box. Outliers from the above scheme are visualized by circles. The median is marked with a thick line.

We used the non-parametric one-sided two-sample Wilcoxon rank-sum test [Hollander/Wolfe 1999, p. 106ff] for evaluation as we did not assume a certain distribution (e.g. a normal distribution for both data sets) in general. Besides the tolerance for different distributions, this test is also recommended for both ordinal and interval scaled data (despite introductions to assume constant spacing for the rating scale, we cannot assume homogenous interval scaled data). The statistical power of the Wilcoxon test is in the worst case 13.6% smaller than the power of the corresponding t-test [Hollander/Wolfe 1999, p. 139]. This reduction factor is considered for all power calculations in this section. However, all statistical power calculations result in values close to ‘1’.

The reason for these high statistical power values can be found in very small p-values (cf. Section 7.4.3).

We only conducted statistical analysis for T_{Find} and T_{Link} , because we did not receive enough comparable data sets for T_{Collect} (cf. Section 7.4.4).

7.4.2 Cross-over analysis

We applied the Hills-Armitage approach [Hills/Armitage 1979, pp. 12-15] to analyze carry-over effects of the experiment. This regression method is recommended to analyze data from cross-over experiments [Senn 2002, p. 47]. The approach is based on treating the experimental sequence order as a dummy variable that, in our case, takes the value ‘<1’ for data points of GA and ‘1’ for data points of GB. The test is legitimate regardless of whether just one or both of the effects (i.e. practice and sequence effect) is present. However, the test is incapable of discriminating between the two effects [Laitenberger et al. 2001, p. 416]. After a regression analysis, the gradient coefficient of the regression line can be seen as an indicator of a carry-over effect. More precisely, one can determine the significance of the effect by calculating the distance of the coefficient from zero. Table B7-2 shows the cross-over analysis results (abbr.: RC (regression coefficient), SE (standard error), t (t-value), P (probability > |t|), TE (treatment effect), CE (carry-over effect). Significance level: 5%).

			RC	SE	t-value	P
T_{Find}	Effort	TE	3.228	0.180	17.018	< 0.001
		CE	-0.366	0.180	-1.929	0.059
	Usability	TE	2.609	0.167	15.599	< 0.001
		CE	0.046	0.167	0.277	0.783
	Time	TE	30.968	2.635	11.754	< 0.001
		CE	0.687	2.635	0.261	0.795
T_{Link}	Effort	TE	1.841	0.111	16.513	< 0.001
		CE	0.059	0.111	0.533	0.596
	Usability	TE	1.130	0.102	11.049	< 0.001
		CE	0.036	0.102	0.356	0.723
	Time	TE	43.541	2.748	15.847	< 0.001
		CE	-2.741	2.748	-0.997	0.323

Table B7-2: Results of the cross-over analysis

Results of the regression analysis presented in Table B7-2 indicate that the mean difference between values of GA and GB (i.e. the values for TE in column RC) is significantly different from zero for all dependent variables (i.e. $p < 0.001$ for effort, usability, and time). This suggests that the semantic wiki is superior to the classical wiki regarding T_{Find} and T_{Link} . A more detailed analysis of the treatment effect is given in the following subsections. A significant carry-over effect could not be determined: All p-

values (cf. values for CE in column P) are greater than the chosen significance level of 0.05.

7.4.3 Analysis of T_{Find} and T_{Link}

For the analysis, we considered 61 subjects (32 in GA and 29 in GB) regarding T_{Find} and 62 subjects (32 in GA and 30 in GB) regarding T_{Link} . Table B7-3 summarizes mean, standard deviation, correlation coefficients, and results of a significance analysis regarding T_{Find} and T_{Link} (abbr.: CW (common wiki), SW (semantic wiki), SD (standard deviation), Corr (Spearman’s rank correlation coefficient)). P-values calculated by the Wilcoxon test all are smaller than 0.001.

		Mean		SD		Corr
		CW	SW	CW	SW	
T_{Find}	Effort	6.279	3.131	1.803	1.954	0.694
	Usability	4.295	6.836	1.551	1.812	0.630
	Time	204.83	175.836	35.203	43.196	0.811
T_{Link}	Effort	3.387	5.065	1.441	1.740	0.791
	Usability	4.484	3.452	1.720	2.092	0.900
	Time	90.548	134.177	27.008	35.103	0.770

Table B7-3: Significance and correlation analysis

Figure B7-3 and Figure B7-4 show box plots of the subjects’ ratings regarding usability and effort for T_{Find} . The box plots show that subjects using a semantic wiki gave higher ratings regarding usability and lower ratings regarding effort than subjects using the classical wiki. These results are consistent with the time measurements: Almost all subjects accomplished T_{Find} faster in the semantic wiki than in the classical wiki (cf. Figure B7-5). All these results are strongly significant ($p < 0.001$ for effort, usability, and time).

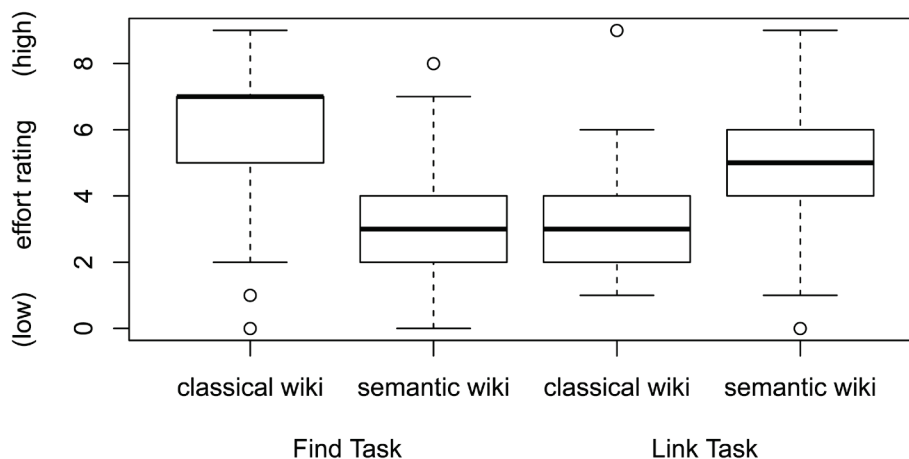


Figure B7-3: Effort ratings for T_{Find} and T_{Link}

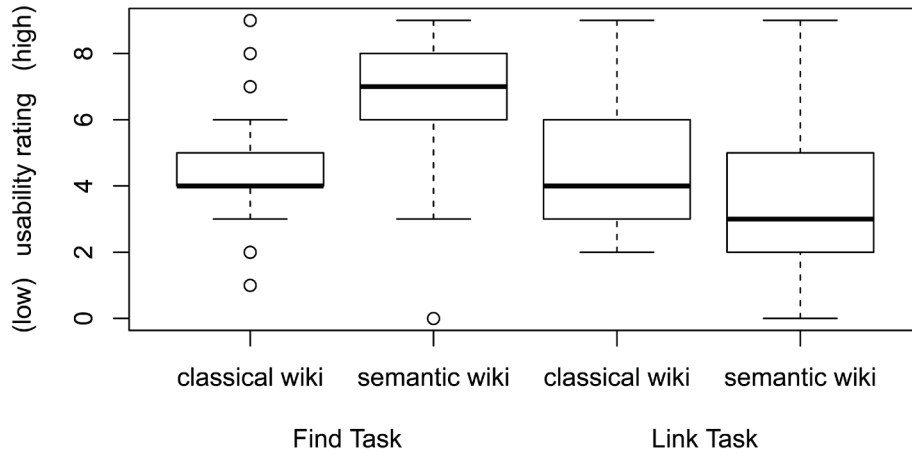


Figure B7-4: Usability ratings for T_{Find} and T_{Link}

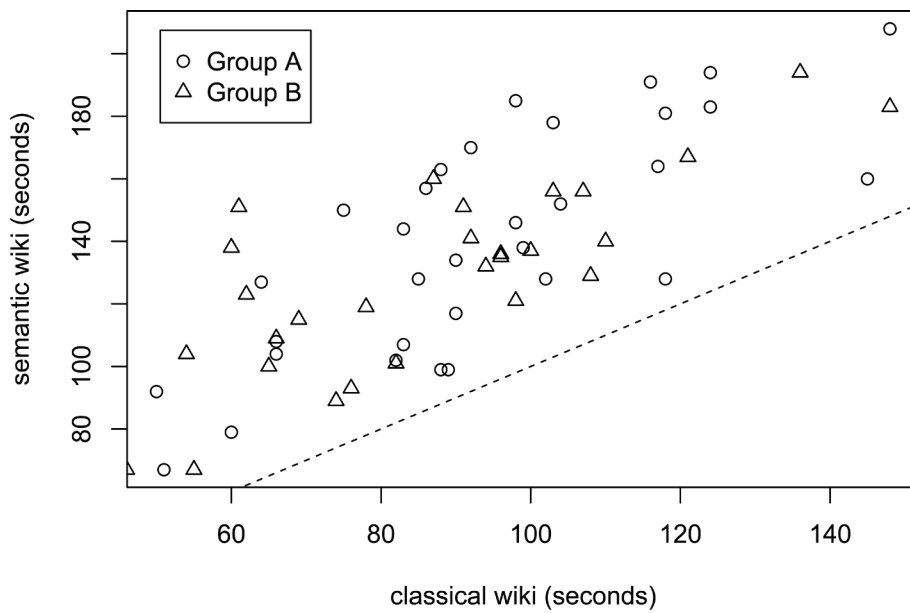


Figure B7-5: Time correlation (0.811) of T_{Find}

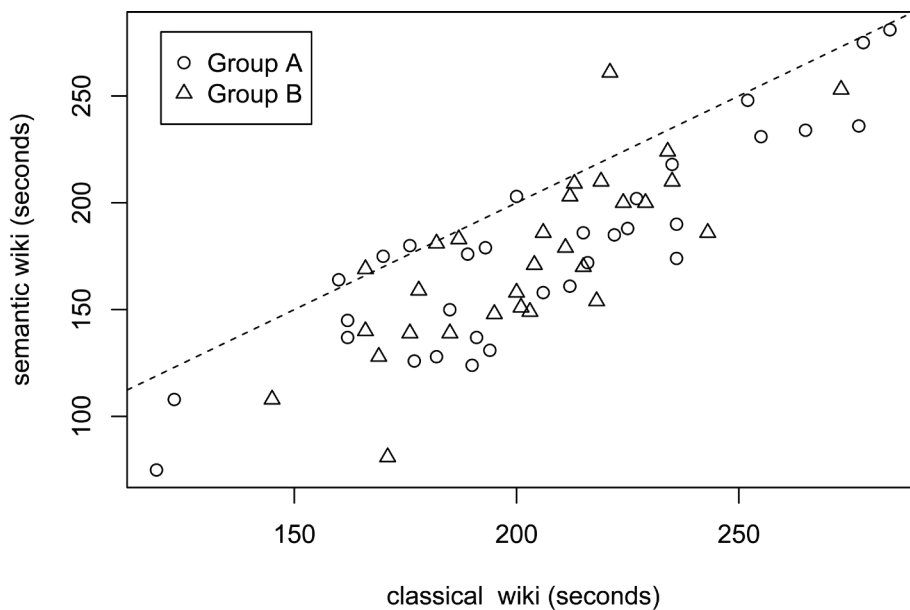


Figure B7-6: Time correlation (0.770) of T_{Link}

Results for T_{Link} are also significant ($p < 0.001$ for effort, usability, and time), yet this time it is the other way round. The box plots in Figure B7-3 and Figure B7-4 and the scatter plot in Figure B7-6 show that subjects using a semantic wiki gave lower ratings and needed more time to accomplish T_{Link} than subjects using a classical wiki. Besides a significant distance, there is a significant positive correlation between each pair of data sets. Values of the Spearman's rank correlation coefficient for all data set pairs are given in Table B7-3.

In summary, all null hypotheses regarding T_{Find} and T_{Link} have to be rejected. Regarding any threats to validity (cf. Section 7.4.5), the semantic wiki performs better for T_{Find} while it shows lower performance for T_{Link} .

7.4.4 Analysis of T_{Collect}

Regarding T_{Collect} , we have too few comparable data sets for significant statistical analysis. Obviously, the given task was too hard to accomplish. Only six subjects tried to accomplish T_{Collect} in the classical wiki. Any of these subjects achieved an acceptable result. With the semantic wiki, 61 subjects (31 in GA and 30 in GB) tried to accomplish the task. It is interesting to note that all seven subjects of GA (remember: classical wiki before semantic wiki) that aborted the experiment were lost while trying to accomplish T_{Collect} in the classical wiki. Subjects of GB also aborted T_{Collect} in the classical wiki but continued with trying to accomplish T_{Link} .

Another interesting finding is the number of 42 correct results for T_{Collect} in the semantic wiki. It has to be noted that the task had to be accomplished by creating an ask statement. Though the semantic wiki provides a query interface to support subjects (cf. Figure B7-7), it is remarkable that 61 subjects tried to use this interface and 42 of them (24 in GA and 18 in GB) succeeded in doing so. Despite this interesting observation, we cannot draw any significant conclusion whether or not to reject the null hypotheses regarding T_{Collect} .

The screenshot displays the SMW query interface. On the left, a 'Query Tree Navigation' pane shows a hierarchy: 'Main Query > Has' followed by 'Topping'. Under 'Topping', there are three items: 'Has Topping' (with a question mark icon), 'Has Spiciness' (with a property icon), and 'String = hot' (with a value icon). On the right, a configuration panel for the 'Has Spiciness' property is shown. It includes buttons for 'Add Category', 'Add Instance', and 'Add Property'. The 'Property name' field contains 'Has Spiciness'. Below it are checkboxes for 'Show in results' and 'Value must be set'. A 'String' dropdown menu is set to '=', and the value field contains 'hot'. At the bottom are 'Add', 'Cancel', and 'Delete' buttons.

Figure B7-7: SMW query interface

7.4.5 Threats to validity

It is the nature of any empirical study that assumptions are made which later on restrict the validity of the results. In this subsection, we discuss these assumptions that impose threats to internal and external validity.

7.4.5.1 *Internal Validity*

Threats to internal validity are those factors that may affect the value of the dependent variables apart from the independent variable [Kitchenham et al. 2002, p. 732]. Effects regarding selection, maturation, instrumentation, and presentation can be considered as not relevant when using a crossover design with random allocation to sequence which is designed to avoid these problems [Senn 2002, p. 35ff]. However, using a cross-over design is problematic if there is an interaction between sequence and treatment. Considering this threat, we conducted a cross-over analysis that is discussed in Section 7.4.2. Results of this analysis do not indicate significant cross-over effects.

7.4.5.2 *External Validity*

Threats to external validity are those factors that may limit the applicability of the experimental results to industry practice [Christensen 2001, p. 360]. Our experiment considered three threats to external validity: participant representativeness, scenario representativeness, and task representativeness.

Participant representativeness is an issue, because all participants are more or less interested in the topic of data quality management. This characteristic cannot be assumed in general for users of wiki based metadata management software. Nevertheless, the familiarity with (semantic) wikis was low: All subjects stated to read on Wikipedia, but nobody had ever written some content. Only twelve subjects (16%) knew the MediaWiki software before the experiment, only one stated to use it. Regarding SMW, only three subjects (4%) stated to know the software.

Of course, the scenario cannot be generalized to industrial practice except to pizzeria businesses. Regarding this threat, we need to discuss whether or not the complexity of the scenario can be viewed as a valid representation for a real-world scenario. One important factor at this point is the knowledge of real-world employees about their specific tasks. For our experiment, we tried to compensate the missing knowledge about the arbitrary scenario with lower complexity. However, one valid argument threatens the results regarding T_{Find} : The semantic wiki can only expedite a user's work process, if the information needed is semantically connected (due to the search and navigation support that is based on semantic annotations).

When discussing the representativeness of the experiment's results, we must admit that the effects identified are certainly dependent on the software that was used. All tasks

defined addressed specific features of SMW (i.e. extracted information to enhance navigation, the query interface, or the SMW markup).

7.4.6 Discussion

We identified a significant difference between the performance of the classical wikis and the performance of the semantic wikis, while certain threats to validity have been taken into consideration. As the experiment was intended to provide decision support as to whether or not using a semantic wiki (i.e. the SMW extension for a MediaWiki based wiki) is beneficial for metadata management, we discuss the results in the context of the Design Research project mentioned in Section 7.1.2.

The experiment's results indicate that information retrieval is faster and easier when extending MediaWiki with SMW. At the same time, the effort for information maintenance increases due to additionally needed semantic annotations in the wiki content. When trying to draw conclusions from these findings, one has to consider specific characteristics of the corporate scenario focused. If the intended effect is to facilitate content creation, a classical wiki would be the better choice. If the wiki is supposed to facilitate information retrieval and the utilization of contained information, a semantic wiki is preferable. Besides this consideration, technical aspects (e.g. integration of wiki content into further applications [Muljadi, 2006 #27; Paoli, 2006 #29]) have to be taken into account.

In the business scenario mentioned in Section 7.1.2, the clear focus is on information retrieval as the number of content users exceeds the number of intended content authors (i.e. 6.000 vs. 90). For that reason, the identified effects concerning T_{Find} are regarded to be more important than increasing effort for content authors. However, the evaluation of these expectations is part of future work. At this point, we just want to outline how the experiment's findings can contribute to given wiki development projects.

7.5 Conclusion and future work

The paper at hand presents the results of a controlled experiment on the effect of using a semantic wiki for metadata management. Taking into consideration threats to validity, we have come to the conclusion that the semantically enhanced wiki engine SMW is superior to the classical variant (i.e. a classical MediaWiki installation) regarding information retrieval tasks. At the same time, the results indicate that in the semantic wiki more effort is needed to build up the semantically annotated knowledge base, which is a precondition for semantically enhanced functionality. These results are being discussed with reference to a real-world metadata management scenario in order to disclose how to draw conclusions for practice.

Further work needs to be done to assess whether the findings are representative for real-world scenarios. The Design Research project mentioned is intended to observe comparable effects for the specific use case of a metadata repository with SAP support information. Besides the effects investigated in our experiment, the effects of deep metadata integration into business processes have to be examined. The possibility to provide a machine processable representation of collaboratively created wiki content that can be easily integrated into enterprise systems (i.e. a SAP system) is supposed to have a positive effect on process performance (i.e. process fidelity, process cycle time).

8 Collaborative management of business metadata

Titel	Collaborative management of business metadata ²⁷
Autoren	Kai M. Hüner; Boris Otto; Hubert Österle Universität St. Gallen, Institut für Wirtschaftsinformatik Müller-Friedberg-Strasse 8, 9000 St. Gallen, Schweiz {kai.huener boris.otto hubert.oesterle}@unisg.ch
Publikationsorgan	International Journal of Information Management 31(4), S. 366-373
Status	Veröffentlicht

Tabelle B8-1: Bibliographische Angaben zum Beitrag „Collaborative management of business metadata“

Abstract

Legal provisions, cross-company data exchange and intra-company reporting or planning procedures require comprehensively, timely, unambiguously and understandably specified business objects (e.g. materials, customers, and suppliers). On the one hand, this business metadata has to cover miscellaneous regional peculiarities in order to enable business activities anywhere in the world. On the other hand, data structures need to be standardized throughout the entire company in order to be able to perform global spend analysis, for example. In addition, business objects should adapt to new market conditions or regulatory requirements as quickly and consistently as possible. Centrally organized corporate metadata managers (e.g. within a central IT department) are hardly able to meet all these demands. They should be supported by key users from several business divisions and regions, who contribute expert knowledge. However, despite the advantages regarding high metadata quality on a corporate level, a collaborative metadata management approach of this kind has to ensure low effort for knowledge contributors as in most cases these regional or divisional experts do not benefit from metadata quality themselves. Therefore, the paper at hand identifies requirements to be met by a business metadata repository, which is a tool that can effectively support collaborative management of business metadata. In addition, the paper presents the results of an evaluation of these requirements with business experts from various companies and of scenario tests with a wiki-based prototype at the company Bayer CropScience AG. The evaluation shows two things: First, collaboration is a success factor when it comes to establishing effective business metadata management and integrating metadata with enterprise systems, and second, semantic wikis are well suited to realizing business metadata repositories.

8.1 Introduction

8.1.1 Motivation and problem statement

A clear and unambiguous understanding of data is indispensable for the effective management of multidivisional companies. At the same time, clear and unambiguous definitions of business objects (e.g. materials, customers, suppliers) are imperative for implementing and benefiting from integrated, automated business processes (Kagermann, 2010 #249). Fundamentally, smooth and efficient business process execution is ensured by metadata describing the data structure of business objects (i.e. their attributes, and relations) and providing information for correct usage of business objects in business processes. High-quality metadata (i.e. metadata that is up to date, accurate and complete) helps to create and maintain a common understanding of business objects and business processes. It minimizes the occurrence of errors in automated activities and reduces waiting times between activities. The following three examples illustrate how poor metadata quality may negatively affect companies' business process performance.

DB Netz AG (40,300 employees, EUR 4.1 billion revenue in 2009), a subsidiary of Deutsche Bahn AG, is responsible for Germany's railway infrastructure. The company has to make sure that its entire infrastructure is accurately kept in an inventory database in order to be able to report to national authorities on an annual basis for the purposes of obtaining public funding. This task is impeded by the fact that there is no company-wide definition of basic business objects (i.e. tracks, switches, and tunnels). For example, while a tunnel in one geographical region is considered as an underground track segment, with the tunnel length being the distance between tunnel entrance and tunnel exit, in another region "tunnel length" refers to the distance between two underground stations. Moreover, crucial knowledge (e.g. on the objects' geographical position, maintenance, and monetary value) is distributed across the entire company. As a result of these problems, any consolidation of infrastructure data needed for inventorization is virtually impossible.

SBB Cargo AG (3,700 employees, CHF 915 million revenue in 2009), a subsidiary of Schweizerische Bundesbahnen, is responsible for Switzerland's railway freight transportation. The trains of SBB Cargo run on tracks that are used by a certain train type within a given period of time. To book a track, SBB Cargo only needs to provide the train type used and the period of time. No further attributes are necessary to specify the business object from a data exchange perspective. For internal purposes (e.g. specifying transportation orders, identifying dangerous goods to be transported, putting in place work schedules for the staff), additional attributes are relevant, the values of which are set by various information systems at various points in time. In order to be able to achieve flexibility in planning and short response times (e.g. in case of disturbance or malfunction), an unambiguous model of the business objects (e.g. track, train,

train conductor) is needed that is understood in every department of the company. Another important requirement relates to business object definitions of business partners, since during any one trip the trains of SBB Cargo may run on tracks of various providers in various countries, making it necessary to take into account different business object definitions when a track is booked.

Bayer CropScience AG (18,700 employees, EUR 6.5 billion revenue in 2009), a market leader in the field of agrochemical products (e.g. for plant protection, insect control, seed growing), is a business division of Bayer AG. Bayer CropScience is conducting a worldwide business process harmonization project, in the course of which master data is to be consolidated. Currently, the structure, meaning and usage of the company's master data vary from region to region. For example, materials are counted by weight in Europe, and by item (e.g. bottle, box) in other regions. In order to enable corporate analyses (e.g. spend analysis, demand planning), data structures and business metadata have to be harmonized as well.

8.1.2 Research objective and paper structure

A common feature of the cases described in Section 8.1.1 is that the knowledge needed to create consistent business metadata is spread across various departments, divisions or lines of business. One possible solution to this problem could be to collaboratively identify, record, discuss and consolidate knowledge on various and differing aspects of business objects (e.g. technical features, accounting attributes, regional characteristics in the case of DB Netz's "tunnel") with the help of business experts. To do so, the use of a collaboration tool following the concept of Wikipedia²⁸ (i.e. a software tool supporting an iterative and collaborative approach for building a repository that any user can edit) could facilitate the process of collaboratively creating and maintaining consistent business metadata.

A comparison of the cases described in Section 8.1.1 with Wikipedia indicates the potential to use Wikipedia as a model for solving the problems described in the cases. Wikipedia even describes concepts like "tunnel", "train" or "train conductor", and provides collaboratively maintained and commonly accepted definitions. If the maintenance of Wikipedia was restricted to employees of DB Netz for example, the provided description of "tunnel" could be used as an accepted, company-wide definition. The question is, however, whether collaboration effects brought about by software similar to the Wikipedia platform can also be accomplished in a business context. Therefore, the research objectives are a) to identify requirements to be met by a repository which supports the collaborative management of business metadata, b) to implement a wiki-based business metadata repository, and c) to evaluate its use in a business context.

²⁸ Wikipedia's English website: <http://en.wikipedia.org>

The next section of the paper defines business metadata according to literature on the subject. Furthermore, advantages and disadvantages of wikis to be used for creating business metadata repositories are discussed, and some scientific findings on the use of wikis in business contexts are presented. Section 8.3 specifies the research approach on which the paper is based and the techniques used for identifying and evaluating requirements to be met by a business metadata repository. Section 8.4 deals with the design of a business metadata repository, starting with the identification of business requirements, followed by the derivation of generic usage processes from these requirements, and finally leading to the implementation of a wiki-based prototype that meets the requirements and supports the usage processes. Section 8.5 presents the results of a requirements evaluation with business experts from various companies and users at Bayer CropScience using a wiki-based prototype. Finally, Section 8.6 presents a summary of the paper and gives recommendations for future research.

8.2 State of the art

8.2.1 Business metadata management

Metadata describes other data, i.e. it gives this data a certain meaning in a certain context. For example, a stick has a length (a physical property) of two meters (a measurement unit). Here, for the description of an object (the stick) a data item (the number “2”) and two metadata items (“length” and “meters”) are given [Sen 2004, p. 151]. Like the stick in this example, a business object is a real or virtual element in a company’s process of providing services or products (e.g. a purchase or production order, a contract, a business partner, a certain material) that is characterized by a number of properties represented by data elements [Martin 1975]. Metadata referring to a business object describe its characteristics, providing both technical aspects (e.g. data structure, data distribution) [Tozer 1999, pp. 184-194] and business-related aspects like information regarding correct use of the business object in certain business processes [Burnett et al. 1999, p. 1213, Marco 2000, pp. 49-52]. Table B8-2 gives an overview of the business objects, data attributes, and metadata mentioned in the examples given in Section 8.1.1, e.g. length of a tunnel in the case of DB Netz. In this context of metadata describing business objects, a business metadata repository is an information system that supports the collaborative management of business metadata and that aims at describing business objects comprehensively, unambiguously and understandably [Schmidt/Otto 2008, p. 215].

	Business objects	Data attributes	Metadata
DB Netz AG	Rails, tunnels, switches, etc.	Monetary value, geographical position, state of maintenance, etc.	Minimum requirements for specifications of geographical information, specification of measures, differentiation of business objects (e.g. rail vs.

	Business objects	Data attributes	Metadata
			switch), documentation of regional differences, etc.
SBB Cargo AG	Transportation orders, tracks, trains, etc.	Track reservation period, track price, train type, etc.	Documentation of different understandings of business objects by different business partners, documentation of synonyms and homonyms (e.g. train conductor, train driver), etc.
Bayer Crop-Science AG	Customers, suppliers, materials, etc.	Name, customer address, material density, material weight, packaging measures, etc.	Specification of unambiguous identifications of business partners, specifications for units of measure, documentation of regional differences (e.g. address data, billing data), etc.
Wikipedia	Animals, companies, history, people, technology, etc.	Categorization, denominators, date of birth, size, etc.	Documentation of historical facts, technical explanations, classifications of animals, people and plants, pictures, etc.

Table B8-2. Examples of business objects, data attributes, and metadata

8.2.2 Systems for collaborative metadata management

Due to the success of Wikipedia, companies increasingly express their desire to use wikis for business purposes [Bughin et al. 2008, p. 2, Bughin et al. 2009, p. 14]. As research has shown, the successful use of wikis in companies is basically possible [Majchrzak et al. 2006], mainly as effective tools for specification processes [Wagner/Majchrzak 2007]. However, the development of Wikipedia has shown that unstructured metadata, as created and used in a classical wiki, may lead to redundancies and inefficient usage of wiki content [Krötzsch et al. 2007].

In general, a wiki is a website providing content which can easily be edited by any (registered or anonymous) user [Cunningham 1995]. Depending on the software used, a wiki may come with different functionality. There are two functions, however, that can be found in practically any wiki [Kane/Fichman 2009, p. 2]: “Open editing” allows any user to change and modify content, and “edit preservation” serves for saving all edits and versions of the content. By means of a “rollback”, a previous version of certain wiki content can easily be reproduced (in order to prevent content vandalism, for example) [Viégas et al. 2004, pp. 578-580].

As mentioned above, metadata is largely unstructured in a classical wiki. While the structuring of wiki content by means of design guidelines and the linking of articles does support human users of a wiki (cf. Wikipedia), it does not provide any prerequisites that allow for efficient machine processing of content [Krötzsch et al. 2007]. It is this shortcoming that semantic wikis address, offering additional functionality for content structuring including enhanced support for navigation and content search [Schaf-

fert et al. 2008]. Semantic MediaWiki²⁹ (SMW) is a software that provides a semantic wiki [Krötzsch et al. 2007]. It is an extension of MediaWiki³⁰ (the wiki engine of Wikipedia) and is capable of processing annotated links between wiki articles as well as generating a machine-processable representation of such relations. In a semantic wiki that is used as a business metadata repository, for example, links between wiki articles on a certain business object, its attributes, or process activities using the business object could be processed. Based on the machine-processable representation of the links, a list containing all attributes of the business objects or a list containing all business objects used in a certain process activity could then be generated and used by other enterprise systems automatically. In a classical wiki, such lists need to be created and updated manually.

An experiment comparing a classical wiki and a semantic wiki with regard to usability and maintenance effort [Hüner/Otto 2009] has brought about two basic findings: On the one hand, users are able to find metadata more quickly, leading to higher satisfaction regarding usability of the system; on the other hand, the effort for managing and maintaining the content of a semantic wiki is considered to be quite substantial since the information from which the machine-processable link structures are generated requires additional attention. As Section 8.4 of this paper shows, however, enhanced usability is a success factor when it comes to fostering the willingness for collaboration across a company and collecting the knowledge and information needed to produce high-quality metadata.

8.3 Research approach

The results presented in this paper have been developed in the context of the Competence Center for Corporate Data Quality (CC CDQ). As part of the research program Business Engineering [Österle 1996], the CC CDQ is a consortium research project [Österle/Otto 2010] aiming at the design of artifacts (e.g. architectures, methods, reference models) which are expected to solve practical problems occurring in corporate data quality management. The project includes collaboration with subject matter experts from various companies, who assist both in the process of problem identification and in the process of evaluating the artifacts.

The research process applied follows the process model for Design Science Research proposed by Peffers et al. [Peffers et al. 2008, pp. 52-56]. Design Science Research is a research paradigm that aims at the design and evaluation of solutions to practical business problems specified by organizations [March/Smith 1995, p. 260, Hevner et al. 2004, p. 76]. Research in this context focuses on “means-end relations”, not on “cause-effect relations” [Winter 2008, p. 470], e.g. a business metadata repository (means) to support business metadata management in a company (end).

²⁹ SMW website: <http://www.semantic-mediawiki.org>

³⁰ MW website: <http://www.mediawiki.org>

This paper describes requirements to be met by a business metadata repository as a Design Science Research construct [March/Smith 1995, p. 256], the elements of which can be used for modeling further artifacts (e.g. generic processes for using the business metadata repository, software models). To validate the requirements identified, a wiki-based prototype was implemented and evaluated in two ways: a focus group interview [Cavana et al. 2001] was conducted to find out whether the prototype meets the requirements and allows for effective metadata management in organizations in general; and test scenarios at Bayer CropScience were put in place to see if the prototype is more capable of supporting metadata management than a business metadata repository already in use and considered inadequate.

8.4 Business metadata repository design

8.4.1 Requirements for effective business metadata management

The focus groups conducted in the course of the CC CDQ discussed numerous cases where metadata of low quality (i.e. inaccurate, inconsistent and/or incomplete) had caused business problems (cf. Section 8.1.1). It was agreed that business metadata management should be supported by the collaboration of as many business experts as possible from as many company divisions, departments, and units as possible. The collaboration of distributed experts and metadata users was identified as a promising alternative to centrally organized corporate metadata managers (e.g. within a central IT department) who are hardly able to provide comprehensive and timely information with regard to miscellaneous regional and divisional peculiarities.

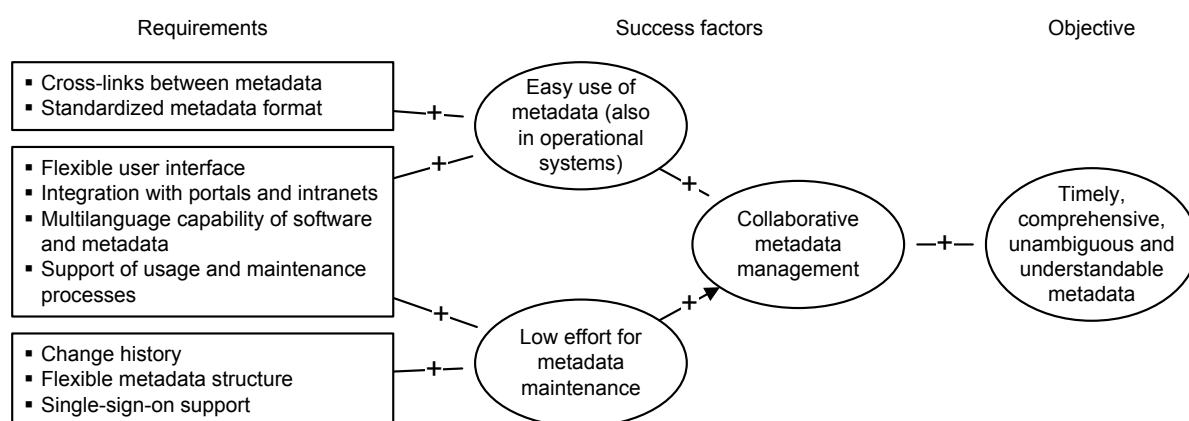


Figure B8-1. Requirements to be met by a business metadata repository

Besides the positive effect of collaboration on metadata quality (i.e. on comprehensiveness, timeliness, understandability and unambiguity), the focus group also discussed the opposite causality direction (cf. Figure B8-1). The participants agreed that the high quality of content as an outcome of a collaborative process might motivate individuals to increase their contribution if they see the contribution of others as a payoff for their own effort. Two success factors for achieving a high level of collaboration were identified: First, the effort for managing and maintaining metadata should

be kept rather low; and second, the tool should facilitate easy use of metadata in as many enterprise systems as possible (e.g. customer relationship management (CRM), enterprise resource planning (ERP) systems).

The outcome of a focus group interview in 2007 (cf. Appendix A for participants) is a list of requirements a business metadata repository should meet (cf. Table B8-3) in order to enable collaborative metadata management. Figure B8-1 shows the interrelations between the requirements to be met by a business metadata repository and the two success factors for effective metadata management mentioned above.

Requirement	Description
R01. Cross-links between metadata	In dictionaries, terms are usually defined with the help of cross-links to other terms, glossaries, or pictures. Likewise, the business metadata repository should contain such cross-links in order to avoid metadata redundancy and allow for efficient navigation.
R02. Standardized metadata format	The metadata kept in the business metadata repository should be highly visible, i.e. it should be available and used by as many enterprise systems as possible, thereby fostering the willingness to take part in the collaboration efforts among as many users as possible. A standardized metadata format would facilitate easy metadata import and export.
R03. Flexible user interface	The business metadata repository's user interface should offer options for flexible configuration (i.e. different roles for different jobs) in order to accomplish effective collaboration, including as many users, departments and divisions as possible.
R04. Integration with portals and intranets	A frequent constraint stipulated by a company's IT strategy is to reduce the amount of systems visible to users and to provide a single front end (e.g. a portal or intranet). For that reason it should be possible to integrate the business metadata repository's front end with existing information systems.
R05. Multilanguage capability of software and metadata	In order to accomplish effective collaboration (cf. R03), metadata must be available in different languages. Also, the business metadata repository's user interface should be multilanguage capable.
R06. Support of usage and maintenance processes	Usage and maintenance processes should be clearly specified and documented (e.g. searching metadata within the business metadata repository, workflows for updating metadata, or reporting defects) in order to make sure the business metadata repository is efficiently used.
R07. Change history	A metadata change history should be available at any time in order to be able to quickly identify modifications made, correct errors, and recover previous states.
R08. Flexible metadata structure	Over time, metadata is likely to require modification (regarding both meaning and structure), caused by e.g. the need to describe new data attributes or to implement additional enterprise systems that use the data. In order to ensure high metadata quality (i.e. up to date and accurate), it should be easy to make modifications.
R09. Single-sign-on support	The business metadata repository should be capable of integrating existing modes for authentication (i.e. no separate authentication should be required).

Table B8-3. Requirements to be met by a business metadata repository

8.4.2 Processes for collaborative management of business metadata

Bayer CropScience has been using a Lotus Notes-based business metadata repository. This repository contains metadata on customer, material and vendor attributes describing, among other things, the attributes' maintenance in the company's ERP system. In its current version, the repository is not used to the extent desired, leading to incomplete and outdated metadata. Table B8-4 lists the shortcomings of the business metadata repository used by Bayer CropScience as identified in interviews with six users of the system. The interviews were conducted using a questionnaire (see Appendix B) and were based on CommonKADS, which is a method supporting the development of knowledge intensive information systems [Schreiber et al. 2000, pp. 14-19]. The interview questions are structured in three parts following a top-down approach to identify general tasks the interviewees use the repository for (questions A1 – A4), to identify usage processes for the tasks identified (B1 – B8), and to address concrete functionality of the repository (user interface, different ways of using metadata, etc.) for each task identified (C1 – C4).

Shortcoming	Requirement	Prototype functionality
No introduction and documentation	R10. Intuitive user interface (cf. R03, R06)	Basic functionality from MediaWiki (e.g. full-text search) is complemented by automatically generated lists (SMW functionality). For clear navigation, metadata is clustered into three groups (i.e. sorted by date of last quality check).
	R11. User instructions and documentation in the wiki	MediaWiki and SMW offer comprehensive user instructions and documentation. Case-specific functions and configurations have not yet been documented.
Low quality of content	R12. Feedback functions (cf. R06)	Adding comments to articles is a standard feature of MediaWiki.
	R13. Easy identification of outdated metadata (cf. R06)	Requirement is met by metadata being clustered into three groups specifying the time they were last updated.
	R14. Easy updating function (cf. R06, R07, R08)	Data entry and editing is facilitated by special forms. Implementing a function for validation required an extension (about 40 lines of PHP code) in the case of Bayer CropScience.
No multilanguage capability	R15. Multilanguage capability of user interface (cf. R05)	The user interface of MediaWiki and SMW is available in a number of languages.
	R16. Support for structuring content available in various languages (cf. R05)	The language the content is presented in can be chosen easily. Cross-links connecting different language versions of the same article have not yet been implemented.
Poor navigation	R17. Clear navigation concept (cf. R03, R06)	Navigation functionality of the prototype is designed along use and management processes that are deduced from user interview results.
Poorly structured search results	R18. Intuitive search function (cf. R06)	The full-text search function provided by MediaWiki offers state-of-the-art search functionality. Special filters or sorting mechanisms have

Shortcoming	Requirement	Prototype functionality
		not yet been implemented.
	R19. Predefined filters (cf. R03)	The classification of metadata into three groups specifying the time they were last updated presents a simple predefined filter.
Unspecific denomination of metadata	R20. Unambiguous denomination	Metadata is clearly denominated by the article titles.
No integration with ERP system	R21. Integration with ERP system (cf. R02, R06)	SMW is capable of exporting annotated content in various formats (e.g. XML, CSV). Bayer CropScience is planning to take advantage of this option in order to be able to use the metadata of their wiki in their ERP system as well.
No documentation of interdependencies between fields	see R11	
	R22. Support of content-related cross-links (cf. R01)	Cross-links between articles are a basic functionality of MediaWiki.
No user-specific views	R23. Support of user-individual content and functions (cf. R03)	MediaWiki offers predefined wiki articles for registered users. Special templates for metadata managers are in preparation.
No cross-links between objects	see R22	

Table B8-4. Functionality of a wiki-based business metadata repository

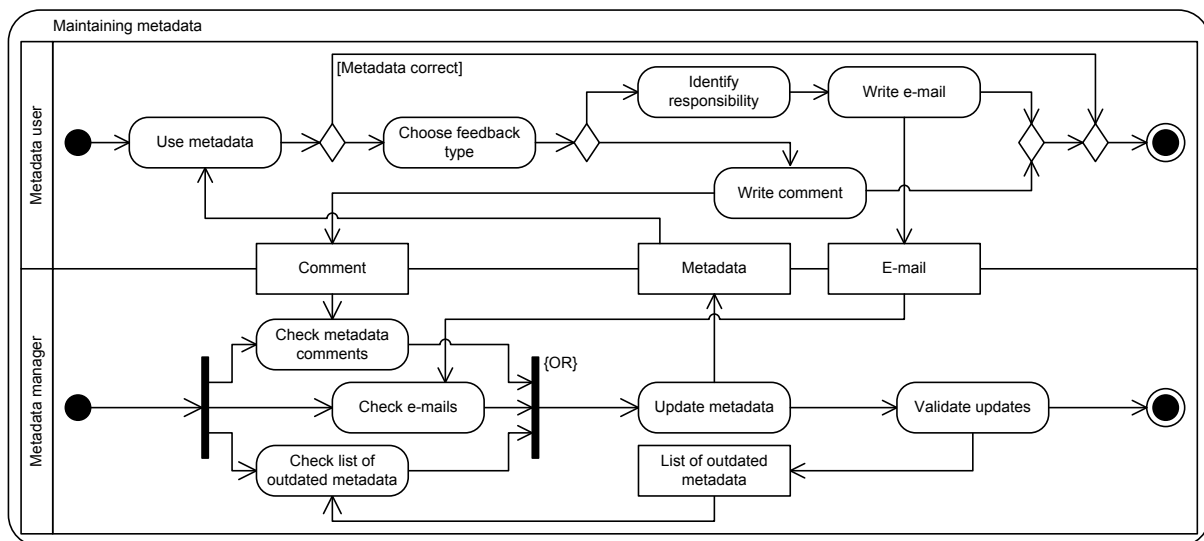


Figure B8-2. Process of maintaining business metadata

From the interview results, three generic usage processes for a new version of the business metadata repository were derived. One of these processes (i.e. for maintaining metadata) is shown in Figure B8-2. It addresses the requirements R10, R12 – R14, R17 – R20, R22 and R23 (and thus also R01, R03 and R06 – R08), and describes the reporting of a metadata defect by a metadata user and the validation of metadata by a metadata manager. A metadata defect can be reported either by adding a comment to the respective metadata item or simply by writing an e-mail to the respective metadata

manager. A metadata update is triggered either by a request for change (i.e. comment or e-mail) or by the metadata manager checking a list of outdated metadata.

The other two processes cover the use of metadata directly in an enterprise system, and the joint use of enterprise systems and a business metadata repository for extended metadata search. As the requirements to be met by the specific repository used at Bayer CropScience are largely identical with the general requirements a repository of this kind should meet (cf. Table B8-3), the usage processes developed can be seen as generic processes for a business metadata repository.

8.4.3 A wiki-based prototype

Taking into consideration the requirements and generic processes identified, a prototype was implemented and evaluated in two scenarios (cf. Section 8.5). This prototype is based on a semantic wiki which was implemented by means of SMW (cf. Section 8.2.2) and a number of SMW extensions (cf. Semantic Forms³¹ for the creation of forms for metadata maintenance). Table B8-4 describes how the prototype's functionality meets the requirements identified at Bayer CropScience. In addition, Figure B8-3 shows some aspects of the prototype's user interface.

The figure displays three screenshots of a web-based business metadata repository interface.
 Screenshot (a) shows the main page with a navigation menu on the left and three columns of metadata sets. The 'Up to date metadata' column lists items like 'LIFNR-KNVP' (Account number of vendor or creditor) and 'GRICD-LFB1' (Distribution Type). The 'Metadata for review' column lists items like 'DIBER-MARC' (Goods receipt times) and 'VRBWK-MARC' (Sales unit). The 'Out-dated metadata' column lists items like 'PARHS-KNVK' (Partner language) and 'WANID-KNVA' (Processing time).
 Screenshot (b) shows a detailed view of a metadata set 'STCEG-KNA1'. It includes a 'Definition' section with text about VAT registration numbers and partner data. A summary box shows 'Responsible: James Miller', 'Last validated: 2009/07/19', and 'Involved: Mary Forder, John Alberts'.
 Screenshot (c) shows the 'Edit Metadata: STCEG-KNA1' form. It has input fields for 'Responsible' (James Miller), 'Last validated' (19 July 2009), 'Involved' (John Alberts), and 'Short description' (Partner Data and VA...).

Figure B8-3. User interface of a business metadata repository

The business metadata repository's main page (see Figure B8-3.a) presents a list of metadata sets in the order in which they were most recently validated and updated by the respective metadata manager (R13). Alternatively, metadata can be found using a full-text search. Figure B8-3.b shows parts of a business metadata repository page including information on who is responsible for the metadata given. Each set of metadata is briefly described and the date of the last validation is shown. By following con-

³¹ Semantic Forms website: http://www.mediawiki.org/wiki/Extension:Semantic_Forms

tent-related cross-links a user is able to identify the person responsible for the article (R22). Validation dates may be colored, depending on the time that has passed since they were last updated (R13). A button labeled “validate” for indicating a content check is only displayed if the user currently logged in is responsible for the metadata item given (R14).

The business metadata repository is configured so that only registered users (i.e. metadata managers) are allowed to edit metadata, while comments can be added by any user (also anonymously). The respective manager is shown the metadata he is involved with or responsible for in predefined lists (R23). Figure B8-3.c shows a form for metadata maintenance that was created by means of Semantic Forms. This extension is capable of analyzing the machine-processable wiki content in order to generate lists supporting data entry and editing, for example.

8.5 Artifact evaluation

8.5.1 Focus group evaluation results

The prototype was evaluated by subject matter experts in a focus group in 2008 (cf. Appendix A for list of participants). In terms of the general requirements identified (cf. Table B8-3), this evaluation produced the following results:

- R01, R03, R07 and R08 were considered as being met by the prototype. Consequently, these requirements were not discussed any further.
- R02 (standardized metadata format) was discussed very extensively. The possibility to integrate business metadata repository content by means of metadata export (a standard function offered by SMW) was greatly appreciated. Bayer CropScience is planning to use this function for integrating the business metadata repository’s content into their central ERP system.
- R04 (easy integration with portals or intranets) was considered as not being met by the prototype. However, it was agreed that this feature could easily be added, as the prototype has been set up as a standard website.
- R05 (multilanguage capability) was considered as being met by the prototype. Multi-language capability of the software is ensured by MediaWiki and SMW both supporting a number of languages, with further languages being easily added, if need be. Multilanguage capability of metadata can easily be achieved by any wiki, either by composing articles in various languages or by linking articles written in different languages with one another.
- R06 (support of usage and maintenance processes) was considered as being partially met by the prototype through the possibility to manage and maintain

metadata by means of structured forms (see Figure B8-3.c). Simple workflows can be supported by sequences of forms.

- R09 (single-sign-on support) was considered as not being met by the prototype, as the prototype requires separate authentication. However, it was agreed that the prototype could easily be extended (e.g. extensions to integrate LDAP directories are available).

The subject matter experts agreed that in view of its intuitive functionality (cf. easy correction of errors and easy adding of comments) a wiki-based business metadata repository is very likely to lead to increased collaboration. They also agreed that semantic wiki features are well suited for supporting non-technical business users (cf. forms for editing metadata and enhanced navigation for retrieving metadata). However, technical details of a semantic wiki (e.g. semantic annotations and semantic queries) should be hidden from users in order to prevent an impediment to the positive collaboration effect of using a wiki, especially in the case of business users.

8.5.2 Results of prototype testing at Bayer CropScience

The prototype was also evaluated in the course of functional test scenarios at Bayer Crop-Science. Following the generic processes developed (cf. Section 8.4.2), a number of tasks were specified. Based on the scenario shown in Figure B8-2, the test users, who had not been trained to work with the prototype, were asked to

- search for a certain wiki article,
- validate the metadata given in this article,
- report potential metadata defects to the persons responsible by adding comments or by sending an e-mail,
- correct the metadata defects identified by the other test users,
- validate the modifications made.

Following the execution of the tasks specified, the test users were asked in a final interview to

- judge the prototype's usability,
- assess whether the prototype meets the identified requirements,
- give their opinion on whether more requirements to be met by a business metadata repository could be identified.

All test users were able to execute all the tasks specified. Also, all test users confirmed that working with the prototype was no problem for them and that the wiki offered very good assistance in performing the tasks successfully. As all test users were familiar with the company's regular business metadata repository, they were able to assess the prototype's functionality in comparison with the old system. Aspects such as easy navigation, flexible and well-structured lists, full-text search, and consistent cross-linking of metadata were considered as particularly positive. Only one negative state-

ment was made with regard to the low usability of linking external files to metadata in the repository.

8.6 Conclusion and future research

The paper identifies requirements to be met by a business metadata repository for supporting the collaborative management of business metadata, specifies generic processes for such a repository, describes the implementation of a wiki-based prototype, and presents the results of an evaluation of the requirements and the prototype. The prototype analysis (see Table B8-4) shows that most requirements can be met by standard wiki functionality (e.g. feedback by means of comments, rollbacks, easy-to-use content updating, and cross-links) and that semantic wiki features are able to meet remaining requirements (e.g. content structuring, metadata export, and predefined content filters).

Further research should deal with the identification and development of detailed functional specifications of a business metadata repository, ensuring in particular metadata integration with enterprise systems. In this respect, the possibility to export metadata in a standardized format could provide the functional basis for using metadata directly in an enterprise system. Empirical studies should investigate whether using a business metadata repository that meets the requirements specified actually leads to the level of collaboration expected and desired (i.e. involving more or all of a company's departments, divisions or lines of business in the process of collaborative business metadata management) and whether metadata quality can be significantly improved over time.

With regard to the case of Bayer CropScience presented in this paper, further research should examine whether the scope of business metadata repository users can really be extended in the long run, with more departments, divisions and lines of business of the company accepting and supporting the concept of collaborative business metadata management. Furthermore, it should be examined whether a wiki – as an “open” approach – is capable of improving metadata quality substantially and with a sustained effect, or if editing rights should possibly be limited to a small number of people. Particular attention should be paid to the aspect of integrating metadata with the company's enterprise systems. In view of the planned use of business metadata repositories in companies currently collaborating with the CC CDQ, it will be possible to compare the analysis of similar cases in future research.

9 Toward a functional reference model for master data quality management

Titel	Toward a functional reference model for master data quality management ³²
Autoren	Boris Otto; Kai M. Hüner, Hubert Österle Universität St. Gallen, Institut für Wirtschaftsinformatik Müller-Friedberg-Strasse 8, 9000 St. Gallen, Schweiz {boris.otto kai.huener hubert.oesterle}@unisg.ch
Publikationsorgan	Information Systems and E-Business Management, doi: 10.1007/s10257-011-0178-0
Status	Veröffentlicht

Tabelle B9-1: Bibliographische Angaben zum Beitrag „Toward a functional reference model for master data quality management“

Abstract

The quality of master data has become an issue of increasing prominence in companies. One reason for that is the growing number of regulatory and legal provisions companies need to comply with. Another reason is the growing importance of information systems supporting decision-making, requiring master data that is up-to-date, accurate and complete. While improving and maintaining master data quality is an organizational task that cannot be encountered by simply implementing a suitable software system, system support is mandatory in order to be able to meet challenges efficiently and make for good results. This paper describes the design process toward a functional reference model for master data quality management (MDQM). The model design process spanned several iterations comprising multiple design and evaluation cycles, including the model's application in a participative case study at consumer goods manufacturer Beiersdorf. Practitioners may use the reference model as an instrument for the analysis, design and implementation of a company's MDQM system landscape. Moreover, the reference model facilitates evaluation of software systems and supports company-internal and external communication. From a scientific perspective, the reference model is a design artifact; hence it represents a theory for designing information systems in the area of MDQM.

9.1 Introduction

9.1.1 Motivation and problem statement

Master data describes the essential business entities of a company, such as suppliers, customers, products, employees, and assets [Dreibelbis et al. 2008, Loshin 2008, Smith/McKeen 2008]. The quality of master data has become an issue of increased prominence in companies. Reasons for that are an increasing number of regulatory and

³² © Springer-Verlag 2011

legal provisions companies need to comply with, the need to have a *single source of truth* for company-wide reporting [Yen 2004, McCann 2010], and the demand for having a *360°-view on the customer* [Pula et al. 2003, Leser/Naumann 2007].

An example for the latter is the shift from long-term fixed-line contracts toward a consumer-centric, information services providing business model in the telecommunications industry. Telecommunications providers are forced to manage large amounts of ordering, complaints, contract, service quality, and billing data around individual customer information, which requires significant changes to the way lifecycles and architectures of customer master data are designed today. Consulting company Deloitte summarizes the importance of master data: “Data ascends from the basement to the board room.” [Deloitte 2009]

To meet such requirements, master data of good quality (i.e. master data that is up-to-date, consistent and complete) is needed. Companies respond to these requirements by establishing data governance [Weber et al. 2009, Kathri/Brown 2010] or master data management (MDM) [Loshin 2008, Smith/McKeen 2008]. Of course, improving and maintaining master data quality is an organizational task that cannot be encountered by simply implementing a software system [English 1999, Kathri/Brown 2010]. However, appropriate system support is mandatory in order to be able to meet challenges regarding master data quality efficiently and make for good results.

The consolidation of the market for data quality management (DQM) and MDM systems that is currently under way has been aggravating decision-making regarding the choice of the right product. IBM acquiring Ascential [IBM 2005] and SAP acquiring BusinessObjects [SAP 2007b] are two examples of this development. Integration of the products acquired with the existing product portfolio in order to be able to offer an integrated solution for quality oriented MDM (in the following master data quality management, MDQM) poses new challenges not just for vendors of MDQM systems but also for user companies:

- *System evaluation.* What basic functionality needs to be provided by an MDQM system and which system fulfills best the imposed business requirements?
- *System landscape analysis and design.* Which MDQM systems are currently providing the required functionality, and which ones should do so in the future?
- *System landscape transformation.* What is a reasonable approach and schedule towards the future MDQM system landscape?
- *Communication.* What terminology is to be used for company-internal and external communication (e.g. with MDQM system vendors, system integrators, other companies)?

Existing literature hardly delivers any answers to these questions. While there are studies on both MDM [Loshin 2008, Smith/McKeen 2008] and DQM [Wang 1998, Batin/Scannapieco 2006, Lee et al. 2006], an integrated analysis cannot be found. Sources addressing practitioners are numerous with regard to the general significance of master data quality [White et al. 2006, Radcliffe/White 2009], recommendations for doing successful MDM [Dreibelbis et al. 2008, Loshin 2008], and comparison of system vendors [Karel 2006, Kokemüller 2009]. General statements, however, on the functionality of MDQM systems are missing.

9.1.2 Research question and contribution

Against this background, the article puts up the following research question: What functionality is to be provided by a system supposed to improve and maintain the quality of master data? To answer this question, the paper follows the principles of Design Science Research (DSR) [Hevner et al. 2004, Peffers et al. 2008] in order to develop and evaluate a functional reference model for MDQM.

In general, a reference model is an information model that can be used not just in one specific context but in various contexts [Becker et al. 2004, Fettke/Loos 2007]. For a specific class of companies, a reference model claims to be generally applicable and to serve as a predefined pattern to cope with practical problems [Schütte 1998, Rosemann/Schütte 1999]. The design of reference models has a long tradition in the Information Systems (IS) research community, particularly in Europe [Frank et al. 2008, p. 391].

From an epistemological perspective, the MDQM reference model is an artifact and, thus, the result of design oriented research [Nunamaker et al. 1991, March/Smith 1995]. DSR aims at designing artifacts according to scientific principles, in order to be able to solve practical problems [March/Smith 1995, Hevner et al. 2004]. Hence, the reference model contributes both to the scientific and the practical state of the art:

- *Contribution to practical state of the art.* The reference model supports companies trying to overcome the challenges listed in the previous section. It helps create a common terminology both for internal and external communication. Furthermore, it offers an instrument for evaluating existing and identifying required MDQM functionality, and it allows comparing of different commercial software systems on a vendor-independent basis.
- *Contribution to scientific state of the art.* The description of the design process and of concrete design decisions allows scientific validation of the artifact presented as well as its extension by aspects previously not sufficiently considered or differentiated. Furthermore, the reference model represents an abstraction of an information system in the field of MDQM. Hence, it forms a “theory for designing” [Gregor 2006].

The remainder of the paper starts with an introduction to DQM and MDM as scientific areas of interest, followed by a brief discussion on related studies. After that, the paper outlines the design process, introduces the reference model, and explains concrete design decisions with regard to one specific function group, namely Cross Functions. For the purpose of demonstration the paper then describes how the reference model was applied at consumer goods manufacturer Beiersdorf in the context of a participative case study [Baskerville 1997]. After that, the results of the case study and the design process are evaluated against a number of dimensions for evaluation [Frank 2007]. The paper concludes with a short summary and recommendations for further research on the topic.

9.2 Related work

9.2.1 Data quality management

Data quality has been the subject of a number of research studies, some of which come up with lists and categories of data quality dimensions as the result of empirical research [Wang/Strong 1996], while others present practitioners' experiences [Redman 1996, English 1999, Loshin 2001] or draw theoretical conclusions [Wand/Wang 1996, Price/Shanks 2005]. Despite all differences, there is consensus in the specification of the term data quality as the data's fitness for use, i.e. evaluating the quality of data depends on the context the data is used in and the user the data is used by.

DQM comprises initiatives for improving the quality of data [Batini/Scannapieco 2006]. In doing so, DQM goes beyond mere reactive initiatives (i.e. identification and fixing of data defects) [Shankaranarayanan/Cai 2006]. Rather, DQM aims at carrying out preventive initiatives in order to be able to ensure data quality by means of an iterative process including steps to define, measure, analyze and improve data quality as well as to design appropriate framework conditions for DQM [Wang 1998, English 1999, Eppler/Helfert 2004]. An overview of various DQM approaches is given by Batini et al. [2009].

9.2.2 Master data management

Master data specifies the essential business entities a company's business activities are based on. Such entities are, for example, business partners (customers, suppliers), products, or employees [Smith/McKeen 2008]. Basically, master data can be differentiated by three concepts: master data class, master data attribute, and master data object [Loshin 2008]. A master data object represents a concrete business object (an automobile manufactured in a certain plant at a certain point in time, for example), and it specifies selected characteristics of this business object (color, features, or price, for example) by means of attributes. Attributes selected for representation of a specific class of business objects (customers or products, for example) constitute a master data

class (which usually is specified by a data model). From the perspective of data modeling, a master data object is an instance of a master data class, which is created by assigning data values (a sequence of figures, for example) to a master data attribute (to the attribute phone number of the data class representing customers, for example).

MDM comprises all activities for creating, modifying or deleting a master data class, a master data attribute, or a master data object [White et al. 2006, Smith/McKeen 2008], i.e. the modeling, provision, quality management, maintenance, and archiving of master data. All these activities aim at providing master data of good quality (i.e. master data that is complete, accurate, timely, and well-structured) for being used in business processes [Karel 2006, Loshin 2008].

9.2.3 MDM and DQM

The fact that data quality has been specified as one objective of MDM has led to DQM being considered in a number of studies as only one of several functions of MDM [Dreibelbis et al. 2008, DAMA 2009]. Taking this view, analysis of DQM application systems is limited to reactive DQM only [White/Radcliffe 2008]. Looking at both areas in a differentiated manner, however, it can be seen that there are activities in both areas mutually affecting each other, so that any subordination of one area to the other is inappropriate. Particularly preventive DQM activities [data governance, for example [Weber et al. 2009, Kathri/Brown 2010] or business metadata management [Burnett et al. 1999, Marco 2000]] help structure a company's MDM as they specify master data or define responsibilities for data maintenance. Hence, preventive DQM does have an effect on the design of a company's MDM. Taking this into consideration the reference model presented in this paper does not aim at ranking the two concepts of MDM and DQM against each other. Rather, it specifies business user requirements on the functionality of an application system supporting both DQM and MDM activities in order to be able to provide master data of good quality for being used in business processes.

9.3 Research approach

9.3.1 Overview

A reference model is defined as the result of a design process performed by a modeler who specifies (at a certain point in time) general elements of a system so that it serves as a reference point in the design of an information system. A reference model represents a class of use cases and can be used for the development of company-specific models [Schütte 1998, pp. 69-74]. Reference models can be distinguished by application domain (e.g. retail industry, MDQM), modeling language (e.g. Entity-Relationship Model, Unified Modeling Language), size, design process, and evaluation strategy [Fettke/Loos 2004].

The paper at hand describes the model design process and the resulting reference model for MDQM functionality. According to the ARIS concept [org. Architektur Integrierter Informationssysteme (German), Architecture of Integrated Information Systems], which distinguishes between four descriptive views (functional, data, control, and organizational view) and three descriptive layers (business design, technical design, and implementation layer) [Scheer 1992, Scheer 1997, Scheer et al. 2005], the functional reference model for MDQM represents the business design of a functional view on MDQM systems. It is therefore purpose-oriented and task-oriented (rather than data-oriented, for example), while it disregards technical and implementation aspects.

The work presented in the paper follows in general the guidelines for DSR proposed by Hevner et al. [2004]. In particular, the design process is based on the principles of the Design Science Research Methodology (DSRM) [Peppers et al. 2008], suggesting a sequential design process comprising iterations of design and evaluation cycles [Simon 1998, Hevner et al. 2004] as well as various points of entry. The model design process followed in this research represents a problem-centered initiation of the DSRM process model, as the need for doing research on the topic was identified by focus groups (see Section 9.3.2). The need for doing research on the topic was neither announced in advance, nor did it result from reported shortcomings of an existing artifact.

The research context is formed by the (research program blinded for review, in the following referred to as the Research Program). Since 2006, the (research organization blinded for review), together with a number of partner companies, has been developing design artifacts in the field of corporate data management.

9.3.2 Research process

As proposed by the DSRM process model, the design of the reference model was carried out in six steps.

The *first step*, which was carried out between January and December 2008, aimed at identifying the problem and motivating the research. As outlined in the Section 9.1, the research described in this paper was mainly motivated by the identification of a number of challenges in the practitioners' community. In 2008 the market for MDQM was characterized by major consolidation activities. IBM, for example, acquired Ascential [IBM 2005], and SAP bought Business Objects [SAP 2007b]. In parallel to this ongoing market consolidation, practitioners who participated in the Research Program continuously articulated the demand for support with regard to the four challenges mentioned above. Managerial publications supported the observed results. Business analyst company Gartner, for example, documented "common questions on data integration

and data quality” from their 2008 MDM summit. Exemplary questions were [Friedman 2009]:

- “What is the relative strength of MDM solution providers with regard to data quality functionality? Are they as competent as the leading best-of-breed data quality tool vendors?”
- “Has market consolidation already reached the point where the advantages of single-vendor stacks for MDM (including data integration and data quality tools) outweigh the advantages of a best-of-breed strategy?”

The *second step* in the research process was about the definition of the objectives of the solution. The objectives of the research resulted from the identification of the practical challenges mentioned above and the realization that the existing knowledge base was not able to deliver appropriate responses to these challenges (see *Introduction*). This gap demanded for a “technique of practical utility” – which is one perception of reference models according to Fettke and Loos [2004].

Moreover, the reference model has to comply with the *Guidelines for Orderly Modeling (GOM)*, i.e. correctness, relevance, economic viability, clarity, comparability, and systematical design [Becker et al. 1995]. The GOM have been used for evaluation of reference models and have been specialized for that purpose [Schütte 1998]. Summarizing, the proposed solution is a reference model which on the one hand provides a solution for overcoming the challenges mentioned above and on the other hand is designed according to the GOM. The proposed solution was confirmed within a focus group interview on December 3, 2008 (focus group B, see Appendix H).

The *third step* comprised the design activities which followed the general principles of reference modeling [Schütte 1998, Becker et al. 2002, Fettke/Loos 2004]. The theoretical foundations of the model’s design itself are outlined below in the section 9.4.

The design process was carried out in three iterations. The first version of the reference model was built on the basis of an integrated state-of-the-art analysis. The analysis included both a literature review and a functional analysis of selected software systems in the MDQM domain, namely from IBM, Oracle, and SAP (see Table B9-2). The system analysis follows a recommendation by Hevner [2004] to look into existing instantiations when analyzing the state of the art. The relevant criterion for the selection of systems was that each system had to offer broad MDQM functionality, because the first design iteration aimed at identifying as many functions as possible in order not to neglect significant aspects.

Sources of Knowledge	Analyzed Items
Books	[Dreibelbis et al. 2008], [English 1999], [Heilig et al. 2006]
Industry Associations	[DAMA 2009]
Functionality of software systems	IBM [IBM 2006], Oracle [BEA 2006, Butler 2007], SAP [SAP

Sources of Knowledge	Analyzed Items
	2008], TIBCO
Reports from analysts, consultancies etc.	[White et al. 2006], [White 2008], [White/Radcliffe 2008]

Table B9-2: Knowledge sources for 1st design iteration

The second and the third design iteration were based on the results from focus groups A (2nd version) and C and D (3rd version). All three focus groups were used for demonstration and evaluation purposes (see below). Focus groups are a useful research method when researchers want to learn about the degree of consensus on a topic [Morgan/Krueger 1993, p. 17]. Therefore, they fit well the purpose of a reference model, which aims to represent common understanding of a topic. A similar use of focus groups in the process of reference model design can be found in the work of Sinha et al. [2004] and Schelp and Winter [2007].

The *fourth step* of the design process aimed at demonstrating the applicability of the reference model. Three different demonstration activities were used:

- Focus groups A, C, and D (see Appendix H for details);
- Participative case study at Beiersdorf (see Section 9.5 and Appendix H for details);
- Applicability check with three software vendors, namely IBM, SAP, and TIBCO.

In total, 31 persons participated in the three focus groups. 55 percent of them represented user companies, 39 percent consultancies, and 6 percent software companies. This means that all relevant groups of the MDQM community were involved in the process. Moreover, the third version of the reference model was reflected against the expertise of three software vendors (IBM, SAP, TIBCO) in order not to neglect of existing instantiations and also not to include functionality which was deemed “technically infeasible” by software providers.

Apart from that, the reference model was applied in a “real life” context during the participative case study [cf. Baskerville 1997] at consumer goods manufacturer Beiersdorf. The case study was carried out between February and October 2010.

In the *fifth step* the reference model was evaluated. Activities included:

- Focus groups A, C, and D;
- Participative case study at Beiersdorf;
- Multi-perspective evaluation according to the guidelines proposed by Frank [2007] (see Section 9.6);

- Compliance with GOM.

Due to the active role the researchers took in the process of adapting the reference model in the course of the participative case study, additional evaluation activities were required. Therefore, the paper uses both the evaluation criteria for reference modeling proposed by Frank [2007] and the GOM [Schütte 1998] for a more inter-subjective artifact evaluation.

The *sixth step* included communication activities. Both Hevner et al. [2004] and Pefers et al. [2008] stipulate that DSR results must be disseminated both in the practitioners' and the scientific community. While the former is addressed by a working paper and a managerial publication (which includes a model overview only), the paper at hand aims at making the research available for the scientific body of knowledge. First, it describes the reference model itself so that it can be used, extended, and evaluated by future research. Second, the paper outlines the research process to make it verifiable and repeatable for other researchers.

Figure B9-1 summarizes the six steps of the research process.

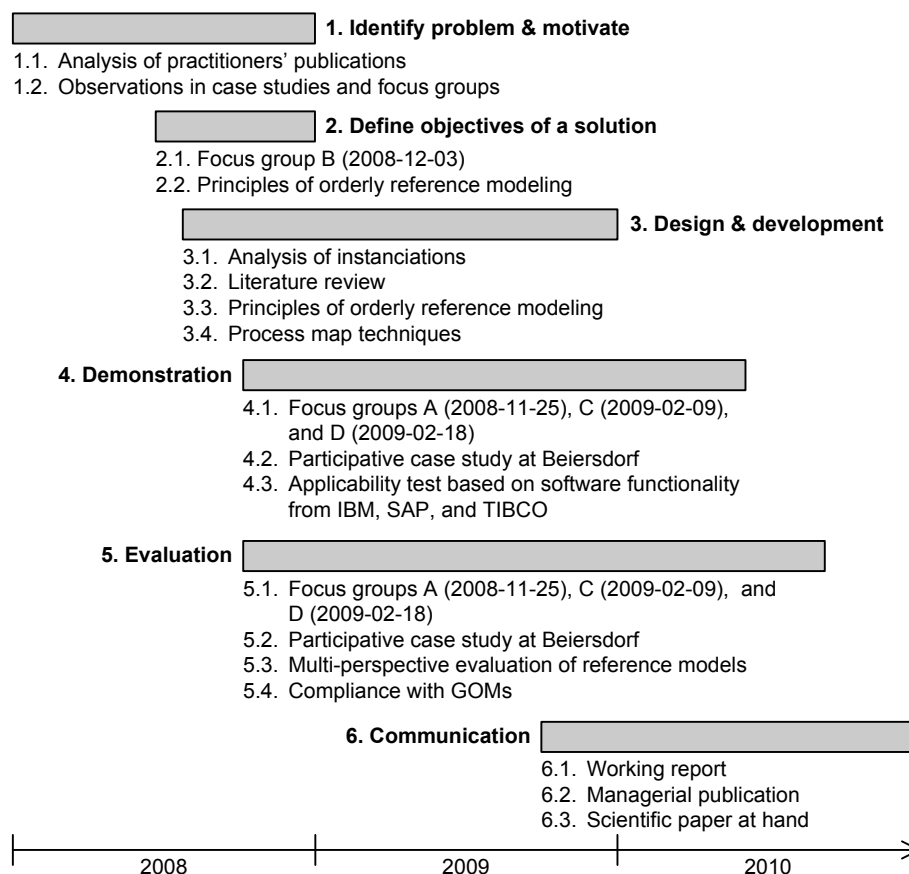


Figure B9-1: Research process overview

9.4 Reference model design

9.4.1 Design foundations

The design of the functional reference model for MDQM follows the ARIS conventions for the functional view of information systems [Scheer 2001, pp. 21-38], according to which functions should be hierarchically structured. The reference model presented in this paper consists of a three-level structure formed by function groups, functions, and sub-functions (see Figure B9-2).

Functions of the reference model for MDQM are grouped in function groups. A function group consists of one or more functions, whereas each function is assigned to only one function group. Functions themselves consist of sub-functions. The cardinalities between functions and sub-functions are analogous to those between function groups and functions. The use of three hierarchical levels conforms with the modeling propositions outlined in ARIS [Scheer 2001, p. 25].

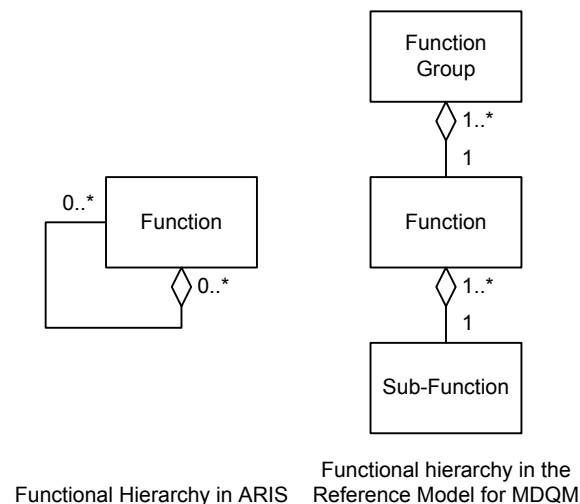


Figure B9-2: Modeling functional hierarchies

In general, there are three criteria to form functional hierarchies, namely performance, object, and process. The reference model for MDQM uses process criteria, grouping functions in function groups and sub-functions in functions, based on their purpose-oriented and task-oriented relationships.

The visual representation of the reference model follows the principles of process maps, which in general aim at identifying and representing similar processes, subprocesses and functions [Heinrich et al. 2009] in a tabular form. This concept is also widely used in the practitioners' community, for example in SAP's business maps. In particular, technology-related business maps (e.g. for SAP NetWeaver) use a tabular design [SAP 2007a]. The tabular presentation format is reused in the present paper because it was deemed appropriate to ensure high comprehensibility and acceptance by potential users of the reference model.

9.4.2 Model overview

In total, the reference model comprises 6 function groups, 19 functions, and 72 sub-functions. Figure B9-3 shows the function group and the function level of the model (all 72 sub-functions are explained in Appendix H).

Master Data Lifecycle Management	Data Creation	Data Maintenance	Data Deactivation	Data Archiving
Metadata Management and Master Data Modeling	Data Modeling	Model Analysis	Metadata Management	
Data Quality Assurance	Data Analysis	Data Enrichment	Data Cleansing	
Master Data Integration	Data Import	Data Transformation	Data Export	
Cross Functions	Automation	Reports	Search	Workflow Management
Administration	Data History Management	User Management		

Figure B9-3: Function groups (left column) and associated functions

The following listing describes the six function groups.

- *Master Data Lifecycle Management*. A master data object's lifecycle starts with its creation during business operations and ends with its deactivation and/or archiving [Redman 1996, Loshin 2001]. *Master Data Lifecycle Management* describes all activities data users or data managers do with master data during its entire lifespan [Reid et al. 2007]. Self-explaining functions, such as Create or Update, are not stated explicitly.
- *Metadata Management and Master Data Modeling*. Basically, metadata specifies data properties and the meaning of data (of master data, for example). In doing so, metadata both specifies data structures and – in the form of unambiguous specifications – ensures correct usage of data throughout an organization [Burnett et al. 1999, Tozer 1999, Marco 2000]. From an MDM perspective, metadata comprises all the information necessary for efficient management and effective usage of master data. According to the definition above, master data modeling means creation of technical metadata (data types, relationship multiplicities, for example).
- *Data Quality Assurance*. This function group comprises functions for preventive (i.e. to prevent potential, future data defects) and reactive (i.e. to repair data defects that have occurred) maintenance and improvement of the quality of master data. The three functions comprise sub-functions for identification of data defects and measuring data quality (*Data Analysis*), for improving data

quality by comparison with and integration of external reference data (*Data Enrichment*) and for repair of data defects identified (*Data Cleansing*).

- *Master Data Integration*. This function group comprises functions supporting transfer (import and export) and structural transformation (e.g. consolidation of fields or tables) of master data.
- *Cross Functions*. This function group comprises functions that cannot be assigned to one of the other groups. Sub-functions under the function *Automation* do not provide additional functionality but offer support for being able to efficiently use other functions by making them machine processable.
- *Administration*. This function group comprises functions for user administration and the tracing of changes and modifications made.

9.4.3 Functional details

Figure B9-4 displays the four functions and the fifteen sub-functions of the function group named Cross Functions. As the definition and structural design of this function group was a frequent issue during the discussions in the focus group interviews and the application of the reference model, Table B9-3 outlines design decisions made in the course of several design iterations, leading to the structural design of the function group as shown in Figure B9-4 (see Appendix H for the description of the sub-functions).

Cross Functions	Automation	Reports	Search	Workflow Management
	Automation	Reports	Search	Workflow Management
	Automated Enrichment	Data Quality Reports	Dynamic Value Search	Bundling of Activities
	Automated Export	Usage Statistics	Free Search	Graphical Workflow Modeling
	Automated Import	Job Monitoring	Fuzzy Search	Create/Maintain Workflows
	Cross-Function Automation	Audit Support		
	Push and Pull Mechanisms			

Figure B9-4: Functions and sub-functions of function group Cross Functions

Design decision	Justification
No explication of cross-cutting functions in the initial version of the reference model.	<ul style="list-style-type: none"> • Simplicity (principle “<i>as simple as possible</i>”). • Analysis of practitioners’ publications (see Table B9-2, Figure B9-1, process step 1.1) • Observations in case studies and focus group discussions (see Figure B9-1, 1.2).
Add functions for analysis and	<ul style="list-style-type: none"> • Analysis of practitioners’ publications [English 1999, Bati-

Design decision	Justification
reporting (e.g. <i>Search, Profiling, Reporting</i>).	<p>ni/Scannapieco 2006, Alur et al. 2007, Dreibelbis et al. 2008].</p> <ul style="list-style-type: none"> • Verified by reverse engineering analysis (see Figure B9-1, 4.3, [Otto/Hüner 2009]).
Add automation functions (e.g. <i>Automated Export / Import, Automated Enrichment</i>).	<ul style="list-style-type: none"> • Analysis of instantiations (see Figure B9-1, 1.2). • Verified by reverse engineering analysis (see Figure B9-14.3, [Otto/Hüner 2009]).
Explication of cross-cutting functions (i.e. functions that support more than one function from different function groups).	<ul style="list-style-type: none"> • Focus group A (see Figure B9-1, 4.1) suggested to group cross-cutting functions (i.e. <i>Automation, Reports, Search</i>) analogous to Corporate Services of SAP's ERP Solution Map [SAP 2011]. • Allows differentiation of end user functionality and supporting functionality (cross functions). • Verified by focus groups C and D (see Figure B9-1, 5.1). • Verified by reverse engineering analysis (see Figure B9-1, 4.3, [Otto/Hüner 2009]).
Assign <i>Data Quality Reports to Cross Functions</i> (previously under <i>Data Quality Assurance</i>), with <i>Graphical Analysis</i> remaining under Data Analysis.	<ul style="list-style-type: none"> • Focus group A decided to leave sub-function Graphical Analysis under <i>Data Quality Assurance</i> due to its close relation with <i>Profiling</i> (for example, when searching for duplicates, identification of potential duplicates is the result of profiling and supports manual analysis in the process of duplicate recognition). • Verified by focus groups C and D (see Figure B9-1, 5.1).
Add <i>Workflow Management to Cross Functions</i> .	<ul style="list-style-type: none"> • Focus group B discussed organizational challenges in the context of MDQM and raised the requirement of supporting workflows for collaborative data creation and maintenance. • Verified by focus groups C and D (see Figure B9-1, 5.1). • Analysis of instantiations (see Figure B9-1, 1.2) did not support this decision (regarding MDQM offerings), but each application system provider offers additional components for workflow management, and focus group members confirmed that these components were required for MDMQ.
Ignore function <i>Automation</i> for as-is functionality ratings by business users (see Section 9.1, use case <i>System landscape analysis and design</i>).	<ul style="list-style-type: none"> • Business experts (see below, model application at Beiersdorf) had difficulties in distinguishing between some <i>Automation</i> sub-functions (i.e. <i>Automated Import</i>) and automated sub-functions (e.g. <i>Delta Import</i>) from other function groups.

Table B9-3: Examples of design decisions and justifications regarding the function group *Cross Functions* (sorted chronologically)

9.5 Reference model application

9.5.1 Case study overview

A participative case study [Baskerville 1997] was used to demonstrate the reference model's applicability in a "real life" context. A participative approach allows for evaluation of the applicability of the model. The case study was conducted at consumer goods manufacturer Beiersdorf AG, headquartered in Hamburg, Germany. Operating in two business segments, namely Consumer (including brands as Nivea) and Medical (including brands like Hansaplast), Beiersdorf is a global company, generates an an-

nual revenue of approx. EUR 5.7 billion in 150 subsidiaries, and employs about 20,000 people all over the world [Beiersdorf 2009].

The master data class investigated in the participative case study is product data, e.g. identifiers, logistics data, and product hierarchies [Saaksvuori/Immonen 2008]. For the management of global product data Beiersdorf is using a central product lifecycle management (PLM) system, which was implemented in 2004. The PLM system at regular intervals (i.e. every three hours) provides new or modified product data to five regional enterprise resource planning (ERP) systems and a number of other global information systems [e.g. a decision support system (BW), a planning system (APO), and a procurement system (EBP)]. As the data is directly committed into the receiving systems, a consistent database is always ensured [Schemm 2008]. Figure B9-5 illustrates the system landscape used at Beiersdorf for the management of product data. The system landscape depicted is typical for a global company, comprising both global systems supporting processes that affect several organizational units and local systems supporting processes within discrete organizational units [Lehmann 2003]. Figure B9-5 gives examples of systems the functionality of which is supposed to be represented by the MDQM reference model.

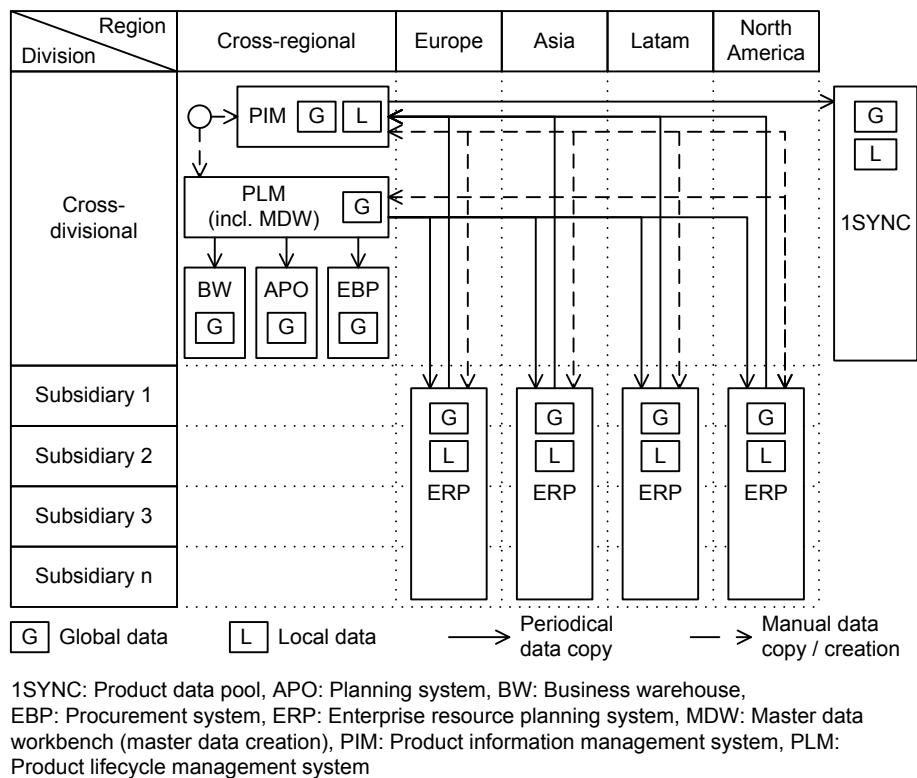


Figure B9-5: Application landscape for managing master data of Beiersdorf

Being part of the PLM system, the Master Data Workbench (MDW) provides functionality for master data creation, thereby ensuring that master data is captured by the PLM system right at the moment it is being created. The users of the system (about 150) work with input masks specifically designed to match with the product innovation process and the product data maintenance process. The whole process of master

data gathering and creation is done by means of one single system. Fast and accurate data processing is ensured, since there is no media breakage and a number of PLM functions (e.g. allocation of unique identifiers, check routines) can already be used in the process of data creation.

Product data distribution to wholesalers and retailers is controlled by means of a central product information management (PIM) system. The PIM system is provided with global and local master data from both the PLM system and local ERP systems. The PIM system also controls the data transfer to 1Sync, Beiersdorf's data pool.

9.5.2 Case Study Approach and Results

At Beiersdorf, planning, maintenance, and implementation of the system landscape is done by a subsidiary of the company, named Beiersdorf Shared Services (BSS). The majority of systems (PLM, ERP, BW, or APO systems) have been implemented with products from SAP. One consequence of this single-source strategy is that complexity regarding interfaces and license contracts could be reduced. Another consequence, however, has been a growing dependence on this particular manufacturer. The consolidation process in the market for DQM and MDM systems in general, together with changes in SAP's portfolio in particular, has raised a number of questions for BSS:

1. Which of the MDQM functions basically available are required by customers of BSS (i.e. the business users at Beiersdorf)?
2. Does BSS offer all required functions in the quality desired?
3. Are standard software systems available from vendors BSS has contractual relations with for functions currently provided by self-developed or modified standard software systems?
4. Are there alternative products available for functions BSS is currently providing not at all or not in the quality desired?

These questions can be assigned to the use cases introduced in the first section of this paper, namely *System evaluation* (3 and 4), *System landscape analysis and design* (1 and 2), and *System landscape transformation* (3 and 4). Thus, the case of Beiersdorf can be taken for evaluation of the reference model [Frank 2007, vom Brocke 2007]. The application of the reference model comprises the following four activities (with Activities 1 to 3 already being completed):

1. *Preparation*. The reference model was not configured. For evaluation of the 72 functions (see Activity 2), a guideline explaining and illustrating each function, an assessment method (see below), and a template for documentation was developed.

2. *Assessment.* In a workshop comprising one expert from BSS and three business users, each of the 72 functions was rated (see below) in terms of 1) to what degree it is demanded by users, and 2) to what degree users are currently satisfied with it.
3. *As-is analysis.* For each of the 72 functions the component of the applications landscape providing the function was identified (PLM system, for example). Furthermore, each component was documented with regard to being a standard application component, an adapted component, or a self-developed component.
4. *To-be concept.* The analysis of the rating results (Activity 2) and the as-is analysis (Activity 3) leads to the identification of functions that are demanded but are provided not at all or not in the quality desired. For each of these functions it has to be stated whether it can be implemented by means of components already in place or whether a new product needs to be integrated.

The assessment method (Activity 2) helps identify functions that are highly demanded but are provided not at all or not in the quality desired. For this purpose, two metrics were defined: demand and satisfaction. Both metrics were rated by the workshop participants for each sub-function of the reference model on a Likert scale ranging from 0 to 5. For aggregation of both ratings, a coefficient matrix C (see below) was used to map the ratings on a scalar value $n = d^T \cdot C \cdot s$, which can be interpreted as a *need for action*. In this formula, d (demand) and s (satisfaction) are unit vectors with 1 indicating the respective rating [e.g. $n = d^T \cdot C \cdot s$ represents a rating of 4].

$$C = \begin{pmatrix} 0/25 & 0/25 & 0/25 & 0/25 & 0/25 & 0/25 \\ 5/25 & 4/25 & 3/25 & 2/25 & 1/25 & 0/25 \\ 10/25 & 8/25 & 6/25 & 4/25 & 2/25 & 0/25 \\ 15/25 & 12/25 & 9/25 & 6/25 & 3/25 & 0/25 \\ 20/25 & 16/25 & 12/25 & 8/25 & 4/25 & 0/25 \\ 25/25 & 20/25 & 15/25 & 10/25 & 5/25 & 0/25 \end{pmatrix}$$

An example: A function rated '4' for demand and '2' for satisfaction is given the value $12/25$. The meaning of the metrics and scales was explained to the workshop participants, so that approximately equidistant rating of functions could be ensured, allowing interpretation of the level of measurement according to a ratio scale.

Figure B9-6 gives an overview of the ratings of the 72 sub-functions (with dark coloring indicating a high need for action) and shows the results of the as-is analysis (smaller part on the right-hand side of each box representing a function).

The as-is analysis (see Figure B9-6) shows the highest need for action in Archiving and Usage statistics. Interestingly, standard software is already available to support archiving functionality, but it is obviously not used. In contrast, for usage statistics no support at all is available at present. Apart from that, Beiersdorf identified a number of functions with relatively high need for action and no software support. An example of that is Glossary/Dictionary, which is used to establish shared definitions for master data in the organization and to provide business context information.

Master Data Lifecycle Management	Data Creation	Data Maintenance	Data Deactivation	Data Archiving
	Conditional Entries	Check-out	Bulk Editing	Archiving
	Bulk Editing	Bulk Editing		History Control
	Plausibility Check	Plausibility Check		
Metadata Management and Master Data Modeling	Data Modeling	Model Analysis	Metadata Management	High demand and low satisfaction
	Data Model Editing	Dependency Analysis	Business Rules Documentation	
	Graphical Modeling	Data Type Recognition	Glossary/Dictionary	No demand or high satisfaction
	Classification	Primary and Secondary Key Recognition	Metadata Import	
	Support of Business Standards	Relationship Recognition	Mandatory Fields Administration	
	Data Model Version Control		Metadata Publication	
			Metadata Transport	
		Metadata Visualization		
Data Quality Assurance	Data Analysis	Data Enrichment	Data Cleansing	
	Compliance Verification	External Reference Data	Delta Import	
	Graphical Analysis	Classification Schemes	Duplicate Recognition	
	Plausibility Lists	Measuring Units	Pattern Recognition	
	Profiling	Multilingual Capability	Plausibility Check	
		Management of Unstructured Data	Spelling Check	
Master Data Integration	Data Import	Data Transformation	Data Export	
	Delta Import	Field Split	Search Based Data Selection	
	Import Formats	Field Merge	Delta Export	
	Connectors	Data Type Conversion	Export Formats	
	Virtual Integration	Pivot Tables	Connectors	
			Limitation	
			Preview	
Cross Functions	Automation	Reports	Search	Workflow Management
	Automated Enrichment	Data Quality Reports	Dynamic Value Search	Bundling of Activities
	Automated Export	Usage Statistics	Free Search	Graphical Workflow Modeling
	Automated Import	Job Monitoring	Fuzzy Search	Create/Maintain Workflows
	Cross-Function Automation	Audit Support		
	Push and Pull Mechanisms			
Administration	Data History Management	User Management		
	Data Lineage	User Interface Design		
	Last User	Roles and Rights		

Standard software
 Custom solution
 Not supported

Figure B9-6: Reference model applied for rating of functions

Besides the identification of MDQM functions and sub-functions to be further developed at Beiersdorf, the application of the reference model has motivated a modification of the reference model itself. Workshop participants had difficulties in distinguishing between some cross functions (i.e. Automation) and sub-functions from other function groups (e.g. Delta Import and Export, Profiling). From a business perspective, it is hardly possible to decide whether a certain data record is in place (and of good quality) because of an automated import and cleansing routine or because of a manual activity performed by IT staff (BSS in the case of Beiersdorf). Therefore, sub-functions from function Automation (of function group Cross Functions) will no longer be considered for as-is functionality ratings by business users (see Section 9.1, use case System landscape analysis and design). However, in responding to other challenges (e.g. System evaluation) the differentiation between function Automation and automated sub-functions has proven to be reasonable (see reverse engineering analysis of MDQM application systems (Figure B9-1, process step 3.1). The differentiation has also been confirmed during the validation by vendors (see Figure B9-1, 4.3).

9.6 Evaluation

9.6.1 Multi-perspective artifact evaluation

With a particular focus on the evaluation of reference models, Frank [2007] has proposed a framework comprising four perspectives of evaluation. This framework is used for evaluation of the reference model presented in this paper, taking into consideration the challenges described in the first section.

- *Economic Perspective.* Due to the simple structure of the reference model (three levels) and clearly defined objectives, the costs for training, adaptation and application (see *Deployment Perspective and Engineering Perspective*) are low (one day for preparation and two days for application in the case presented). Tools supporting the processes of rating and analysis (see Figure B9-6) can be created at low effort (MS Excel based templates for documentation in the case of Beiersdorf, for example). Using the model does not lead to direct cost savings. However, the as-is analysis might identify unnecessary application system licenses or potentials for consolidation. The analysis of the ratings as well as comparison with other companies on the basis of the reference model might lead to better support of required functionality. Both the focus group interviews and the final validation by application system providers have shown that the reference model is capable of substantially simplifying exchange of knowledge and comparison of products.
- *Deployment Perspective.* The focus group interviews, the validation of application system providers, and the application of the reference model at Beiersdorf have shown that the model is easy to understand and well applicable. Any re-

jection of the model due to the fact that it was developed externally (the not-invented-here-syndrome) could not be observed.

- *Engineering Perspective.* Applying the reference model at Beiersdorf has shown that the model can be used to overcome the challenges described. The model's simple structure (see *Deployment Perspective*) ensures its easy adaptability [cf. vom Brocke 2007]. After the model had been applied at Beiersdorf, two functions (Data Sourcing and Data Delivery) for cross-company exchange of master data (product data for external production of artworks and electronic delivery of them, for example) were added under the function group *Master Data Integration* in order to be able to take into account company specific requirements of Beiersdorf (administration of subscriptions at the interfaces of different artwork agencies, for example) for future ratings.
- *Epistemological Perspective.* The validation by application system providers (see Figure B9-1, 4.3.) as well as the application of the reference model at Beiersdorf (see Figure B9-1, 5.2.) has shown that the model is capable of abstracting and representing reality (i.e. MDQM systems). Critical distance is ensured by explication of use cases. By taking into account only one of four ARIS views and one of three ARIS levels, the development and implementation of an application system, for example, is no adequate use case. Moreover, explication of the model design process ensures that scientific principles are followed (such as verifiability and reproducibility of the artifact).

9.6.2 Compliance with GOM

GOM can be used to evaluate the validity of artifact design. They allow for assessing whether the reference model complies with the specifications developed in the second step of the design process (see above).

GOM	Description	Assessment
Correctness	<ul style="list-style-type: none"> • Model consistency, completeness • Semantic correctness • Naming conventions 	<p>The three focus groups did not identify any missing components. However, the design of the cross-functions remains questionable.</p> <p>The reference model follows conventions of the ARIS functional view.</p>
Relevance	<ul style="list-style-type: none"> • Appropriateness of model boundaries • Appropriate level of abstraction 	<p>The focus on MDQM is relevant, as the Motivation section of the paper shows. The level of abstraction has proven appropriate both in the focus groups and the participative case study.</p>
Economic viability	<ul style="list-style-type: none"> • Reasonable modeling effort 	<p>The model has been developed without the need for significant expenditures. Its application fulfils the criteria of the economic perspective according to Frank [2007] (see above).</p>
Clarity	<ul style="list-style-type: none"> • Comprehensibility, 	<p>Becker et al. [1995, p. 438] concede the limited verifia-</p>

GOM	Description	Assessment
	Readability	bility of this guideline. The reference model, however, was understood well by subject matter experts in the focus groups, by vendor representatives, and by case study participants.
Comparability	<ul style="list-style-type: none"> • Compatibility to other models • Support of “as is” and “to be” models 	<p>The compatibility to other models has not been evaluated.</p> <p>The support of comparing “as is” and “to be” situations in a company has been demonstrated in the case study at Beiersdorf.</p>
Systematical Design	<ul style="list-style-type: none"> • Decomposition into views, layers etc. 	The reference model has a hierarchical structure consisting of three levels. This design has been used in similar cases, e.g. in SAP’s business maps [SAP 2007a].

Table B9-4: GOM compliance

9.7 Conclusion and future research

The paper describes a functional reference model for MDQM. The model design process spanned the six steps as proposed by DSRM and includes several design and evaluation cycles.

The reference model is beneficial with regard to both the advancement of the scientific state of the art and the state of the art in practice. Practitioners can use the reference model as an instrument supporting the analysis and design of their MDQM system landscapes. Furthermore, the model facilitates evaluation of software products, and it facilitates both inner-company and cross-company communication. From a researcher’s perspective, the reference model is a representation of an information system, as it specifies business user requirements for MDQM system functionality, and as it delivers new knowledge in terms of representing reality [cf. Frank 2007]. Explication of the research process allows verification, correction, and differentiation of this representation.

The reference model has its limitations due to its focus on the business layer of the functional view of MDQM [Scheer 1992, Scheer 1997] and due to the fact that other ARIS views and levels were not modeled. Hence, the application of the reference model is restricted to use cases similar to the one described. Besides scientific validation of the reference model, further research on the topic should aim at extending the model and adding to it more views and levels. The authors of this paper think that especially the control view and the organizational view offer potential for designing relevant artifacts. Based on case studies, generic characteristics of DQM and MDM organizations (roles and responsibilities, for example) could then be identified to constitute the basis for conceptualizing rights and roles as the reference model’s organizational view. Besides, interdependencies between individual functions of the reference model could be identified and typical MDQM activities (adapting the master data ar-

chitecture, identifying and clearing duplicates, measuring data quality, for example) could be described.

Literaturverzeichnis

- [Alur et al. 2007]
Alur, N., Joseph, R., Mehta, H., Nielsen, J. o. r. T., Vasconcelos, D., IBM WebSphere Information Analyzer and Data Quality Assessment, IBM, 2007
- [Amram/Kulatilaka 1998]
Amram, M., Kulatilaka, N., Real Options: Managing Strategic Investment in an Uncertain World, Harvard Business School Press, Boston, Massachusetts, 1998
- [Ballou et al. 1998]
Ballou, D., Wang, R. Y., Pazer, H., Kumar.Tayi, G., Modeling Information Manufacturing Systems to Determine Information Product Quality, in: Management Science, 44, 1998, Nr. 4, S. 462-484
- [Baskerville 1997]
Baskerville, R. L., Distinguishing action research from participative case studies, in: Journal of Systems and Information Technology, 1, 1997, Nr. 1, S. 24-43
- [Baskerville/Wood-Harper 1996]
Baskerville, R. L., Wood-Harper, A. T., A Critical Perspective on Action Research as a Method for Information Systems Research, in: Journal of Information Technology, 11, 1996, Nr. 3, S. 235-246
- [Baskerville/Wood-Harper 1998]
Baskerville, R. L., Wood-Harper, A. T., Diversity in information systems action research methods, in: European Journal of Information Systems, 7, 1998, S. 90-107
- [Batini et al. 2007]
Batini, C., Barone, D., Mastrella, M., Maurino, A., Ruffini, C., A Framework and a Methodology for Data Quality Assessment and Monitoring, in: Robbert, M. A., O'Hare, R., Markus, M. L., Klein, B. (Hrsg.), Proceedings of the 12th International Conference on Information Quality, Cambridge, Massachusetts, 2007, S. 333-346
- [Batini et al. 2009]
Batini, C., Cappiello, C., Francalanci, C., Maurino, A., Methodologies for Data Quality Assessment and Improvement, in: ACM Computing Surveys, 41, 2009, Nr. 3, S. 1-52
- [Batini/Scannapieco 2006]
Batini, C., Scannapieco, M., Data Quality. Concepts, Methodologies and Techniques, Springer, Berlin, Deutschland et al., 2006
- [BEA 2006]
BEA, BEA AquaLogic™ Data Services Platform, BEA Systems, Inc., San Jose, 2006
- [Beck/Weitzel 2005]
Beck, R., Weitzel, T., Some Economics of Vertical Standards: Integrating SMEs in EDI Supply Chains, in: Electronic Markets, 15, 2005, Nr. 4, S. 313-322

- [Becker et al. 2002]
Becker, J., Algermissen, L., Delfmann, P., Knackstedt, R.,
Referenzmodellierung, in: Das Wirtschaftsstudium (WISU), 30, 2002, Nr. 11,
S. 1392–1395
- [Becker et al. 2003]
Becker, J., Holten, R., Knackstedt, R., Niehaves, B., Forschungsmethodische
Positionierung in der Wirtschaftsinformatik – Epistemologische, ontologische
und linguistische Leitfragen, Institut für Wirtschaftsinformatik Westfälische
Wilhelms-Universität Münster, Münster, Deutschland, 2003
- [Becker et al. 2004]
Becker, J., Holten, R., Knackstedt, R., Niehaves, B., Epistemologische
Positionierungen in der Wirtschaftsinformatik am Beispiel einer
konsensorientierten Informationsmodellierung, in: Frank, U. (Hrsg.),
Wissenschaftstheorie in Ökonomie und Wirtschaftsinformatik, Deutscher
Universitäts-Verlag, Wiesbaden, Deutschland, 2004, S. 335-366
- [Becker et al. 1995]
Becker, J., Rosemann, M., Schütte, R., Grundsätze ordnungsmäßiger
Modellierung (Guidelines for Orderly Modeling), in: Wirtschaftsinformatik, 37,
1995, Nr. 5, S. 435-445
- [Beiersdorf 2009]
Beiersdorf, Annual Report 2009, Beiersdorf AG, Hamburg, Deutschland, 2009
- [Beiersdorf 2010]
Beiersdorf, Annual Report 2010, Beiersdorf AG, Hamburg, Deutschland, 2010
- [Berson/Dubov 2007]
Berson, A., Dubov, L., Master Data Management and Customer Data
Integration for a Global Enterprise, McGraw-Hill, Two Penn Plaza, New York,
2007
- [Bitterer 2007]
Bitterer, A., Gartner's Data Quality Maturity Model, G00139742, Gartner
Research, Stamford, Connecticut, 2007
- [Black/Scholes 1973]
Black, F., Scholes, M., The Pricing of Options and Corporate Liabilities, in:
Journal of Political Economy, 81, 1973, Nr. 3, S. 637-654
- [Boisot/Canals 2004]
Boisot, M., Canals, A., Data, information and knowledge: have we got it right?,
in: Journal of Evolutionary Economics, 14, 2004, Nr. 1, S. 43-67
- [Bourdreau/Couillard 1999]
Bourdreau, A., Couillard, G., Systems Integration and Knowledge
Management, in: Information Systems Management, 16, 1999, Nr. 4, S. 24-32
- [Breya 2009]
Breya, M., Decision-Making Without Borders, in: Information Management,
19, 2009, Nr. 8, S. 32
- [Brinkkemper 1996]
Brinkkemper, S., Method Engineering: Engineering of Information Systems
Development Methods and Tools, in: Information and Software Technology,
38, 1996, Nr. 4, S. 275-280

- [Bucher 2009]
Bucher, T., Ausrichtung der Informationslogistik auf operative Prozesse – Entwicklung und Evaluation einer situativen Methode, Dissertation, Institut für Wirtschaftsinformatik, Universität St. Gallen, St. Gallen, 2009
- [Buffa et al. 2008]
Buffa, M., Gandon, F., Ereteo, G., Sander, P., Faron, C., SweetWiki: A semantic wiki, in: Journal of Web Semantics, 6, 2008, Nr. 1, S. 84-97
- [Bughin et al. 2008]
Bughin, J., Chui, M., Miller, A., Building the Web 2.0 Enterprise. McKinsey Global Survey Results, in: McKinsey Quarterly, 2008, 2008, Nr. July, S. 1-10
- [Bughin et al. 2009]
Bughin, J., Chui, M., Miller, A., How companies are benefiting from Web 2.0: McKinsey Global Survey Results, in: McKinsey Quarterly, 2009, 2009, Nr. September, S. 10-17
- [Buhl et al. 2010]
Buhl, H. U., Mertens, P., Schumann, M., Urbach, N., Smolnik, S., Riempp, G., Letter to the Editor: Statements on the Contribution by Urbach et al. from Issue 4/2009, in: Business & Information Systems Engineering, 2, 2010, Nr. 2, S. 109-120
- [Bund 2004]
Bund, Telekommunikationsgesetz, Bundesrepublik Deutschland, 2004
- [Bund 2009]
Bund, Leistungs- und Finanzierungsvereinbarung, Bundesrepublik Deutschland, 2009
- [Burgess et al. 2004]
Burgess, M. S. E., Gray, W. A., Fiddian, N. J., Quality Measures and the Information Consumer, in: Chengalur-Smith, S., Raschid, L., Long, J., Seko, C. (Hrsg.), Proceedings of the 9th International Conference on Information Quality, Boston, Massachusetts, 2004, S. 373-388
- [Burnett et al. 1999]
Burnett, K., Ng, K. B., Park, S., A Comparison of the Two Traditions of Metadata Development, in: Journal of the American Society for Information Science, 50, 1999, Nr. 13, S. 1209-1217
- [Butler 2007]
Butler, D., MDM as a Foundation for SOA, Oracle Corporation, Redwood Shores, California, 2007
- [Caballero et al. 2008]
Caballero, I., Calero, C., Piattini, M., Verbo, E., MMPro: A Methodology based on ISO/IEC 15939 to Draw up Data Quality Measurement Processes, in: Neely, P., Pipino, L., Slone, J. P. (Hrsg.), Proceedings of the 13th International Conference on Information Quality, Cambridge, Massachusetts, 2008
- [Caballero et al. 2007]
Caballero, I., Verbo, E., Calero, C., Piattini, M., A Data Quality Measurement Information Model based on ISO/IEC 15939, in: Robbert, M. A., O'Hare, R., Markus, M. L., Klein, B. (Hrsg.), Proceedings of the 12th International Conference on Information Quality, Cambridge, Massachusetts, 2007, S. 393-408

[Cachon/Lairiviere 2000]

Cachon, G. P., Lairiviere, M. A., Supply Chain Inventory Management and the Value of Shared Information, in: Management Science, 46, 2000, Nr. 8, S. 1032-1048

[Cai/Ziad 2003]

Cai, Y., Ziad, M., Evaluating Completeness of an Information Product, in: Galletta, D., Ross, J. (Hrsg.), Proceedings of the 9th Americas Conference on Information Systems, Tampa, Florida, 2003

[Cavana et al. 2001]

Cavana, R., Delahaye, B. L., Sekaran, U., Applied Business Research: Qualitative and Quantitative Methods, Wiley, Milton, Australien, 2001

[Chen et al. 2000]

Chen, F., Drezner, Z., Ryan, J. K., Simchi-Levi, D., Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of Forecasting, Lead Times, and Information, in: Management Science, 46, 2000, Nr. 3, S. 436-443

[Christensen 2001]

Christensen, L. B., Experimental Methodology, 8. Aufl., Allyn and Bacon, Boston, Massachusetts et al., 2001

[Crosby 1979]

Crosby, P. B., Quality Is Free: The Art of Making Quality Certain, McGraw-Hill, New York et al., 1979

[Cunningham 1995]

Cunningham, W., What Is Wiki, <http://www.wiki.org/wiki.cgi?WhatIsWiki> (zuletzt geprüft: 11. Juni 2010)

[DAMA 2009]

DAMA, The DAMA Guide to the Data Management Body of Knowledge, Technics Publications, Bradley Beach, New Jersey, 2009

[DataFlux 2008]

DataFlux, The Data Governance Maturity Model, DataFlux Corporation, 2008

[Davenport 1993]

Davenport, T. H., Process Innovation: Reengineering Work through Information Technology, Harvard Business School Press, Boston, Massachusetts, 1993

[Davenport/Prusak 1998]

Davenport, T. H., Prusak, L., Working Knowledge: How Organizations Manage What They Know, Harvard Business School Press, Boston, Massachusetts, 1998

[de Bruin et al. 2005]

de Bruin, T., Freeze, R., Kulkarni, U., Rosemann, M., Understanding the Main Phases of Developing a Maturity Assessment Model, in: Campbell, B., Underwood, J., Bunker, D. (Hrsg.), Proceedings of the 16th Australasian Conference on Information Systems, Sydney, Australien, 2005

[Deloitte 2009]

Deloitte, Telecommunications Predictions - TMT Trends 2009, Deloitte Touche Tohmatsu, London, England, 2009

- [DeLone/McLean 1992]
DeLone, W., McLean, E., Information Systems Success: The Quest for the Dependent Variable, in: Information Systems Research, 3, 1992, Nr. 1, S. 60-95
- [DeLone/McLean 2003]
DeLone, W. H., McLean, E. R., The DeLone and McLean Model of Information Systems Success: A Ten-Year Update, in: Journal of Management Information Systems, 19, 2003, Nr. 4, S. 9-30
- [Deming 1986]
Deming, W. E., Out of the Crisis, MIT-Press, Cambridge, Massachusetts, 1986
- [DIN/EN/ISO 2000]
DIN/EN/ISO, Qualitätsmanagementsysteme: Leitfaden zur Leistungsverbesserung (ISO 9004:2000), DIN EN ISO 9004:2000 D, Deutsches Institut für Normung, 2000
- [Dippold et al. 2005]
Dippold, R., Meier, A., Schnider, W., Unternehmensweites Datenmanagement, Vieweg, Braunschweig, Deutschland et al., 2005
- [Dreibelbis et al. 2008]
Dreibelbis, A., Hechler, E., Milman, I., Oberhofer, M., van Run, P., Wolfson, D., Enterprise Master Data Management: An SOA Approach to Managing Core Information, Pearson Education, Boston, Massachusetts, 2008
- [DTAG 2008]
DTAG, 2008 Annual Report, Deutsche Telekom AG, 2008
- [Dyché/Levy 2006]
Dyché, J., Levy, E., Customer Data Integration, Wiley, Hoboken, New Jersey, 2006
- [Ebner et al. 2011]
Ebner, V., Hüner, K. M., Otto, B., Fallstudie Bayer CropScience AG – Entwurf und Implementierung geschäftsorientierter Datenqualitätskennzahlen, BE HSG / CC CDQ / 23, Institut für Wirtschaftsinformatik Universität St. Gallen, St. Gallen, Schweiz, 2011
- [EFQM 2003a]
EFQM, Assessing for Excellence, European Foundation for the Quality Management, 2003
- [EFQM 2003b]
EFQM, Introducing Excellence, European Foundation for the Quality Management, 2003
- [EFQM 2011]
EFQM, EFQM Framework for Corporate Data Quality Management: Assessing the Organization's Data Quality Management Capabilities, European Foundation for Quality Management, Institut für Wirtschaftsinformatik Universität St. Gallen, St. Gallen, Schweiz, 2011
- [English 1999]
English, L. P., Improving Data Warehouse and Business Information Quality, Wiley, New York et al., 1999

[English 2002]

English, L. P., Process Management and Information Quality: How Improving Information Production Processes Improves Information (Product) Quality, in: Fisher, C., Davidson, B. N. (Hrsg.), Proceedings of the 7th International Conference on Information Quality, Cambridge, Massachusetts, 2002, S. 206-209

[Eppler 2006]

Eppler, M. J., Managing Information Quality, 2. Aufl., Springer, Berlin, Deutschland et al., 2006

[Eppler/Helfert 2004]

Eppler, M. J., Helfert, M., A Classification and Analysis of Data Quality Costs, in: Chengalur-Smith, S., Long, J., Raschid, L., Seko, C. (Hrsg.), Proceedings of the 9th International Conference on Information Quality, Cambridge, Massachusetts, 2004, S. 311-325

[EU 2006]

EU, Verordnung (EG) Nr. 1907/2006 des Europäischen Parlaments und des Rates, Europäisches Parlament und Rat der Europäischen Union, 2006

[Even/Kaiser 2009]

Even, A., Kaiser, M., A Framework for Economic-driven Assessment of Data Quality Decisions, in: Bowen, P., Elmagarmid, A. K., Österle, H., Sattler, K.-U. (Hrsg.), Proceedings of the 14th International Conference on Information Quality, Potsdam, Deutschland, 2009

[Even/Shankaranarayanan 2005]

Even, A., Shankaranarayanan, G., Value-Driven Data Quality Assessment, in: Naumann, F., Gertz, M., Madnick, S. (Hrsg.), Proceedings of the 10th International Conference on Information Quality, Cambridge, Massachusetts, 2005

[Even et al. 2007]

Even, A., Shankaranarayanan, G., Berger, P. D., Economics-Driven Data Management: An Application to the Design of Tabular Data Sets, in: IEEE Transactions on Knowledge & Data Engineering, 19, 2007, Nr. 6, S. 818-831

[Fan et al. 2008]

Fan, W., Geerts, F., Jia, X., Kementsietsidis, A., Conditional Functional Dependencies for Capturing Data Inconsistencies, in: ACM Transactions on Database Systems, 33, 2008, Nr. 2, S. 1-48

[Farrukh et al. 2003]

Farrukh, C., Fraser, P., Gregory, M., Development of a structured approach to assessing practice in product development collaborations, in: Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 217, 2003, Nr. 8, S. 1131-1144

[Fettke/Loos 2003a]

Fettke, P., Loos, P., Classification of reference models: a methodology and its application, in: Information Systems and e-Business Management, 1, 2003, Nr. 1, S. 35-53

- [Fettke/Loos 2003b]
Fettke, P., Loos, P., Multiperspective Evaluation of Reference Models – Towards a Framework, Conceptual Modeling for Novel Application Domains, Springer, Berlin et al., 2003, S. 80-91
- [Fettke/Loos 2004]
Fettke, P., Loos, P., Referenzmodellierungsforschung, in: WIRTSCHAFTSINFORMATIK, 46, 2004, Nr. 5, S. 331-340
- [Fettke/Loos 2007]
Fettke, P., Loos, P., Perspectives on Reference Modeling, in: Fettke, P., Loos, P. (Hrsg.), Reference Modeling for Business Systems Analysis, Idea Group, Hershey, Pennsylvania et al., 2007, S. 1-21
- [Fisher et al. 2003]
Fisher, C. W., Chengalur-Smith, I., Ballou, D. P., The Impact of Experience and Time on the Use of Data Quality Information in Decision Making, in: Information Systems Research, 14, 2003, Nr. 2, S. 170-188
- [Fisher/Kingma 2001]
Fisher, C. W., Kingma, B. R., Criticality of data quality as exemplified in two disasters, in: Information & Management, 39, 2001, Nr. 2, S. 109-116
- [Foshay et al. 2007]
Foshay, N., Mukherjee, A., Taylor, A., Does Data Warehouse End-User Metadata Add Value?, in: Communications of ACM, 50, 2007, Nr. 11, S. 70-77
- [Frank 1997]
Frank, U., Erfahrung, Erkenntnis und Wirklichkeitsgestaltung. Anmerkungen zur Rolle der Empirie in der Wirtschaftsinformatik, in: Grün, O., Heinrich, L. J. (Hrsg.), Wirtschaftsinformatik – Ergebnisse empirischer Forschung, Springer, Wien, Österreich, 1997, S. 21-35
- [Frank 2000]
Frank, U., Evaluation von Artefakten in der Wirtschaftsinformatikforschung, in: Häntschel, I., Heinrich, L. J. (Hrsg.), Evaluation und Evaluationsforschung in der Wirtschaftsinformatik, Oldenbourg, München, Deutschland, 2000, S. 35-48
- [Frank 2007]
Frank, U., Evaluation of Reference Models, in: Fettke, P., Loos, P. (Hrsg.), Reference Modeling for Business Systems Analysis, Idea Group, Hershey, Pennsylvania et al., 2007, S. 118-139
- [Frank et al. 2008]
Frank, U., Schauer, C., Wigand, R. T., Different Paths of Development of Two Information Systems Communities: A Comparative Study Based on Peer Interviews, in: Communications of the AIS, 22, 2008, Nr. 21, S. 319-412
- [Frankreich 2010]
Frankreich, Project de loi: Portant engagement national pour l'environnement (Texte définitif), Französische Nationalversammlung, Paris, Frankreich, 2010
- [Fraser et al. 2002]
Fraser, P., Moultrie, J., Gregory, M., The use of maturity models / grids as a tool in assessing product development capability, Proceedings of the IEEE International Engineering Management Conference, Cambridge, England, 2002, S. 244-249

[Friedman 2006]

Friedman, T., Gartner Study on Data Quality Shows That IT Still Bears the Burden, G00137680, Gartner Research, Stamford, Connecticut, 2006

[Friedman et al. 2007]

Friedman, T., Feinberg, D., Beyer, M. A., Gassman, B., Bitterer, A., Newman, D., Radcliffe, J., Bell, T., Gilbert, M. R., Blechar, M. J., White, A., Hype Cycle for Data Management, 2007, G00148704, Gartner Research, Stamford, Connecticut, 2007

[Friedman 2009]

Friedman, T., Q&A: Common Questions on Data Integration and Data Quality From Gartner's MDM Summit, G00164983, Gartner Research, Stamford, Connecticut, 2009

[Gebauer et al. 2005]

Gebauer, M., Caspers, P., Weigel, N., Reproducible Measurement of Data Field Quality, in: Naumann, F., Gertz, M., Madnick, S. (Hrsg.), Proceedings of the 10th International Conference on Information Quality, Cambridge, Massachusetts, 2005

[Gelinas/Dull 2009]

Gelinas, U. J., Dull, R. B., Accounting Information Systems, 8. Aufl., South-Western, Mason, Ohio, 2009

[Goh et al. 1999]

Goh, C. H., Bressan, S., Madnick, S. E., Siegel, M., Context interchange: new features and formalisms for the intelligent integration of information, in: ACM Transactions on Information Systems, 17, 1999, Nr. 3, S. 270-293

[Gorla/Krishnan 2002]

Gorla, N., Krishnan, K., Information quality: modeling accuracy- timeliness in transaction processing systems, in: Becker, J. D., Willis, G. W. K. (Hrsg.), Proceedings of the 8th Americas Conference on Information Systems, Dallas, Texas, 2002

[Greenwald 1976]

Greenwald, A. G., Within-Subject Designs: To Use or Not To Use?, in: Psychological Bulletin, 83, 1976, Nr. 2, S. 314-320

[Gregor 2006]

Gregor, S., The Nature of Theory in Information Systems, in: Management Information Systems Quarterly, 30, 2006, Nr. 3, S. 611-642

[GS1 2010]

GS1, GS1 General Specifications, Version 10, Issue 1, Global Standards One, 2010

[Gustavsson 2006]

Gustavsson, M., Information Quality Measurement. Considerations when defining Measures supporting MPC decision making, in: Talburt, J., Pierce, E., Wu, N., Campbell, T. (Hrsg.), Proceedings of the 11th International Conference on Information Quality, Cambridge, Massachusetts, 2006

[Gutzwiller 1994]

Gutzwiller, T. A., Das CC RIM-Referenzmodell für den Entwurf von betrieblichen, transaktionsorientierten Informationssystemen, Physica, Heidelberg, Deutschland, 1994

- [Hansen/Neumann 2005]
Hansen, H. R., Neumann, G., *Wirtschaftsinformatik 1*, 9. Aufl., Lucius & Lucius, Stuttgart, Deutschland, 2005
- [Heilig et al. 2006]
Heilig, L., Karch, S., Pfennig, R., Hofmann, C., *SAP NetWeaver Master Data Management*, SAP Press, Bonn, Deutschland, 2006
- [Heinrich et al. 2008]
Heinrich, B., Kaiser, M., Klier, M., Does the EU Insurance Mediation Directive Help to Improve Data Quality? A Metric-Based Analysis, in: Golden, W., Acton, T., Conboy, K., van der Heijden, H., Tuuainen, V. (Hrsg.), *Proceedings of 16th European Conference on Information Systems*, Galway, Irland, 2008
- [Heinrich et al. 2007]
Heinrich, B., Klier, M., Kaiser, M., DQ metrics: A Novel Approach to Quantify Timeliness and its Application In CRM, in: Robbert, M. A., O'Hare, R., Markus, M. L., Klein, B. (Hrsg.), *Proceedings of the 12th International Conference on Information Quality*, Cambridge, Massachusetts, 2007
- [Heinrich et al. 2009]
Heinrich, B., Klier, M., Kaiser, M., A Procedure to Develop Metrics for Currency and its Application in CRM, in: *Journal of Data and Information Quality*, 1, 2009, Nr. 1, S. 1-28
- [Heinrich 2002]
Heinrich, L. J., *Informationsmanagement: Planung, Überwachung und Steuerung der Informationsinfrastruktur*, 7. Aufl., Oldenbourg, München, Deutschland et al., 2002
- [Hevner et al. 2004]
Hevner, A. R., March, S. T., Park, J., Ram, S., Design Science in Information Systems Research, in: *Management Information Systems Quarterly*, 28, 2004, Nr. 1, S. 75-105
- [Heym/Österle 1993]
Heym, M., Österle, H., Computer-Aided Methodology Engineering, in: *Information and Software Technology*, 35, 1993, Nr. 6/7, S. 345-354
- [Hills/Armitage 1979]
Hills, M., Armitage, P., The two-period cross-over clinical trial, in: *British Journal of Clinical Pharmacology*, 8, 1979, Nr. 1, S. 7-20
- [Hipp et al. 2007]
Hipp, J., Müller, M., Hohendorff, J., Naumann, F., Rule-based Measurement of Data Quality in Nominal Data, in: Robbert, M. A., O'Hare, R., Markus, M. L., Klein, B. (Hrsg.), *Proceedings of the 12th International Conference on Information Quality*, Cambridge, Massachusetts, 2007, S. 364-379
- [Hollander/Wolfe 1999]
Hollander, M., Wolfe, D. A., *Nonparametric Statistical Methods*, 2. Aufl., Wiley, New York, 1999
- [Höning 2009]
Höning, F., *Methodenkern des Business Engineering. Metamodell, Vorgehensmodell, Techniken, Ergebnisdokumente und Rollen*, Dissertation, Institut für Wirtschaftsinformatik Universität St. Gallen, St. Gallen, Schweiz, 2009

- [Huang et al. 1999]
Huang, K.-T., Lee, Y. W., Wang, R. Y., Quality Information and Knowledge, Prentice Hall, Upper Saddle River, New Jersey, 1999
- [Hubbard 2009]
Hubbard, D. W., The Failure of Risk Management: Why It's Broken and How to Fix It, Wiley, Hoboken, New Jersey, 2009
- [Hult/Lennung 1980]
Hult, M., Lennung, S.-A., Towards a Definition of Action Research: A Note and a Bibliography, in: Journal of Management Studies, 17, 1980, Nr. 2, S. 241-250
- [Humphrey 1989]
Humphrey, W. S., Managing the Software Process, Addison-Wesley, Reading, Massachusetts, 1989
- [Hüner 2011]
Hüner, K. M., Methode zur Spezifikation geschäftsorientierter Datenqualitätskennzahlen, BE HSG / CC CDQ / 13, Institut für Wirtschaftsinformatik der Universität St. Gallen, St. Gallen, Schweiz, 2011
- [Hüner et al. 2011a]
Hüner, K. M., Brauer, B., Otto, B., Österle, H., Fachliches Metadatenmanagement mit einem semantischen Wiki, in: HMD – Praxis der Wirtschaftsinformatik, 2011
- [Hüner et al. 2009]
Hüner, K. M., Ofner, M., Otto, B., Towards a Maturity Model for Corporate Data Quality, in: Shin, D. (Hrsg.), Proceedings of the 2009 ACM Symposium on Applied Computing, Waikiki Beach, Hawaii, 2009, S. 231-238
- [Hüner/Otto 2009]
Hüner, K. M., Otto, B., The Effect of Using a Semantic Wiki for Metadata Management: A Controlled Experiment, in: Sprague, R. H. (Hrsg.), Proceedings of the 42nd Hawaii International Conference on System Sciences, Waikoloa, Hawaii, 2009
- [Hüner et al. 2008]
Hüner, K. M., Otto, B., Österle, H., Metadatenmanagement mit einem Semantischen Wiki, in: IM – Fachzeitschrift für Information Management & Consulting, 23, 2008, Nr. 2, S. 34-40
- [Hüner et al. 2011b]
Hüner, K. M., Otto, B., Österle, H., Collaborative management of business metadata, in: International Journal of Information Management, 31, 2011, Nr. 4, S. 366-373
- [Hüner et al. 2011c]
Hüner, K. M., Schierning, A., Otto, B., Österle, H., Product data quality in supply chains: the case of Beiersdorf, in: Electronic Markets, 21, 2011, Nr. 2, S. 141-154
- [Hüner et al. 2011d]
Hüner, K. M., Otto, B., Österle, H., Specifying data quality metrics that matter to business, in: Journal of the Association for Information Systems, 2011 (in 2. Begutachtung)

- [IBM 2005]
IBM, IBM Completes Acquisition Of Ascential Software, IBM Corporation, Armonk, New York, 2005
- [IBM 2006]
IBM, IBM Information Server Version 8.0: Information Server Introduction, IBM Corporation, Armonk, New York, 2006
- [IBM 2007]
IBM, The IBM Data Governance Council Maturity Model: Building a roadmap for effective data governance, IBM Software Group, 2007
- [IEEE 1990]
IEEE, IEEE Standard Glossary of Software Engineering Terminology, IEEE Std 610.121990, The Institute of Electrical and Electronics Engineers, 1990
- [IEEE 1998]
IEEE, IEEE Standard for a Software Quality Metrics Methodology, IEEE Std 1061-1998, The Institute of Electrical and Electronics Engineers, 1998
- [Inmon et al. 2008]
Inmon, W. H., O'Neil, B. K., Fryman, L., Business Metadata: Capturing Enterprise Knowledge, Morgan Kaufmann, Burlington, Massachusetts, 2008
- [ISO/IEC 2007]
ISO/IEC, ISO/IEC 15939. Systems and software engineering – Measurement process, ISO/IEC 15939:2007(E), International Organization for Standardization, 2007
- [ITGI 2007]
ITGI, CobiT 4.1, IT Governance Institute, 2007
- [Joshi/Rai 2000]
Joshi, K., Rai, A., Impact of the quality of information products on information system users' job satisfaction: an empirical investigation, in: Information Systems Journal, 10, 2000, Nr. 4, S. 323-345
- [Jupp et al. 2007]
Jupp, S., Moulton, G., Rector, A., Stevens, R., Wroe, C., A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools, The University of Manchester, Manchester, England, 2007
- [Juran 1988]
Juran, J. M., Juran on Planning for Quality, The Free Press, New York, 1988
- [Kagermann et al. 2010]
Kagermann, H., Österle, H., Jordan, J. M., IT-Driven Business Models: Global Case Studies in Transformation, Wiley, Hoboken, New Jersey, 2010
- [Kahn et al. 2002]
Kahn, B. K., Strong, D. M., Wang, R. Y., Information Quality Benchmarks: Product and Service Performance, in: Communications of the ACM, 45, 2002, Nr. 4, S. 184-192
- [Kaiser et al. 2007]
Kaiser, M., Klier, M., Heinrich, B., How to Measure Data Quality? - A Metric-Based Approach, in: Rivard, S., Webster, J. (Hrsg.), Proceedings of the 28th International Conference on Information Systems, Montreal, Kanada, 2007

[Kane/Fichman 2009]

Kane, G. C., Fichman, R. G., The Shoemaker's Children: Using Wikis for Information Systems Teaching, Research, and Publication, in: *Management Information Systems Quarterly*, 33, 2009, Nr. 1, S. 1-17

[Kaplan/Norton 1992]

Kaplan, R. S., Norton, D. P., The Balanced Scorecard: Measures that Drive Performance, in: *Harvard Business Review*, 70, 1992, Nr. 1, S. 71-79

[Karel 2006]

Karel, R., *Introducing Master Data Management*, Forester Research, Cambridge, Massachusetts, 2006

[Keren 1993]

Keren, G., Between- or Within-Subjects Design: A Methodological Dilemma, in: Keren, G., Lewis, C. (Hrsg.), *A Handbook for Data Analysis in the Behavioral Sciences – Methodological Issues*, Erlbaum, Hillsdale, New Jersey et al., 1993, S. 257-272

[Khatri/Brown 2010]

Khatri, V., Brown, C. V., Designing Data Governance, in: *Communications of the ACM*, 53, 2010, Nr. 1, S. 148-152

[Kitchenham et al. 2002]

Kitchenham, B. A., Pfleeger, S. L., Hoaglin, D. C., Rosenberg, J., Preliminary Guidelines for Empirical Research in Software Engineering, in: *IEEE Transactions on Software Engineering*, 28, 2002, Nr. 8, S. 721-734

[Kokemüller 2009]

Kokemüller, J., *Stammdatenmanagementsysteme 2009*, Fraunhofer-Verlag, Stuttgart, Deutschland, 2009

[Krcmar 2005]

Krcmar, H., *Informationsmanagement*, 4. Aufl., Springer, Berlin, Deutschland et al., 2005

[Kröttsch et al. 2005]

Kröttsch, M., Vrandečić, D., Völkel, M., Wikipedia and the Semantic Web – The Missing Links, in: Voss, J., Lih, A., Klein, S., Ma, C. (Hrsg.), *Proceedings of Wikimania 2005*, Frankfurt, Deutschland, 2005

[Kröttsch et al. 2007]

Kröttsch, M., Vrandečić, D., Völkel, M., Haller, H., Studer, R., Semantic Wikipedia, in: *Journal of Web Semantics*, 5, 2007, Nr. 4, S. 251-261

[Kulp et al. 2004]

Kulp, S. C., Lee, H. L., Ofek, E., Manufacturer Benefits from Information Integration with Retail Customers, in: *Management Science*, 50, 2004, Nr. 4, S. 431-444

[Laitenberger et al. 2001]

Laitenberger, O., El Emam, K., Harbich, T., An internally Replicated Quasi-Experimental Comparison of Checklist and Perspective-based Reading of Code Documents, in: *IEEE Transactions on Software Engineering*, 27, 2001, Nr. 5, S. 387-421

- [Lau et al. 2002]
Lau, J. S. K., Hunag, G. Q., Mak, K. L., Web-based simulation portal for investigating impacts of sharing production information on supply chain dynamics from the perspective of inventory allocation, in: *Integrated Manufacturing Systems*, 13, 2002, Nr. 5, S. 345-358
- [Lee et al. 1997]
Lee, H. L., Padmanabhan, V., Whang, S., Information Distortion in a Supply Chain: The Bullwhip Effect, in: *Management Science*, 43, 1997, Nr. 4, S. 546-558
- [Lee et al. 2006]
Lee, Y. W., Pipino, L. L., Funk, J. D., Wang, R. Y., *Journey to Data Quality*, MIT Press, Cambridge, Massachusetts, 2006
- [Lee et al. 2002]
Lee, Y. W., Strong, D. M., Kahn, B. K., Wang, R. Y., AIMQ: a methodology for information quality assessment, in: *Information & Management*, 40, 2002, Nr. 2, S. 133-146
- [Legner/Schemm 2008]
Legner, C., Schemm, J., Toward the Inter-organizational Product Information Supply Chain - Evidence from the Retail and Consumer Goods Industries, in: *Journal of the Association for Information Systems*, 9, 2008, Nr. 4, S. 119-150
- [Lehmann 2003]
Lehmann, H., An object-oriented architecture model for international information systems?, in: *Journal of Global Information Management*, 11, 2003, Nr. 3
- [Leist/Zellner 2006]
Leist, S., Zellner, G., Evaluation of Current Architecture Frameworks, in: Liebrock, L. M. (Hrsg.), *Proceedings of the 21st Annual ACM Symposium on Applied Computing*, Dijon, Frankreich, 2006, S. 1546-1553
- [Lenk 1979]
Lenk, H., *Pragmatische Vernunft: Philosophie zwischen Wissenschaft und Praxis*, Reclam, Stuttgart, Deutschland, 1979
- [Leser/Naumann 2007]
Leser, U., Naumann, F., *Informationsintegration: Architekturen und Methoden zur Integration verteilter und heterogener Datenquellen*, dpunkt, Heidelberg, Deutschland, 2007
- [Lipsey 1990]
Lipsey, M. W., *Design Sensitivity: Statistical Power for Experimental Research*, Sage, Newbury Park, California et al., 1990
- [Loos 1999]
Loos, P., Grunddatenverwaltung und Betriebsdatenerfassung als Basis der Produktionsplanung und -steuerung, in: Corsten, H., Friedl, B. (Hrsg.), *Produktionscontrolling*, Vahlen, München, Deutschland, 1999, S. 227-252
- [Loshin 2001]
Loshin, D., *Enterprise Knowledge Management: The Data Quality Approach*, Morgan Kaufmann, San Diego, California, 2001

- [Loshin 2008]
Loshin, D., Master Data Management, Morgan Kaufmann, Burlington, Massachusetts, 2008
- [Louridas/Loucopoulos 2000]
Louridas, P., Loucopoulos, P., A generic model for reflective design, in: ACM Transactions on Software Engineering and Methodology, 9, 2000, Nr. 2, S. 199-237
- [Lunze 2008]
Lunze, J., Regelungstechnik 1. Systemtheoretische Grundlagen, Analyse und Entwurf einschleifiger Regelungen, 7. Aufl., Springer, Berlin, Deutschland et al., 2008
- [Madnick 1995]
Madnick, S. E., Integrating Information From Global Systems: Dealing with the "On- and Off-Ramps" of the Information SuperHighway, in: Journal of Organizational Computing and Electronic Commerce, 5, 1995, Nr. 2, S. 69–82
- [Majchrzak et al. 2006]
Majchrzak, A., Wagner, C., Yates, D., Corporate Wiki Users: Results of a Survey, in: Riehle, D. (Hrsg.), Proceedings of the 2006 International Symposium on Wikis, Odense, Dänemark, 2006, S. 99-104
- [March/Smith 1995]
March, S. T., Smith, G. F., Design and natural science research on information technology, in: Decision Support Systems, 15, 1995, Nr. 4, S. 251-266
- [March/Storey 2008]
March, S. T., Storey, V. C., Design Science in the Information Systems Discipline: An Introduction to the Special Issue on Design Science Research, in: Management Information Systems Quarterly, 32, 2008, Nr. 4, S. 725-730
- [Marco 2000]
Marco, D., Building and Managing the Meta Data Repository. A Full Lifecycle Guide, Wiley, New York et al., 2000
- [Martin 1975]
Martin, J., Computer Data-Base Organization, Prentice-Hall, Englewood Cliffs, New Jersey, 1975
- [McCann 2010]
McCann, D., If you build it... in: CFO, 26, 2010, Nr. 3, S. 29-31
- [McCarthy 1962]
McCarthy, J., Towards a Mathematical Science of Computation, in: Popplewell, C. M. (Hrsg.), Proceedings of the IFIP Congress 62, München, Deutschland, 1962, S. 21-28
- [McKinney Jr./Yoos 2010]
McKinney Jr., E. H., Yoos, C. J., Information About Information: A Taxonomy of Views, in: Management Information Systems Quarterly, 34, 2010, Nr. 2, S. 329-344
- [McKnight 2005]
McKnight, W., What Does it all Mean?, in: DM Review, 15, 2005, Nr. 6, S. 80
- [Mende 1995]
Mende, M., Ein Führungssystem für Geschäftsprozesse, Dissertation, Institut für Wirtschaftsinformatik Universität St. Gallen, St. Gallen, Schweiz, 1995

- [Mertens 2007]
Mertens, P., Integrierte Informationsverarbeitung 1: Operative Systeme in der Industrie, 16. Aufl., Gabler, 2007
- [Mertens 2009]
Mertens, P., Integrierte Informationsverarbeitung 1: Operative Systeme in der Industrie, 17. Aufl., Gabler, Wiesbaden, Deutschland, 2009
- [Mertens et al. 2005]
Mertens, P., Bodendorf, F., König, W., Picot, A., Schumann, M., Hess, T., Grundzüge der Wirtschaftsinformatik, 9. Aufl., Springer, Berlin, Deutschland et al., 2005
- [Moody 2005]
Moody, D. L., Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions, in: Data & Knowledge Engineering, 55, 2005, Nr. 3, S. 243-276
- [Morgan/Krueger 1993]
Morgan, D. L., Krueger, R. A., When to use Focus Groups and why?, in: Morgan, D. L. (Hrsg.), Successful Focus Groups, Sage, Newbury Park, California, 1993, S. 3-19
- [Morris 1938]
Morris, C. W., Foundations of the Theory of Signs, in: Neurath, O., Bohr, N., Dewey, J., Russell, B., Carnap, R., Morris, C. W. (Hrsg.), International Encyclopedia of Unified Science, University of Chicago Press, London, England, 1938
- [Muljadi et al. 2006]
Muljadi, H., Takeda, H., Ando, K., Development of a Semantic Wiki-based Feature Library for the Extraction of Manufacturing Feature and Manufacturing Information, in: International Journal of Computer Science, 1, 2006, Nr. 4, S. 265-273
- [Mure/Zwick 2009]
Mure, J., Zwick, T., Der Einfluss erst seit kurzem verfügbarer Datenzugänge auf die empirische Forschungsagenda der BWL, in: Scherer, A. G., Kaufmann, I. M., Patzer, M. (Hrsg.), Methoden in der Betriebswirtschaftslehre, Gabler, Wiesbaden, Deutschland, 2009, S. 113-115
- [Muschter 1998]
Muschter, S., IS-gestütztes Prozessmanagement, Dissertation, Institut für Wirtschaftsinformatik Universität St. Gallen, St. Gallen, Schweiz, 1998
- [Nakatani et al. 2006]
Nakatani, K., Chuang, T.-T., Zhou, D., Data Synchronization Technology: Standards, Business Values and Implications, in: Communications of AIS, 17, 2006, Nr. 1, S. 962-994
- [Naumann/Rolker 2000]
Naumann, F., Rolker, C., Assessment Methods for Information Quality Criteria, in: Klein, B. D., Rossin, D. F. (Hrsg.), Proceedings of the 5th Conference on Information Quality, Cambridge, Massachusetts, 2000, S. 148-162

- [Newman/Logan 2006]
Newman, D., Logan, D., Achieving Agility: How Enterprise Information Management Overcomes Information Silos, G00137817, Gartner Research, Stamford, Connecticut, 2006
- [Nicolaou/McKnight 2006]
Nicolaou, A. I., McKnight, D. H., Perceived Information Quality in Data Exchanges: Effects on Risk, Trust, and Intention to Use, in: Information Systems Research, 17, 2006, Nr. 4, S. 332-351
- [Nixon/Simperl 2006]
Nixon, L. J. B., Simperl, E. P. B., Makna and MultiMakna: towards semantic and multimedia capability in wikis for the emerging web, in: Schaffert, S., Sure, Y. (Hrsg.), Proceedings of the Semantics 2006, Wien, Österreich, 2006
- [Nunamaker et al. 1991]
Nunamaker, J. F., Chen, M., Purdin, T. D. M., Systems Development in Information Systems Research, in: Journal of Management Information Systems, 7, 1991, Nr. 3, S. 89-106
- [Nuseibeh et al. 1996]
Nuseibeh, B. A., Finkelstein, Anthony, Kramer, J., Method Engineering for Multi-Perspective Software Development, in: Information and Software Technology, 38, 1996, Nr. 4, S. 267-274
- [Ofner et al. 2009]
Ofner, M. H., Hüner, K. M., Otto, B., Dealing with Complexity: A Method to Adapt and Implement a Maturity Model for Corporate Data Quality Management, in: Nelson, M. L., Shaw, M. J., Strader, T. J. (Hrsg.), Proceedings of the 15th Americas Conference on Information Systems, San Francisco, California, 2009
- [Olle et al. 1988]
Olle, W. T., Hagelstein, J., Macdonald, I. G., Rolland, C., Sol, H. G., van Assche, F. J. M., Verrijn-Stuart, A. A., Information Systems Methodologies - A Framework for Understanding, Addison-Wesley, Wokingham, England, 1988
- [Olson 2003]
Olson, J. E., Data Quality: The Accuracy Dimension, Morgan Kaufmann, San Francisco, California, 2003
- [OMG 2006]
OMG, Meta Object Facility (MOF) Core Specification: OMG Available Specification Version 2.0, formal/06-01-01, Object Management Group, 2006
- [OMG 2009]
OMG, OMG Unified Modeling Language (OMG UML), Infrastructure: Version 2.2, formal/09-02-04, Object Management Group, 2009
- [Orr 1998]
Orr, K., Data quality and systems theory, in: Communications of the ACM, 41, 1998, Nr. 2, S. 66-71
- [Österle 1995]
Österle, H., Business Engineering: Prozeß- und Systementwicklung, Band 1: Entwurfstechniken, 2. Aufl., Springer, Berlin, Deutschland, 1995

[Österle 1996]

Österle, H., Business Engineering: Transition to the Networked Enterprise, in: EM – Electronic Markets, 6, 1996, Nr. 2, S. 14-16

[Österle et al. 2011]

Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Loos, P., Mertens, P., Oberweis, A., Sinz, E. J., Memorandum gestaltungsorientierter Wirtschaftsinformatik, in: European Journal of Information Systems, 2011

[Österle/Blessing 2003]

Österle, H., Blessing, D., Business Engineering Modell, in: Österle, H., Winter, R. (Hrsg.), Business Engineering, 2. Aufl., Springer, Berlin, Deutschland, 2003, S. 65-85

[Österle/Blessing 2005]

Österle, H., Blessing, D., Ansätze des Business Engineering, in: HMD – Praxis der Wirtschaftsinformatik, 42, 2005, Nr. 241, S. 7-17

[Österle/Otto 2010]

Österle, H., Otto, B., Konsortialforschung: Eine Methode für die Zusammenarbeit von Forschung und Praxis in der gestaltungsorientierten Wirtschaftsinformatikforschung, in: WIRTSCHAFTSINFORMATIK, 52, 2010, Nr. 5

[Österle et al. 2007]

Österle, H., Winter, R., Höning, F., Kurpjuweit, S., Osl, P., Business Engineering: Core-Business-Metamodell, in: Wisu – Das Wirtschaftsstudium, 36, 2007, Nr. 2, S. 191-194

[Otto/Hüner 2009]

Otto, B., Hüner, K. M., Funktionsarchitektur für unternehmensweites Stammdatenmanagement, BE HSG / CC CDQ / 14, Institut für Wirtschaftsinformatik Universität St. Gallen, St. Gallen, 2009

[Otto et al. 2009a]

Otto, B., Hüner, K. M., Österle, H., Identification of Business Oriented Data Quality Metrics, in: Bowen, P., Elmagarmid, A. K., Österle, H., Sattler, K.-U. (Hrsg.), Proceedings of the 14th International Conference on Information Quality, Potsdam, Deutschland, 2009, S. 122-134

[Otto et al. 2009b]

Otto, B., Hüner, K. M., Österle, H., Unternehmensweite Stammdatenqualität, in: ERP Management, 2009, 2009, Nr. 3, S. 19-21

[Otto et al. 2010]

Otto, B., Hüner, K. M., Österle, H., A Cybernetic View on Data Quality Management, in: Leidner, D. E., Elam, J. J. (Hrsg.), Proceedings of the 16th Americas Conference on Information Systems, Lima, Peru, 2010

[Otto/Österle 2010]

Otto, B., Österle, H., Relevance Through Consortium Research? A Case Study, in: Alexander, T., Turpin, M., van Deventer, J. (Hrsg.), Proceedings of the European Conference on Information Systems, Pretoria, Südafrika, 2010

[Otto et al. 2011]

Otto, B., Hüner, K. M., Österle, H., Toward a functional reference model for master data quality management, in: Information Systems and e-Business Management, 2011, doi: 10.1007/s10257-011-0178-0

[Otto et al. 2008]

Otto, B., Wende, K., Schmidt, A., Hüner, K. M., Vogel, T., Unternehmensweites Datenqualitätsmanagement: Ordnungsrahmen und Anwendungsbeispiele, in: Dinter, B., Winter, R. (Hrsg.), Integrierte Informationslogistik, Springer, Berlin, Deutschland, 2008, S. 211-230

[Otto et al. 2007]

Otto, B., Wende, K., Schmidt, A., Osl, P., Towards a Framework for Corporate Data Quality Management, in: Toleman, M., Cater-Steel, A., Roberts, D. (Hrsg.), Proceedings of 18th Australasian Conference on Information Systems, Toowoomba, Australien, 2007, S. 916-926

[Paoli/Loregian 2006]

Paoli, F. D., Loregian, M., On Integrating a Semantic Wiki in a Knowledge Management System, in: Tummarello, G., Bouquet, P., Signore, O. (Hrsg.), Proceedings of the 3rd Italian Semantic Web Workshop (3rd Semantic Web Applications and Perspectives), Pisa, Italien, 2006

[Parssian et al. 1999]

Parssian, A., Sarkar, S., Jacob, V., Assessing Data Quality for Information Products, in: De, P., DeGross, J. I. (Hrsg.), Proceedings of the 20th International Conference on Information Systems, Charlotte, North Carolina, 1999

[Parssian et al. 2004]

Parssian, A., Sarkar, S., Jacob, V. S., Assessing Data Quality for Information Products: Impact of Selection, Projection, and Cartesian Product, in: Management Science, 50, 2004, Nr. 7, S. 967-982

[Parssian et al. 2009]

Parssian, A., Sarkar, S., Jacob, V. S., Impact of the Union and Difference Operations on the Quality of Information Products, in: Information Systems Research, 20, 2009, Nr. 1, S. 99-120

[Peppers et al. 2008]

Peppers, K., Tuunanen, T., Rothenberger, M. A., Chatterjee, S., A Design Science Research Methodology for Information Systems Research, in: Journal of Management Information Systems, 24, 2008, Nr. 3, S. 45-77

[Peppard et al. 2007]

Peppard, J., Ward, J., Daniel, E., Managing the Realization of Business Benefits from IT Investments, in: MIS Quarterly Executive, 6, 2007, Nr. 1, S. 1-11

[Pierce 2004]

Pierce, E. M., Assessing Data Quality with Control Metrics, in: Communications of the ACM, 47, 2004, Nr. 2, S. 82-86

[Pipino et al. 2002]

Pipino, L. L., Lee, Y. W., Wang, R. Y., Data Quality Assessment, in: Communications of the ACM, 45, 2002, Nr. 4, S. 211-218

[Pipino et al. 2005]

Pipino, L. L., Wang, R. Y., Kocpcso, D., Rybold, W., Developing Measurement Scales for Data-Quality Dimensions, in: Wanf, R. Y., Pierce, E. M., Madnick, S. E., Fisher, C. W. (Hrsg.), Information Quality. Advances in Management Information Systems, M.E. Sharp, Armonk, New York, 2005, S. 37-51

- [Plötner et al. 2010]
Plötner, O., Sieben, B., Kummer, T.-F., Kosten- und Erlösrechnung: Anschaulich, kompakt, praxisnah, 2. Aufl., Springer, Berlin, Deutschland et al., 2010
- [Porter 1985]
Porter, M. E., Competitive Advantage: Creating and Sustaining Superior performance Free Press, New York, 1985
- [Potts 1989]
Potts, C., A Generic Model for Representing Design Methods, Proceedings of the 11th International Conference on Software Engineering, Pittsburgh, Pennsylvania, 1989, S. 217-226
- [Potts/Runs 1988]
Potts, C., Runs, G., Recording the Reasons for Design Decisions, Proceedings of the 10th International Conference on Software Engineering, Singapur, 1988, S. 418-427
- [Price et al. 2008]
Price, R., Neiger, D., Shanks, G., Developing a Measurement Instrument for Subjective Aspects of Information Quality, in: Communications of AIS, 22, 2008, Nr. 1, S. 49-74
- [Price/Shanks 2005]
Price, R., Shanks, G., A semiotic information quality framework: development and comparative analysis, in: Journal of Information Technology, 20, 2005, Nr. 2, S. 88-102
- [Probst 2003]
Probst, C., Referenzmodell für IT-Service-Informationssysteme, Logos, Berlin, Deutschland, 2003
- [Pula et al. 2003]
Pula, E. N., Stone, M., Foss, B., Customer data management in practice: An insurance case study, in: Journal of Database Marketing, 10, 2003, Nr. 4, S. 327-341
- [Radcliffe/White 2009]
Radcliffe, J., White, A., Key Issues for Master Data Management, 2009, G00165723, Gartner Research, Stamford, Connecticut, 2009
- [Raghunathan 1999]
Raghunathan, S., Impact of information quality and decision-maker quality on decision quality: a theoretical model and simulation analysis, in: Decision Support Systems, 26, 1999, Nr. 4, S. 275-286
- [Rautenstrauch 1993]
Rautenstrauch, C., Integration Engineering: Konzeption, Entwicklung und Einsatz integrierter Softwaresysteme, Addison-Wesley, Bonn, Deutschland et al., 1993
- [Redman 1996]
Redman, T. C., Data Quality for the Information Age, Artech House, Boston, Massachusetts et al., 1996
- [Redman 1998]
Redman, T. C., The Impact of Poor Data Quality on the Typical Enterprise, in: Communications of the ACM, 41, 1998, Nr. 2, S. 79-82

- [Redman 2001]
Redman, T. C., *Data Quality: The Field Guide*, Digital Press, Boston, Massachusetts, 2001
- [Redman 2005]
Redman, T. C., *Measuring Data Accuracy. A Framework and Review*, in: Wanf, R. Y., Pierce, E. M., Madnick, S. E., Fisher, C. W. (Hrsg.), *Information Quality. Advances in Management Information Systems*, M.E. Sharp, Armonk, New York, 2005, S. 21-36
- [Reid/Catterall 2005]
Reid, A., Catterall, M., *Invisible data quality issues in a CRM implementation*, in: *Journal of Database Marketing & Customer Strategy Management*, 12, 2005, Nr. 4, S. 305-314
- [Reid et al. 2007]
Reid, R., Fraser-King, G., Schwaderer, W. D., *Data Lifecycles: Managing Data for Strategic Advantage*, Wiley, Chichester, England, 2007
- [Robbert/Fletcher 2007]
Robbert, M.-A., Fletcher, D., *Data Quality Begins with the Business Model*, *Proceedings of the 13th Americas Conference on Information Systems*, 2007
- [Rosemann/de Bruin 2005]
Rosemann, M., de Bruin, T., *Application of a Holistic Model for Determining BPM Maturity*, *Proceedings of the AIM Pre-ICIS Workshop*, Washington, D.C., 2005
- [Rosemann et al. 2004]
Rosemann, M., de Bruin, T., Hueffner, T., *A Model for Business Process Management Maturity*, in: Wenn, A., McGrath, M., Burstein, F. (Hrsg.), *Proceedings of the 13th Australasian Conference on Information Systems*, Melbourne, Australien, 2004
- [Rosemann/Schütte 1999]
Rosemann, M., Schütte, R., *Multiperspektivische Referenzmodellierung*, in: Becker, J., Rosemann, M., Schütte, R. (Hrsg.), *Referenzmodellierung - State-of-the-Art und Entwicklungsperspektiven*, Physica, Heidelberg, Deutschland, 1999, S. 22-44
- [Rosemann/Vessey 2008]
Rosemann, M., Vessey, I., *Toward Improving the Relevance of Information Systems Research to Practice: The Role of Applicability Checks*, in: *Management Information Systems Quarterly*, 32, 2008, Nr. 1, S. 1-22
- [Rossi et al. 2004]
Rossi, M., Ramesh, B., Lyytinen, K., Tolvanen, J.-P., *Managing Evolutionary Method Engineering by Method Rationale*, in: *Journal of the Association for Information Systems*, 5, 2004, Nr. 9, S. 356-391
- [Ryu et al. 2006]
Ryu, K.-S., Park, J.-S., Park, J.-H., *A Data Quality Management Maturity Model*, in: *Journal of the Electronics and Telecommunications Research Institute*, 28, 2006, Nr. 2, S. 191-204
- [Saaksvuori/Immonen 2008]
Saaksvuori, A., Immonen, A., *Product Lifecycle Management*, 3. Aufl., Springer, Berlin, Deutschland et al., 2008

- [SAP 2007a]
SAP, SAP NetWeaver - Technology Map Edition 2007,
<http://www50.sap.com/businessmaps/004F8486F21E4522943507897144041B.htm> (zuletzt geprüft: 19. Januar 2009)
- [SAP 2007b]
SAP, SAP to Acquire Business Objects in Friendly Takeover, SAP AG,
Walldorf, Deutschland, 2007
- [SAP 2008]
SAP, SAP Library - SAP NetWeaver Master Data Management (MDM),
http://help.sap.com/saphelp_mdm550/helpdata/en/43/D7AED5058201B4E1000000A11466F/frameset.htm (zuletzt geprüft: 19. Februar 2009)
- [SAP 2011]
SAP, SAP ERP - SAP Solution Map,
[http://www.sap.com/global/scripts/jump_frame.epx?content=/businessmaps/82087E7A8EB94E1F885C5A1F62C789FB.htm&CloseLabel=Fenster schliessen](http://www.sap.com/global/scripts/jump_frame.epx?content=/businessmaps/82087E7A8EB94E1F885C5A1F62C789FB.htm&CloseLabel=Fenster%20schliessen)
(zuletzt geprüft: 8. Februar 2009)
- [Schaffert 2006]
Schaffert, S., IkeWiki: A Semantic Wiki for Collaborative Knowledge Management, in: Mehandjiev, N. (Hrsg.), Proceedings of the 15th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, Manchester, England, 2006, S. 388-396
- [Schaffert et al. 2008]
Schaffert, S., Bry, F., Baumeister, J., Kiesel, M., Semantic Wikis, in: IEEE Software, 25, 2008, Nr. 4, S. 8-11
- [Scheer 1991]
Scheer, A.-W., Architektur integrierter Informationssysteme: Grundlagen der Unternehmensmodellierung, Springer, Berlin, Deutschland et al., 1991
- [Scheer 1992]
Scheer, A.-W., Architecture of Integrated Information Systems - Foundations of Enterprise Modelling, Springer, Berlin, Deutschland, 1992
- [Scheer 1997]
Scheer, A.-W., Wirtschaftsinformatik: Referenzmodelle für industrielle Geschäftsprozesse, 7. Aufl., Springer, Berlin, Deutschland et al., 1997
- [Scheer 1998]
Scheer, A.-W., ARIS – Vom Geschäftsprozess zum Anwendungssystem, 3. Aufl., Springer, Berlin et al., 1998
- [Scheer 2001]
Scheer, A., -W., ARIS - Modellierungsmethoden, Metamodelle, Anwendungen, Springer, Berlin, Deutschland, 2001
- [Scheer et al. 2005]
Scheer, A.-W., Thomas, O., Adam, O., Process Modeling Using Event-Driven Process Chains, in: Dumas, M., van der Aalst, W., ten Hofstede, A. H. M. (Hrsg.), Process-Aware Information Systems: Bridging People and Software Through Process Technology, Wiley, Hoboken, New Jersey, 2005, S. 119-145

- [Schelp/Winter 2007]
Schelp, J., Winter, R., Towards a Methodology for Service Construction, Proceedings of the 40th Hawaii International Conference on Systems Sciences (HICSS-40), Los Alamitos, California, 2007
- [Schemm 2008]
Schemm, J. W., Stammdatenmanagement zwischen Handel und Konsumgüterindustrie: Referenzarchitektur für die überbetriebliche Datensynchronisation, Dissertation, Institut für Wirtschaftsinformatik, Universität St. Gallen, St. Gallen, Schweiz, 2008
- [Schmidt 2010]
Schmidt, A., Entwicklung einer Methode zur Stammdatenintegration, Dissertation, Institut für Wirtschaftsinformatik Universität St. Gallen, St. Gallen, Schweiz, 2010
- [Schmidt et al. 2010]
Schmidt, A., Hüner, K. M., Grewe, A., Fallstudie Deutsche Telekom AG – Einheitliche Datenarchitektur als Grundlage für unternehmensweites Datenqualitätsmanagement, BE HSG / CC CDQ / 23, Institut für Wirtschaftsinformatik Universität St. Gallen, St. Gallen, Schweiz, 2010
- [Schmidt/Otto 2008]
Schmidt, A., Otto, B., A Method for the Identification and Definition of Information Objects, in: Neely, P., Pipino, L., Slone, J. P. (Hrsg.), Proceedings of the 13th International Conference on Information Quality, Cambridge, Massachusetts, 2008, S. 214-228
- [Scholz/Tietje 2002]
Scholz, R. W., Tietje, O., Embedded Case Study Methods. Integrating quantitative and qualitative Knowledge, Sage, Thousand Oaks, 2002
- [Schreiber et al. 2000]
Schreiber, G., Akkermans, H., Anjewierden, A., de Hoog, R., Shadbolt, N., de Velde, W. V., Wielinga, B., Knowledge Engineering and Management: The CommonKADS Methodology, MIT Press, Cambridge, Massachusetts, 2000
- [Schütte 1998]
Schütte, R., Grundsätze ordnungsmässiger Referenzmodellierung: Konstruktion konfigurations- und anpassungsorientierter Modelle, Gabler, Wiesbaden, Deutschland, 1998
- [Schütte/Rotthowe 1998]
Schütte, R., Rotthowe, T., The Guidelines of Modeling – An Approach to Enhance the Quality in Information Models, Proceedings of the 17th International Conference on Conceptual Modeling, Singapur, 1998, S. 240-254
- [Schwaber/Beedle 2002]
Schwaber, K., Beedle, M., Agile Software Development with Scrum, Prentice Hall, Upper Saddle River, New Jersey, 2002
- [SEI 2006a]
SEI, Appraisal Requirements for CMMI, Version 1.2 (ARC, V1.2), Carnegie Mellon University, Pittsburgh, Pennsylvania, 2006

- [SEI 2006b]
SEI, CMMI for Development, Version 1.2, CMU/SEI-2006-TR-008, Software engineering Institute Carnegie Mellon University, Pittsburgh, Pennsylvania, 2006
- [Sen 2004]
Sen, A., Metadata management: past, present and future, in: Decision Support Systems, 37, 2004, Nr. 1, S. 151-173
- [Senn 2002]
Senn, S., Cross-over Trials in Clinical Research, 2. Aufl., Wiley, Chichester, England, 2002
- [Shankaranarayanan/Cai 2006]
Shankaranarayanan, G., Cai, Y., Supporting data quality management in decision-making, in: Decision Support Systems, 42, 2006, Nr. 1, S. 302-317
- [Shankaranarayanan et al. 2000]
Shankaranarayanan, G., Wang, R. Y., Ziad, M., IP-MAP: Representing the Manufacture of an Information Product, in: Klein, B. D., Rossin, D. F. (Hrsg.), Cambridge, Massachusetts, 2000, S. 1-16
- [Shankaranarayanan et al. 2003]
Shankaranarayanan, G., Ziad, M., Wang, R. Y., Managing Data Quality in Dynamic Decision Environments: An Information Product Approach, in: Journal of Database Management, 14, 2003, Nr. 4, S. 14-32
- [Sheng 2003]
Sheng, Y. P., Exploring the Mediating and Moderating Effects of Information Quality on Firms? Endeavor on Information Systems, in: Eppler, M., Helfert, M., Pipino, L., Segev, A. (Hrsg.), Proceedings of the 8th International Conference on Information Quality, Cambridge, Massachusetts, 2003, S. 344-352
- [Sheng/Mykytyn 2002]
Sheng, Y. P., Mykytyn, P. P., Jr., Information Technology Investment and Firm Performance: A Perspective of Data Quality, in: Fisher, C., Davidson, B. N. (Hrsg.), Proceedings of the 7th International Conference on Information Quality, Cambridge, Massachusetts, 2002, S. 132-141
- [Sheu et al. 2006]
Sheu, C., Yen, H. R., Chae, B., Determinants of supplier-retailer collaboration: evidence from an international study, in: International Journal of Operations & Production Management, 26, 2006, Nr. 1, S. 24-49
- [Shewhart 1931]
Shewhart, W. A., Economic Control of Quality of Manufactured Product, Van Nostrand, New York, 1931
- [Simon 1996]
Simon, H. A., The sciences of the artificial, 3. Aufl., MIT Press, Cambridge, Massachusetts et al., 1996
- [Sinha et al. 2004]
Sinha, P. R., Whitman, L. E., Malzahn, D., Methodology to mitigate supplier risk in an aerospace supply chain, in: Supply Chain Management: An International Journal, 9, 2004, Nr. 2, S. 154-168

- [Smith/McKeen 2008]
Smith, H. A., McKeen, J. D., Developments in Practice XXX: Master Data Management: Salvation or Snake Oil?, in: Communications of the AIS, 23, 2008, Nr. 4, S. 63-72
- [Smith/Potter 2009]
Smith, M., Potter, K., Preliminary Findings: 2009 IT Spending and Staffing Report, G00164940, Gartner Research, Stamford, Connecticut, 2009
- [Souzis 2005]
Souzis, A., Building a Semantic Wiki, in: IEEE Intelligent Systems, 20, 2005, Nr. 5, S. 87-91
- [Spiegler 2000]
Spiegler, I., Knowledge management: a new idea or a recycled concept?, in: Communications of the AIS, 3, 2000, Nr. 4, S. 1-24
- [Stahlknecht/Hasenkamp 2005]
Stahlknecht, P., Hasenkamp, U., Einführung in die Wirtschaftsinformatik, 11. Aufl., Springer, Berlin, Deutschland et al., 2005
- [Stewart et al. 2007]
Stewart, D. W., Shamdasani, P. N., Rook, D. W., Focus Groups: Theory and Practice, 2. Aufl., Sage, Thousand Oaks, California, 2007
- [Strong et al. 1997a]
Strong, D. M., Lee, Y. W., Wang, R. Y., 10 Potholes in the Road to Information Quality, in: IEEE Computer, 30, 1997, Nr. 8, S. 38-46
- [Strong et al. 1997b]
Strong, D. M., Lee, Y. W., Wang, R. Y., Data Quality in Context, in: Communications of the ACM, 40, 1997, Nr. 5, S. 103-110
- [Susman/Evered 1978]
Susman, G. I., Evered, R. D., An Assessment of the Scientific Merits of Action Research, in: Administrative Science Quarterly, 23, 1978, Nr. 4, S. 582-603
- [Tayi/Ballou 1998]
Tayi, G. K., Ballou, D. P., Examining Data Quality, in: Communications of the ACM, 41, 1998, Nr. 2, S. 54-57
- [Tellkamp et al. 2004]
Tellkamp, C., Angerer, A., Fleisch, E., Corsten, D., From Pallet to Shelf: Improving Data Quality in Retail Supply Chains Using RFID, in: Cutter IT Journal, 17, 2004, Nr. 9, S. 19-24
- [Thoo et al. 2010]
Thoo, E., Friedman, T., Feinberg, D., Beyer, M. A., Hype Cycle for Data Management, 2010, G00200878, Gartner Research, Stamford, Connecticut, 2010
- [Tozer 1999]
Tozer, G., Metadata Management for Information Control and Business Success, Artech House, Boston, Massachusetts, 1999
- [Tukey 1977]
Tukey, J. W., Exploratory Data Analysis, Addison-Wesley, Reading, Massachusetts, 1977

[Tuomi 1999]

Tuomi, I., Data Is More Than Knowledge: Implications of the Reversed Knowledge Hierarchy for Knowledge Management and Organizational Memory, in: *Journal of Management Information Systems*, 16, 1999, Nr. 3, S. 103-117

[Ulrich 1984]

Ulrich, H., Die Betriebswirtschaftslehre als anwendungsorientierte Sozialwissenschaft, in: Dyllick, T., Probst, G. J. B. (Hrsg.), *Management*, Haupt, Bern, Schweiz et al., 1984, S. 168-199

[Urbach et al. 2009]

Urbach, N., Smolnik, S., Riempp, G., The State of Research on Information Systems Success, in: *Business & Information Systems Engineering*, 1, 2009, Nr. 4, S. 315-325

[Vaas 2006]

Vaas, L., IBM debuts new server type, in: *eWeek*, 23, 2006, Nr. 42, S. 36

[van der Blonk 2003]

van der Blonk, H., Writing case studies in information systems research, in: *Journal of Information Technology*, 18, 2003, Nr. 1, S. 45-52

[Vermeer 2000]

Vermeer, B. H. P. J., How Important is Data Quality for Evaluating the Impact of EDI on Global Supply Chains?, in: Sprague, R. (Hrsg.), *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, Maui, Hawaii, 2000, S. 7068

[Viégas et al. 2004]

Viégas, F. B., Wattenberg, M., Dave, K., Studying Cooperation and Conflict between Authors with history flow Visualizations, in: Dykstra-Erickson, E., Tscheligi, M. (Hrsg.), *Wien, Österreich*, 2004, S. 575 - 582

[Viscusi et al. 2009]

Viscusi, G., Maurino, A., Grega, S., An Ontology Based Approach to Data Quality Initiatives Cost-Benefit Evaluation, in: Nelson, M. L., Shaw, M. J., Strader, T. J. (Hrsg.), *Proceedings of the 15th Americas Conference on Information Systems*, San Francisco, 2009

[vom Brocke 2003]

vom Brocke, J., *Referenzmodellierung. Gestaltung und Verteilung von Konstruktionsprozessen*, Logos Verlag, Berlin, Deutschland, 2003

[vom Brocke 2007]

vom Brocke, J., Design Principles for Reference Modeling: Reusing Information Models by Means of Aggregation, Specialisation, Instantiation, and Analogy, in: Fettke, P., Loos, P. (Hrsg.), *Reference Modeling for Business Systems Analysis*, Idea Group, Hershey, Pennsylvania, 2007, S. 47-75

[vom Brocke et al. 2010]

vom Brocke, J., Simons, A., Cleven, A., Towards a Business Process-Oriented Approach to Enterprise Content Management (ECM): The ECM-Blueprinting Framework, in: *Information Systems and e-Business Management*, 2010, <http://www.springerlink.com/content/g523683j71254672/> (zuletzt geprüft: 23. November 2010)

[Wagner/Majchrzak 2007]

Wagner, C., Majchrzak, A., Enabling Customer-Centricity Using Wikis and the Wiki Way, in: *Journal of Management Information Systems*, 23, 2007, Nr. 3, S. 17-43

[Walsham 1995]

Walsham, G., Interpretive case studies in IS research: nature and method, in: *European Journal of Information Systems*, 4, 1995, Nr. 2, S. 74-81

[Wand/Wang 1996]

Wand, Y., Wang, R. Y., Anchoring Data Quality Dimensions in Ontological Foundations, in: *Communications of the ACM*, 39, 1996, Nr. 11, S. 86-95

[Wang 1998]

Wang, R. Y., A product perspective on total data quality management, in: *Communications of the ACM*, 41, 1998, Nr. 2, S. 58-65

[Wang et al. 1998]

Wang, R. Y., Lee, Y. W., Pipino, L. L., Strong, D. M., Manage Your Information as a Product, in: *Sloan Management Review*, 39, 1998, Nr. 4, S. 95-105

[Wang et al. 1995]

Wang, R. Y., Storey, V. C., Firth, C. P., A Framework for Analysis of Data Quality Research, in: *IEEE Transactions on Knowledge and Data Engineering*, 7, 1995, Nr. 4, S. 623-640

[Wang/Strong 1996]

Wang, R. Y., Strong, D. M., Beyond Accuracy: What Data Quality Means to Data Consumers, in: *Journal of Management Information Systems*, 12, 1996, Nr. 4, S. 5-33

[Wang et al. 2006]

Wang, S., Archer, N. P., Zheng, W., An Exploratory Study of Electronic Marketplace Adoption: A Multiple Perspective View, in: *Electronic Markets*, 16, 2006, Nr. 4, S. 337-348

[Weber 2009]

Weber, K., Data Governance-Referenzmodell: Organisatorische Gestaltung des unternehmensweiten Datenqualitätsmanagements, Dissertation, Institut für Wirtschaftsinformatik Universität St. Gallen, St. Gallen, Schweiz, 2009

[Weber et al. 2009a]

Weber, K., Otto, B., Österle, H., One Size Does Not Fit All—A Contingency Approach to Data Governance, in: *ACM Journal of Data and Information Quality*, 1, 2009, Nr. 1, S. 4:1-4:27

[Weber et al. 2009b]

Weber, K., Schmidt, A., Hüner, K. M., Ofner, M., Data Quality Management – Definitions and State of the Art, BE HSG / CC CDQ / 1, Institut für Wirtschaftsinformatik Universität St. Gallen, St. Gallen, Schweiz, 2009

[Wedekind 2001]

Wedekind, H., Stammdaten, in: Mertens, P., Back, A., Becker, J., König, W., Krallmann, H., Rieger, B., Scheer, A.-W., Seibt, D., Stahlknecht, P., Strunz, H., Thome, R., Wedekind, H. (Hrsg.), *Lexikon der Wirtschaftsinformatik*, 4. Aufl., Springer, Berlin, Deutschland et al., 2001, S. 72

[Wende 2007]

Wende, K., A Model for Data Governance – Organising Accountabilities for Data Quality Management, in: Toleman, M., Cater-Steel, A., Roberts, D. (Hrsg.), Proceedings of 18th Australasian Conference on Information Systems, Toowoomba, Australien, 2007, S. 417-425

[White et al. 2006]

White, A., Newman, D., Logan, D., Radcliffe, J., Mastering Master Data Management, G00136958, Gartner Research, Stamford, Connecticut, 2006

[White et al. 2008]

White, C., Plotnick, L., Addams-Moring, R., Turoff, M., Hiltz, S. R., Leveraging A Wiki To Enhance Virtual Collaboration In The Emergency Domain, in: Sprague, R. H., Jr. (Hrsg.), Proceedings of the 41st Hawaii International Conference on System Sciences, Waikoloa, Hawaii, 2008, S. 322

[Wilde/Hess 2007]

Wilde, T., Hess, T., Forschungsmethoden der Wirtschaftsinformatik. Eine empirische Untersuchung, in: WIRTSCHAFTSINFORMATIK, 49, 2007, Nr. 4, S. 280-287

[Winer et al. 1991]

Winer, B. J., Brown, D. R., Michels, K. M., Statistical Principles In Experimental Design, 2. Aufl., McGraw-Hill, Boston, Massachusetts, 1991

[Winkler 2004]

Winkler, W. E., Methods for evaluating and creating data quality, in: Information Systems, 29, 2004, Nr. 7, S. 531-550

[Winter 2008]

Winter, R., Design Science Research in Europe, in: European Journal of Information Systems, 17, 2008, Nr. 5, S. 470-476

[WKWI/GI-FB WI 2008]

WKWI/GI-FB WI, WI-Orientierungslisten, in: WIRTSCHAFTSINFORMATIK, 50, 2008, Nr. 2, S. 155-163

[Yang et al. 2008]

Yang, D., Wu, D., Koolmanojwong, S., Brown, A. W., Boehm, B. W., WikiWinWin: A Wiki Based System for Collaborative Requirements Negotiation, in: Sprague, R. H., Jr. (Hrsg.), Proceedings of the 41st Hawaii International Conference on System Sciences, Waikoloa, Hawaii, 2008, S. 24

[Yen 2004]

Yen, B. K., Pulling It All Together, in: Financial Executive, 20, 2004, Nr. 6, S. 60-62

[Yin 2002]

Yin, R. K., Case study Research: Design and Methods, 3. Aufl., Sage, Thousand Oaks, 2002

[Zhao 2007]

Zhao, H., Semantic Matching Across Heterogeneous Data Sources, in: Communications of the ACM, 50, 2007, Nr. 1, S. 44-50

Anhang A Komplette Publikationsliste des Autors

Tabelle A-1 listet die Publikationen der kumulativen Dissertation auf. Tabelle A-2 zeigt weitere Publikationen des Autors. Die Spalte *Typ* gibt neben dem Publikationstyp die von der Wissenschaftlichen Kommission Wirtschaftsinformatik im Verband der Hochschullehrer für Betriebswirtschaft e. V. (WKWI) und dem Fachbereich Wirtschaftsinformatik der Gesellschaft für Informatik (GI-FB WI) angegebene Kategorie der jeweiligen Zeitschrift oder des jeweiligen Konferenzbands an [WKWI/GI-FB WI 2008]. Kategorie A repräsentiert die höchste Bewertung. Für Zeitschriften sind die Kategorien A und B definiert, für Konferenzbände die Kategorien A (Annahmequote $\leq 30\%$), B (Annahmequote $\leq 50\%$) und C (Annahmequote $\leq 70\%$).

Typ	Publikation	Status
K (B ³³)	Hüner, K. M., Otto, B., The Effect of Using a Semantic Wiki for Metadata Management: A Controlled Experiment, in: Sprague, R. H. (Hrsg.): Proceedings of the 42nd Hawaii International Conference on System Sciences, Waikoloa, Hawaii, 2009	Veröffentlicht
K (A ³⁴)	Hüner, K. M., Ofner, M., Otto, B., Towards a Maturity Model for Corporate Data Quality, in: Shin, D. (Hrsg.): Proceedings of the 2009 ACM Symposium on Applied Computing, Waikiki Beach, Hawaii, 2009	Veröffentlicht
K	Otto, B., Hüner, K. M., Österle, H., Identification of Business Oriented Data Quality Metrics, in: Bowen, P., Elmagarmid, A. K., Österle, H., Sattler, K.-U. (Hrsg.): Proceedings of the 14th International Conference on Information Quality, Potsdam, Deutschland, 2009	Veröffentlicht
K (B ³⁵)	Otto, B., Hüner, K. M., Österle, H., A Cybernetic View on Data Quality Management, in: Leidner, D. E., Elam, J. J. (Hrsg.): Proceedings of the 16th Americas Conference on Information Systems, Lima, Peru, 2010	Veröffentlicht
AB	Hüner, K. M., Methode zur Spezifikation geschäftsorientierter Datenqualitätskennzahlen, BE HSG / CC CDQ / 13, Institut für Wirtschaftsinformatik der Universität St. Gallen, St. Gallen, Schweiz, 2011	Veröffentlicht
Z (A)	Hüner, K. M., Otto, B., Österle, H., Collaborative management of business metadata, in: International Journal of Information Management, 31, 2011, Nr. 4, S. 366-373	Veröffentlicht
Z (A)	Hüner, K. M., Schierning, A., Otto, B., Österle, H., Product data quality in supply chains: the case of Beiersdorf, in: Electronic Markets, 21, 2011, Nr. 2, S. 141-154	Veröffentlicht
Z (B)	Otto, B., Hüner, K. M., Österle, H., Toward a functional reference model for master data quality management, in: Information Systems and e-Business Management, 2011, doi: 10.1007/s10257-011-0178-0	Veröffentlicht
Z (A)	Hüner, K. M., Otto, B., Österle, H., Specifying data quality metrics that matter to business, in: Journal of the Association for Information Systems, 2011	In zweiter Begutachtung

AB: Publizierter Arbeitsbericht, K: Publikation in Konferenzband, Z: Publikation in Zeitschrift

Tabelle A-1: Publikationen der kumulativen Dissertation

³³ Annahmequote des Tracks *Knowledge Management Systems*: 43%.

³⁴ Annahmequote des Tracks *Organizational Engineering*: 28%.

³⁵ Annahmequote der Konferenz (für Track *Data and Information Quality* unbekannt): 46%.

Typ	Publikation	Status
BB	Hüner, K. M., Wende, K., INTERSPORT Schweiz AG: SAP-Einführung mit ExpertRetail, in: Wölfle, R., Schubert, P. (Hrsg.): Business Collaboration – Standortübergreifende Prozesse mit Business Software, Hanser, München, Deutschland, 2007, S. 99-111	Veröffentlicht
Z	Hüner, K. M., Otto, B., Österle, H., Metadatenmanagement mit einem Semantischen Wiki, in: IM – Fachzeitschrift für Information Management & Consulting, 23, 2008, Nr. 2, S. 34-40	Veröffentlicht
BB	Otto, B., Wende, K., Schmidt, A., Hüner, K. M., Vogel, T., Unternehmensweites Datenqualitätsmanagement: Ordnungsrahmen und Anwendungsbeispiele, in: Dinter, B., Winter, R. (Hrsg.): Integrierte Informationslogistik, Springer, Berlin, Deutschland, 2008, S. 211-230	Veröffentlicht
K	Otto, B., Hüner, K. M., A Meta-model for Data Quality Management Simulation, in: Bowen, P., Elmagarmid, A. K., Österle, H., Sattler, K.-U. (Hrsg.): Proceedings of the 14th International Conference on Information Quality, Potsdam, Deutschland, 2009	Veröffentlicht
AB	Wende, K., Schmidt, A., Hüner, K. M., Ofner, M., Data Quality Management – Definitions and State of the Art, BE HSG / CC CDQ / 1, Institut für Wirtschaftsinformatik Universität St. Gallen, St. Gallen, 2009	Veröffentlicht
K (C ³⁶)	Ofner, M. H., Hüner, K. M., Otto, B., Dealing with Complexity: A Method to Adapt and Implement a Maturity Model for Corporate Data Quality Management, in: Nelson, M. L., Shaw, M. J., Strader, T. J. (Hrsg.): Proceedings of the 15th Americas Conference on Information Systems, San Francisco, California, 2009	Veröffentlicht
K	Cleven, A., Gubler, P., Hüner, K. M., Design Alternatives for the Evaluation of Design Science Research Artifacts, Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology, Malvern, Pennsylvania, 2009	Veröffentlicht
AB	Otto, B., Hüner, K. M., Funktionsarchitektur für unternehmensweites Stammdatenmanagement, BE HSG / CC CDQ / 14, Institut für Wirtschaftsinformatik Universität St. Gallen, St. Gallen, 2009	Veröffentlicht
Z	Otto, B.; Hüner, K. M.; Österle, H., Unternehmensweite Stammdatenqualität, in: ERP Management, 2009, 2009, Nr. 3, S. 19-21.	Veröffentlicht
F	Ebner, V., Hüner, K. M., Otto, B., Case Study Bayer CropScience AG – Master Data Quality Cockpit, BE HSG / CC CDQ / 23, Institut für Wirtschaftsinformatik Universität St. Gallen, St. Gallen, 2011	Veröffentlicht
Z (B)	Hüner, K. M., Brauer, B., Otto, B., Österle, H., Fachliches Metadatenmanagement mit einem semantischen Wiki, in: HMD – Praxis der Wirtschaftsinformatik, 46, 2010, Nr. 6	Veröffentlicht
F	Schmidt, A., Hüner, K. M., Grewe, A., Fallstudie Deutsche Telekom AG – Einheitliche Datenarchitektur als Grundlage für unternehmensweites Datenqualitätsmanagement, BE HSG / CC CDQ / 23, Institut für Wirtschaftsinformatik Universität St. Gallen, St. Gallen, 2010	Veröffentlicht

AB: Publizierter Arbeitsbericht, BB: Buchbeitrag, F: Publierte Fallstudie,
K: Publikation in Konferenzband, Z: Publikation in Zeitschrift

Tabelle A-2: Weitere Publikationen

³⁶ Annahmequote der Konferenz (für Track *Data and Information Quality* unbekannt): 64%.

Anhang B Gestaltungsprozesse

Die Tabellen dieses Anhangs beschreiben die Gestaltungsprozesse der in der vorliegenden Dissertation vorgestellten Artefakte. Dabei werden die in den jeweiligen Phasen angewendeten Techniken und erzeugten Ergebnisse aufgeführt. Neben dem Techniktyp beschreiben die Tabellen auch die konkrete Anwendung der Techniken, also z. B. Datum und Teilnehmer eines Fokusgruppeninterviews oder Dauer und Unternehmen einer partizipativen Fallstudie.

Phase	Techniken		Ergebnis
Analyse	Fokusgruppeninterview	Identifikation des Bedarfs an geschäftsorientierten DQ-Kennzahlen (2. April 2008, 18 Fachexperten aus 10 Unternehmen) Identifikation des Bedarfs an einer Methode zur Spezifikation unternehmensspezifischer DQ-Kennzahlen (11. Juni 2008, 18 Fachexperten aus 10 Unternehmen)	<i>Forschungsziel:</i> Gestaltung einer Methode a) zur Identifikation von Kausalitäten zwischen Datendefekten, Geschäftsproblemen und strategischen Zielen und b) zur Spezifikation von DQ-Kennzahlen für die Überwachung der identifizierten Datendefekte
	Literaturrecherche	Analyse von 427 wissenschaftlichen Publikation (vgl. Kapitel B-2) und Publikationen von Beratungs- und Berufserfahrung (vgl. Kapitel B-3).	
Gestaltung	Methoden-Engineering	Spezifikation der Ergebnisdokumente, des Metamodells, der Rollen, der Techniken und des Vorgehensmodells [Gutzwiller 1994, S. 11-15]	<i>Methodenentwurf:</i> Version der Kennzahlenmethode nach der jeweiligen Gestaltungsiteration
	Fokusgruppeninterview	Identifikation von Anforderungen an geschäftsorientierte DQ-Kennzahlen (11. Juni 2008, s. o.)	
	Partizipative Fallstudie	Methodenanwendung bei Deutsche Telekom AG (Juli 2007 bis September 2007) Methodenanwendung bei ETA S.A. (August 2007 bis März 2008) Methodenanwendung bei Robert Bosch GmbH (März 2008 bis Juni 2008)	
Evaluation	Fokusgruppeninterview	Diskussion eines Grobkonzepts (Aktivitäten, Metamodell, Rollen) der Methode (4. September 2008, 20 Fachexperten aus 11 Unternehmen) Diskussion des Metamodells, der Rollen und des Vorgehensmo-	<i>Methode:</i> Version der Kennzahlenmethode nach Abschluss der letzten angegebenen Evaluation im Juni 2010

Phase	Techniken		Ergebnis
		<p>dells der Methode (18. Februar 2009, 13 Fachexperten aus 7 Unternehmen)</p> <p>Diskussion der Konzeptualisierung des Geschäftsbezugs von DQ-Kennzahlen durch Kausalketten (23. April 2009, 12 Fachexperten aus 7 Unternehmen)</p> <p>Diskussion der Technik Datenqualitätssimulation (24. Juni 2009, 11 Fachexperten aus 9 Unternehmen)</p> <p>Diskussion der gesamten Methode (9. Februar 2010, 11 Fachexperten aus 10 Unternehmen)</p>	
	Partizipative Fallstudie	<p>Methodenanwendung bei Deutsche Telekom AG (Juni 2009 bis Februar 2010)</p> <p>Methodenanwendung bei Elektrizitätswerke des Kantons Zürich (November 2009 bis Juni 2010)</p> <p>Methodenanwendung bei Beiersdorf AG (Januar 2010 bis Juni 2010)</p>	
Diffusion	Publikation in der Praxis		[Hüner 2011] s. Kapitel B-3
	Wissenschaftliche Publikation		<p>[Otto et al. 2009a] s. Kapitel B-1</p> <p>[Otto et al. 2010] s. Kapitel B-5</p> <p>[Hüner et al. 2011c] s. Kapitel B-4</p> <p>[Hüner et al. 2011d] s. Kapitel B-2</p>

Tabelle B-1: Gestaltungsprozess der Kennzahlenmethode

Phase	Techniken		Ergebnis
Analyse	Fokusgruppeninterview	Identifikation des Bedarfs an einem Reifegradmodell für CDQM (25. Juni 2007, 10 Fachexperten aus 5 Unternehmen)	<i>Forschungsziel:</i> Gestaltung eines Referenzmodells zur Bewertung des Reifegrads von CDQM in einem Unternehmen
	Literaturrecherche	Analyse von Publikationen zu MDM- und DQM-Referenzmodellen von Analysten- und Beratungsunternehmen und von wissenschaftlichen Publikationen zu Reifegradmodellen (vgl. Kapitel B-1)	
Gestaltung	Referenzmodellierung	Analyse und Wiederverwen-	<i>Referenzmodellentwurf:</i>

Phase	Techniken		Ergebnis
		<p>dung ähnlicher Referenzmodelle (vgl. Kapitel B-1)</p> <p>Gestaltungsprinzipien bei der Referenzmodellierung [vom Brocke 2007, S. 51-68]</p> <p>Funktionsdekomposition [Scheer 1997, S. 19-23, Probst 2003, S. 253-256]</p>	Version des Reifegradmodells nach der jeweiligen Gestaltungsiteration
	Fokusgruppeninterview	Identifikation von Erfolgsfaktoren für CDQM (2. April 2008, 20 Fachexperten aus 12 Unternehmen)	
Evaluation	Fokusgruppeninterview	Diskussion des CDQM-Ordnungsrahmens und eines Bewertungsmodells (11. Juni 2008, 18 Fachexperten aus 11 Unternehmen)	<i>Referenzmodell:</i> Version des Reifegradmodells nach Abschluss der letzten angegebenen Evaluation im September 2008
	Partizipative Fallstudie	<p>Anwendung des Reifegradmodells bei Robert Bosch GmbH (März 2008 bis April 2008)</p> <p>Anwendung des Reifegradmodells bei ZF Friedrichshafen AG (Juli 2008 bis September 2008)</p>	
Diffusion	Publikation in der Praxis		[EFQM 2011]
	Wissenschaftliche Publikation		<p>[Hüner et al. 2009] s. Kapitel B-6</p> <p>[Ofner et al. 2009]</p>

Tabelle B-2: Gestaltungsprozess des Reifegradmodells

Phase	Techniken		Ergebnis
Analyse	Fokusgruppeninterview	Identifikation des Bedarfs an Applikationsunterstützung für kollaboratives fachliches Metadatenmanagement (15. November 2007, 13 Fachexperten aus 9 Unternehmen)	<i>Forschungsziel:</i> Prototypische Entwicklung eines Wiki-basierten fachlichen Metadatenkatalogs
	Literaturrecherche	Analyse von Publikationen zu Produkten für fachliches Metadatenmanagement von Software-Herstellern und von wissenschaftlichen Publikationen zur Nutzung von Wikis in Unternehmen (vgl. Kapitel B-8)	
Gestaltung	Software-Engineering	<p>Agile Software-Entwicklung [Schwaber/Beedle 2002]</p> <p>CommonKADS-Methode [Schreiber et al. 2000]</p>	<i>Prototypentwurf:</i> Version des BDD-Wikis nach der jeweiligen Gestaltungsiteration

Phase	Techniken		Ergebnis
	Fokusgruppeninterview	Identifikation von Anforderungen an ein BDD (16. November 2007, 14 Fachexperten aus 10 Unternehmen)	
	Kontrolliertes Experiment	Bearbeitung von Aufgaben (Metadatensuche und -pflege) in einem semantischen und einem normalen Wiki durch 62 Fachexperten	
Evaluation	Fokusgruppeninterview	Demonstration und Diskussion des Prototypen (3. Dezember 2008, 9 Fachexperten aus 7 Unternehmen)	<i>Prototyp</i> : Version des BDD-Wikis nach Abschluss der letzten angegebenen Evaluation im Dezember 2008
	Funktionstest	Test und Bewertung des Prototypen nach der Bearbeitung von Aufgaben in 3 Szenarien durch 5 Mitarbeiter von Bayer CropScience AG (September 2008)	
Diffusion	Publikation in der Praxis		[Hüner et al. 2008] [Hüner et al. 2011a]
	Wissenschaftliche Publikation		[Hüner/Otto 2009] s. Kapitel B-7 [Hüner et al. 2011b] s. Kapitel B-8

Tabelle B-3: Gestaltungsprozess des BDD-Wikis

Phase	Techniken		Ergebnis
Analyse	Partizipative Fallstudie	Einführung des Anwendungssystems <i>NetWeaver MDM</i> von SAP bei ETA S.A. (November 2007 bis März 2008)	<i>Forschungsziel</i> : Gestaltung eines fachkonzeptionellen Referenzmodells der Funktionssicht auf ein IS zur Unterstützung von CDQM
	Fokusgruppeninterview	Identifikation des Bedarfs an einer Übersicht zu Applikationsfunktionen für CDQM (20. September 2007, 14 Fachexperten aus 7 Unternehmen)	
	Produktevaluation	Analyse angebotener Software-Produkte zur Unterstützung von CDQM von IBM, Oracle, SAP und Tibco (Oktober 2007, nicht abgeschlossen, da kein Raster für eine vergleichende Bewertung verfügbar gewesen ist)	
Gestaltung	Referenzmodellierung	Fachkonzeptionelle Modellierung von IS [Scheer 1991, S. 61-135, Scheer 1997, S. 19-64] Funktionale Dekomposition	<i>Referenzmodellentwurf</i> : Version des Funktionsreferenzmodells nach der jeweiligen Gestaltungsiteration

Phase	Techniken		Ergebnis
		Funktionsdekomposition [Scheer 1997, S. 19-23, Probst 2003, S. 253-256]	
	Literaturrecherche ³⁷	Analyse von Publikationen von Analysten- und Bera- tungsunternehmen und Soft- ware-Herstellern (vgl. Kapitel B-9)	
Evaluation	Fokusgruppeninterview	Diskussion der ersten Version des Funktionsreferenzmodells (25. November 2008, 17 Fachexperten aus 17 Unter- nehmen) Diskussion der zweiten Ver- sion des Funktionsreferenz- modells (9. Februar 2009, 5 Fachexperten aus 5 Unter- nehmen) Diskussion der dritten Versi- on des Funktionsreferenzmo- dells (18. Februar 2009, 9 Fachexperten aus 8 Unter- nehmen)	<i>Referenzmodell</i> : Version des Funktionsreferenzmo- dells nach Abschluss der letzten angegebenen Evalu- ation im Mai 2010
	Produktanalyse	Analyse angebotener Soft- ware-Produkte zur Unterstüt- zung von CDQM von IBM, Oracle, SAP und Tibco (April 2009)	
	Partizipative Fallstudie	Anwendung des Funktionsre- ferenzmodells bei Beiersdorf AG (April 2010 bis Mai 2010)	
Diffusion	Publikation in der Praxis		[Otto/Hüner 2009] [Otto et al. 2009b]
	Wissenschaftliche Publikation		[Otto et al. 2011] s. Kapitel B-9

Tabelle B-4: Gestaltungsprozess des Funktionsreferenzmodells

³⁷ Die Literaturrecherche ist in diesem Fall als Gestaltungstechnik klassifiziert, da verschiedene Analysen in den einzelnen Gestaltungsiterationen durchgeführt und die Ergebnisse (u. a. Bezeichner) direkt bei der Gestaltung verwendet worden sind.

Anhang C Erweiterung des BE-Methodenkerns

Tabelle C-1 definiert die Metaentitätstypen, die neben den Entitätstypen des Metamodells des BE-Methodenkerns [Höning 2009, S. 252-257] und des Metamodells der Methode zur Stammdatenintegration [Schmidt 2010, S. 221-225] die vorgestellten Artefakte beschreiben (vgl. Kapitel A-3).

Metaentitäts-typ	Beschreibung und Relationen
CDQM-Prozess	<p>Ein CDQM-Prozess ist ein Unterstützungsprozess mit dem Ziel, die Qualität von Konzerndaten zu sichern und zu verbessern. Ein CDQM-Prozess umfasst Aufgaben aus den Bereichen Datenmanagement, DQM, MDM und Metadatenmanagement [Otto et al. 2007, Schmidt 2010, S. 21-24, EFQM 2011].</p> <p>Beziehungen Ein CDQM-Prozess ist ein spezieller Unterstützungsprozess. Ein CDQM-Prozess erzeugt Datenqualität als Prozessleistung. Ein CDQM-Prozess hat hohe Datenqualität als Ziel.</p>
CDQM-Erfolgsfaktor	<p>CDQM-Erfolgsfaktoren beschreiben Charakteristika eines Unternehmens, die den Erfolg der CDQM-Funktion des Unternehmens gewährleisten und somit die Erzeugung hoher Datenqualität als Ziel von CDQM unterstützen.</p> <p>Beziehungen Ein CDQM-Erfolgsfaktor ist ein spezieller Erfolgsfaktor. CDQM-Erfolgsfaktoren bestimmen die CDQM-Reifegradskala.</p>
CDQM-Informationssystem	<p>Ein CDQM-IS umfasst Applikationen, deren Applikationsfunktionen Mitarbeiter einer CDQM-Funktion bei der Ausführung von CDQM-Aktivitäten (z. B. Messen von Datenqualität, Modellierung von Datenobjekttypen) unterstützen.</p> <p>Beziehungen Ein CDQM-IS ist ein spezielles IS. CDQM-IS unterstützen CDQM.</p>
CDQM-Reifegradskala	<p>Eine CDQM-Reifegradskala ermöglicht die quantitative Bewertung von Strukturen einer CDQM-Funktion. Bei einer Reifegradmessung bewerten Mitarbeiter des Unternehmens oder externe Assessoren, wie ausgeprägt bestimmte Strukturen vorhanden sind. So kann ein Unternehmen den Fortschritt des Aufbaus oder der Umstrukturierung seiner CDQM-Funktion bewerten oder eigene Strukturen mit anderen Unternehmen vergleichen. Der CDQM-Ordnungsrahmen beschreibt die bewerteten Strukturen als Erfolgsfaktoren für CDQM.</p> <p>Beziehungen Eine CDQM-Reifegradskala ist eine spezielle Kennzahl. CDQM-Reifegradskalen können hierarchisch strukturiert sein.</p>
Datenqualität	<p>Datenqualität ist ein durch einen Nutzer von Daten wahrgenommenes Phänomen. Datenqualität bewertet die Erfüllung von Anforderungen, die ein Datennutzer an Daten stellt [Wang/Strong 1996, S. 6]. Ein Datennutzer kann bei der Ausführung von Aktivitäten unterschiedliche Anforderungen an ein (gleiches) Datenelement stellen (Kontextabhängigkeit).</p> <p>Beziehungen Datenqualität ist die Leistung eines CDQM-Prozesses. Datenqualität ist ein Bestandteil der Prozessleistung. Datenqualität bewertet Datenelemente und Metadaten.</p>
DQ-Dimension	<p>DQ-Dimensionen repräsentieren bestimmte Aspekte (z. B. Aktualität, Glaubwürdigkeit, Vollständigkeit) des Phänomens Datenqualität [Wang/Strong 1996, S. 6]</p>

Metaentitäts- typ	Beschreibung und Relationen
	<p>und konzeptualisieren Datenqualität durch die Beschreibung dieser Aspekte.</p> <p>Beziehungen Eine DQ-Dimension ist ein spezieller Erfolgsfaktor. DQ-Dimensionen können hierarchisch strukturiert sein. DQ-Dimensionen konzeptualisieren Datenqualität.</p>
DQ-Kennzahl	<p>Eine DQ-Kennzahl ist ein quantitatives Mass für Datenqualität. Das Messsystem der Kennzahl misst einen Messwert (z. B. durch die Überprüfung von Datenelementen mit Validierungsregeln) an einem Messpunkt (z. B. ein Datenbehälter) zu einem durch eine Messfrequenz definierten Zeitpunkt. Der Messwert wird durch die Abbildung der Messung auf einer Messskala und die Zuweisung einer Masseneinheit dargestellt [IEEE 1998, Batini/Scannapieco 2006, S. 19, ISO/IEC 2007].</p> <p>Beziehungen Eine DQ-Kennzahl ist eine spezielle Kennzahl. DQ-Kennzahlen können hierarchisch strukturiert sein. DQ-Kennzahlen operationalisieren DQ-Dimensionen.</p>
Konzerndaten- klasse	<p>Eine Konzerndatenklasse gruppiert Konzerndatenobjekte des gleichen Typs (z. B. Kundendaten, Produktdaten, Vertragsdaten).</p> <p>Beziehungen Eine Konzerndatenklasse ist ein spezieller Datenobjekttyp. Konzerndatenklassen können hierarchisch strukturiert sein.</p>
Konzernda- tenobjekt	<p>Konzerndatenobjekte (kurz: Konzerndaten, z. B. Kontodaten, Produktdaten, Rechnungsdaten) sind Datenobjekte, die Geschäftsobjekte (z. B. Konten, Kunden, Produkte, Rechnungen) und betriebliche Vorgänge (z. B. Adressänderung, Kontoüberweisung, Lieferung, Reservierung) eines Unternehmens repräsentieren, und die durch Datenobjekttypen und Metadaten für mehrere Applikationen, Geschäftsprozesse und Organisationseinheiten (meist unternehmens- oder konzernweit) einheitlich und unmissverständlich spezifiziert sind (oder sein sollten).</p> <p>Beziehungen Ein Konzerndatenobjekt ist ein spezielles Datenobjekt.</p>
präventive CDQM Mass- nahme	<p>Eine präventive CDQM-Massnahme (z. B. Spezifikation und Nutzung von DQ-Kennzahlen) senkt die Wahrscheinlichkeit, dass bestimmte Datendefekte auftreten oder mindert die Wirkung bestimmter Datendefekte [Otto et al. 2010].</p> <p>Beziehungen Eine präventive CDQM-Massnahme ist eine spezielle Massnahme. Eine präventive CDQM-Massnahme gestaltet Strukturen einer CDQM-Funktion.</p>
reaktive CDQM- Massnahme	<p>Eine reaktive CDQM-Massnahme (z. B. Identifizieren und Entfernen von Dubletten) behebt Datendefekte nach deren Auftreten [Otto et al. 2010].</p> <p>Beziehungen Eine reaktive CDQM-Massnahme ist eine spezielle Massnahme. Eine reaktive CDQM-Massnahme verbessert Datenqualität.</p>
Unterstüt- zungsprozess	<p>Ein Unterstützungsprozess (z. B. Beschaffung, IT, Personalentwicklung) unterstützt Primäraktivitäten (z. B. Marketing, Produktion, Vertrieb) bei der Leistungserstellung durch die Bereitstellung benötigter Ressourcen (z. B. Infrastruktur, Material, Personal) und Managementstrukturen (z. B. Finanzplanung, juristische Beratung, Qualitätsmanagement) [Porter 1985, S. 37-48].</p> <p>Beziehungen Eine Unterstützungsprozess ist ein spezieller Geschäftsprozess.</p>

Tabelle C-1: Definitionen ergänzter Metaentitätstypen

Anhang D Anhänge zu Publikation B-2

Research activity	Focus group and interview outcome
Problem identification, definition of objectives	<p>April 2, 2008 (FG.A, see Figure B2-1): The focus group identifies data quality metrics as a core component of DQM. BASF SE, Bayer CropScience AG, Robert Bosch GmbH, Daimler AG, DB Netz AG, Deutsche Post AG, Deutsche Telekom AG, E.ON AG, ETA S.A., ZF Friedrichshafen AG</p> <p>June 11, 2008 (FG.B): The focus group determines the need for a structured procedure to specify data quality metrics and defines design objectives. B. Braun Melsungen AG, Bayer CropScience AG, Beiersdorf AG, Daimler AG, DB Netz AG, Deutsche Telekom AG, E.ON AG, ETA S.A., Nestlé S.A., Tchibo GmbH</p>
Design	<p>June 11, 2008 (FG.B): The focus group defines requirements for data quality metrics (cf. Activity II.1). B. Braun Melsungen AG, Bayer CropScience AG, Beiersdorf AG, Daimler AG, DB Netz AG, Deutsche Telekom AG, E.ON AG, ETA S.A., Nestlé S.A., Tchibo GmbH</p>
Demonstration, evaluation	<p>September 4, 2008 (FG.C): The focus group discusses a basic concept of the method (i.e. concepts of method engineering, preliminary procedure model, meta-model and roles). Bayer CropScience AG, Robert Bosch GmbH, Ciba AG, Daimler AG, DB Netz AG, Deutsche Telekom AG, E.ON AG, ETA S.A., Nestlé S.A., WestLB AG, ZF Friedrichshafen AG</p> <p>February 18, 2009 (FG.D): The focus group discusses and evaluates the procedure model, the meta-model, and the roles of the method. ABB Ltd., B. Braun Melsungen AG, Bayer CropScience AG, Beiersdorf AG, DB Netz AG, Nestlé AG, Syngenta AG</p> <p>April 23, 2009 (FG.E): The focus group discusses and evaluates the concept of business orientation (i.e. causality chains) and its consideration by the method. Bayer CropScience AG, Beiersdorf AG, Corning, Inc., DB Netz AG, Hilti AG, Novartis AG, Syngenta AG</p> <p>June 24, 2009 (FG.F): The focus group discusses and evaluates the technique Data Quality Simulation Bayer CropScience AG, Beiersdorf AG, DB Netz AG, Deutsche Telekom AG, Kühne + Nagel International AG, Migros-Genossenschaftsbund, Novartis AG, PostFinance, Syngenta AG</p> <p>February 9, 2010 (FG.G): The focus group discusses and evaluates the entire method. Bayer CropScience AG, Beiersdorf AG, DB Netz AG, Festo AG & Co. KG, GRUNDFOS A/S, Henkel AG & Co. KGaA, Metro Group AG, Rewe-Zentral AG, Siemens Enterprise Communications GmbH & Co. KG, Syngenta AG</p>

Table D-1: Focus groups participating in the design of the method

ID	Artifact	Design step	Argument
Initial design rationales			
DR01	Meta-model	Definition of entities data, data quality metric, process activity and process metric.	<ul style="list-style-type: none"> • Design objectives DO2 and DO5. • Literature: [Martin 1975], [IEEE 1990], [Wand/Wang 1996], [Wang/Strong 1996], [IEEE 1998], [Price/Shanks, 2005], [ISO/IEC 2007] and [Caballero et al. 2007].
DR02	Procedure model	Definition of activity specify data quality metrics.	<ul style="list-style-type: none"> • Design objectives DO1 and DO5. • Literature: [Strong et al. 1997], [Kahn et al. 2002], [Lee et al. 2002], [Winkler 2004], [Batini/Scannapieco 2006] and [Kaiser et al. 2007].
DR03	Roles	Definition of role chief data steward.	<ul style="list-style-type: none"> • Design objective DO1. • Literature: [Weber et al. 2009].
DR04	Documentation model	Definition of result document list of data quality metric specifications.	<ul style="list-style-type: none"> • Design objective DO1.
DR05	Techniques	Definition of technique conceptual design documentation.	<ul style="list-style-type: none"> • Design objective DO1.
Design rationales during preparation of case study 1 (CS.1, see Figure B2-1)			
DR06	Meta-model	Definition of entities business problem and data defect.	<ul style="list-style-type: none"> • Design objective DO7.
DR07	Procedure model	Definition of activity identify causal chains in front of activity specify data quality metrics.	<ul style="list-style-type: none"> • Design objective DO1. • Literature: [Batini et al. 2007] and [Caballero et al. 2008].
DR08	Documentation model	Definition of result document list with causal chains.	<ul style="list-style-type: none"> • Design objective DO1.
DR09	Roles	Definition of role data user.	<ul style="list-style-type: none"> • Design objective DO3 and DO7.
DR10	Techniques	Definition of technique workshop, and preparation of an interview guideline for the identification of causal chains.	<ul style="list-style-type: none"> • Design objective DO7. • Literature: [Wang/Strong 1996], [Lee et al. 2002], [Nicolaou/McKnight 2006] and [Price et al. 2008].
Design decisions upon finalization of CS.1			
DR11	Procedure model	Definition of activity define the scope.	<ul style="list-style-type: none"> • Experiences from CS.1. • Literature: [Even/Kaiser 2009].
DR12	Roles	Definition of role process owner.	<ul style="list-style-type: none"> • Design objective DO1.
DR13	Documentation model	Definition of result document list of scopes.	<ul style="list-style-type: none"> • Experiences from CS.1.
DR14	Techniques	Definition of technique interview and preparation of an interview guideline for the identification of appropriate scopes.	<ul style="list-style-type: none"> • Experiences from CS.1. • Literature: [Lee et al. 2002], [Nicolaou/McKnight 2006] and [Price et al. 2008].
DR15	Roles	Definition of role sponsor.	<ul style="list-style-type: none"> • Experiences from CS.1.
Design decisions during preparation of CS.2			

ID	Artifact	Design step	Argument
DR16	Procedure model	Definition of activity identify data and IT systems.	<ul style="list-style-type: none"> Design objective DO05.
DR17	Documentation model	Definition of result document list of data and IT systems.	<ul style="list-style-type: none"> Design objective DO1.
DR18	Techniques	Preparation of an interview guideline for technique interview for the identification of data and IT systems used in the focused process activities.	<ul style="list-style-type: none"> Design objective DO03.
DR19	Roles	Definition of roles business data steward and technical data steward.	<ul style="list-style-type: none"> Experiences from CS.2. Literature: [Weber et al. 2009].
DR20	Meta-model, documentation model	Extension of entity data quality metric by attributes measuring point, measuring technique, scale type, unit of measurement and measuring frequency, and appropriate extension of result document list of data quality metrics.	<ul style="list-style-type: none"> Design objective DO05. Literature: [Batini/Scannapieco 2006].
DR21	Procedure model	Definition of activity document metrics specifications.	<ul style="list-style-type: none"> Design objective DO05.
DR22	Documentation model	Definition of result document conceptual design of data quality metrics.	<ul style="list-style-type: none"> Design objective DO1.
DR23	Techniques	Extension of technique documentation by a conceptual design template for business-oriented data quality metrics.	<ul style="list-style-type: none"> Design objectives DO3 and DO05.
Design decisions upon finalization of CS.2			
DR24	Meta-model	Definition of entities preventive initiative and reactive initiative, and appropriate extension of result document list of causal chains for (optional) documentation of discussed DQM initiatives.	<ul style="list-style-type: none"> Experiences from CS.2.
DR25	Techniques	Definition of technique data quality analysis.	<ul style="list-style-type: none"> Literature: [Winkler 2004] and [Alur et al. 2007].
DR26	Techniques	Extension of technique interview/workshop by a repository of cases (i.e. data quality metrics and causal chains).	<ul style="list-style-type: none"> Experiences from CS.2. Literature: [Ballou et al. 1998], [Redman 1998], [Joshi/Rai 2000], [Fisher/Kingma 2001], [Gorla/Krishnan 2002], [Fisher et al. 2003], [Heinrich et al. 2007], [Heinrich et al. 2009] and [Heinrich et al. 2009].
DR27	Techniques	Definition of technique validation rule design.	<ul style="list-style-type: none"> Experiences from CS.2. Literature: [Parssian et al. 1999], [Shankaranarayanan et al. 2003], [Parssian et al. 2004], [Pierce 2004],

ID	Artifact	Design step	Argument
			[Shankaranarayanan/Cai 2006], [Heinrich et al. 2007], [Hipp et al. 2007], [Fan et al. 2008] and [Parssian et al. 2009].
Design decisions during preparation of CS.3			
DR28	Procedure model	Definition of activity define and prioritize requirements.	<ul style="list-style-type: none"> Experiences (i.e. requirement) from CS.3.
DR29	Documentation model	Definition of result document list of requirements.	<ul style="list-style-type: none"> Design objective DO1. Literature: [Heinrich et al. 2007].
DR30	Techniques	Extension of technique interview/workshop by a repository of generic requirements for data quality metrics.	<ul style="list-style-type: none"> Literature: [Even et al. 2007] and [Kaiser et al. 2007].
DR31	Procedure model	Definition of activity check metrics against requirements. Result document list of requirements can be used for the check activity as well.	<ul style="list-style-type: none"> Experiences (see DR28) from CS.3.
Design rationales upon finalization of CS.3			
DR32	Meta-model	Definition of entity strategic objective.	<ul style="list-style-type: none"> Experiences from CS.3. Literature: [Even/Shankaranarayanan 2005], [Lee et al. 2006], [Robbert and Fletcher 2007], [Batini et al. 2009], [Even/Kaiser 2009] and [Viscusi et al. 2009].
DR33	Techniques	Definition of technique data quality simulation and implementation of a simulation tool.	<ul style="list-style-type: none"> Experiences from CS.3. Literature: [Raghunathan 1999] and [Lee et al. 2006].

Table D-2: Design rationales

Technique	Description
Documentation	<p>As companies usually use templates that have been agreed with their IT department or external IT providers for business design documents, the method only contains a generic structure for a business design with notes on using the result documents produced during the course of applying the method. Using the technique data quality simulation makes it possible to illustrate the elements identified and specified (causal chains, validation rules, calculation formulas, etc.). Modeling various scenarios and causal chains is useful, particularly for consultation with the sponsor.</p> <p>Supportive tools:</p> <ul style="list-style-type: none"> Documentation templates for each activity of the procedure model. Generic structure for a business design document.
Data quality analysis	<p>This technique utilizes statistical methods to identify unusual data elements (i.e. which significantly deviate from the mean or are rarely used). Various software manufacturers offer applications to support this technique [Alur et al., 2007]. Providing that the corporate data steward considers its application useful, this technique should be used to prepare for a workshop in order to find points of reference for possible causal chains. In the case of identified deviations, however, it is important to check whether these really do involve</p>

Technique	Description
	<p>a data defect and, if so, whether the data defect is attributable to a known business problem (bottom-up approach).</p> <p>Supportive tools:</p> <ul style="list-style-type: none"> • Software for data analysis, e.g. WebSphere Information Analyzer provided by IBM [Alur et al. 2007].
Data quality simulation	<p>This technique enables data quality metrics values to be simulated on the basis of a modeled DQM scenario. A scenario of this kind encompasses various causal chains (i.e. data defects, DQ metrics, business problems) and DQM measures, and defines assumptions in respect of the probability of data defects, the cause of business problems and the impact of DQM measures [Otto et al. 2010].</p> <p>Figure D-1 shows a tool for the modeling and simulation of DQM scenarios. The tool enables the definition of data defects and a probability can be given for each data defect. A scenario also encompasses business problems and the assignment of defined data defects as the cause of these business problems. A simulation step simulates the occurrence of data defects and the number of business problems caused. Reactive DQM measures can repair data defects and thus reduce the number of business problems caused in the next simulation step.</p> <p>Supportive tools:</p> <ul style="list-style-type: none"> • Data quality management simulation tool (see Figure D-1).
Interview/workshop	<p>The aim of an interview is to collect information on or evaluate a specific set of facts in a conversation with one or more people. A workshop, on the other hand, serves less for collecting information and more for developing or discussing outcomes (e.g. defining requirements). It is advisable to conduct a workshop with several people as discussions between the participants open up new perspectives and facilitate the identification of otherwise hidden correlations (e.g. interdepartmental causalities between data defects and business problems).</p> <p>Supportive tools:</p> <ul style="list-style-type: none"> • Interview guideline for the identification of appropriate scopes. • Interview guideline for the identification of data and IT systems used in the focused process activities. • Interview guideline for the identification of business problems and causing data defects. • Repository of causal chains and data quality metrics. • Repository of generic requirements for data quality metrics.
Specification of validation rules	<p>This technique assists the corporate data steward with the specification of validation rules as measuring technique for DQ metrics and uses the result documents produced in Activity I.3. The goal is to specify one or more validation rules for each business-critical data defect identified in Activity I.3.</p> <p>Supportive tools:</p> <ul style="list-style-type: none"> • Patterns for validation rules. • Spreadsheet-based template for validation rule definition.
Specification of calculation formulas	<p>When data objects are verified by means of validation rules, a calculation formula links the results of individual validation rules to a metric value. As already mentioned above, it is advisable to use precisely one calculation formula per metric. In this case, the calculation formula of the DQ metric calculates the value of the metric on the basis of the results of validation rules which are assigned to that metric.</p>

Technique	Description
	Supportive tools: <ul style="list-style-type: none"> Patterns for calculation formulas.

Table D-3: Techniques of the method

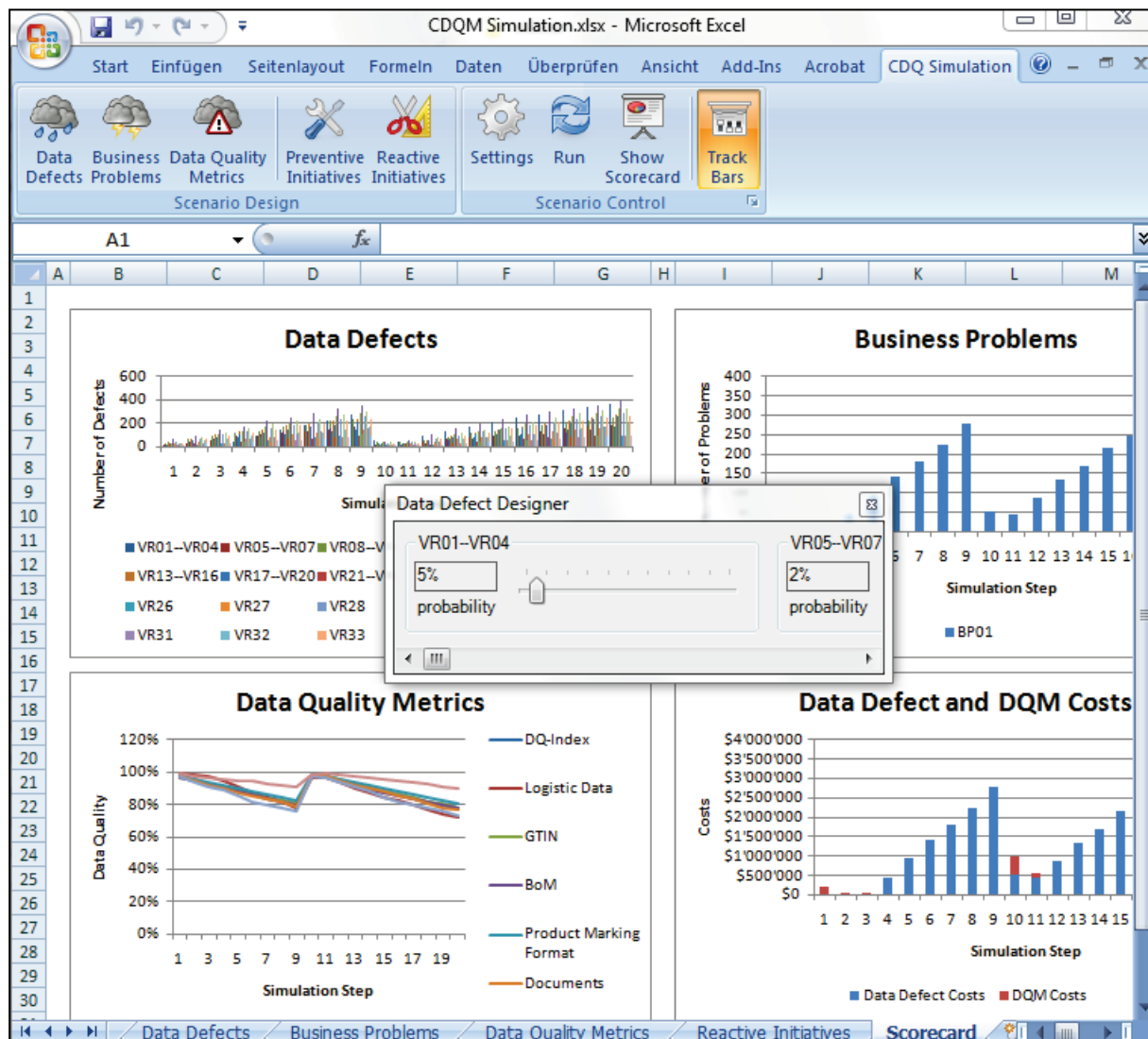


Figure D-1: Spreadsheet-based simulation tool

Company	Function
Bayer CropScience AG	Integration Manager Enterprise Master Data Management
Beiersdorf AG	Head of Supply Chain Data Process Management
Beiersdorf AG	Head of Team Beiersdorf Shared Services Master Data
DB Netz AG	Strategic Infrastructure Data Management, Railway Geographical Data
Festo AG & Co. KG	Head of Product Lifecycle Management
Grundfos A/S	Supply Chain Coordinator
Henkel AG & Co. KGaA	Project Manager Data Quality Strategy
Metro Group AG	Head of Business Intelligence Support
Rewe-Zentral AG	Project Manager Data Quality

Company	Function
Siemens Enterprise Communications GmbH & Co. KG	Vice President Master Data
Syngenta Crop Protection AG	Lead Steward Material Master Data

Table D-4: Subject matter experts

Relevance perspective
<ul style="list-style-type: none"> • Does the method's objective (i.e. identifying causalities between data defects and business problems, and specifying data quality metrics based on these causalities) address a relevant task from a company's point of view? • Does the method produce a relevant result? • Does the method's output (i.e. the metrics) serve the intended purpose?
Economic perspective
<ul style="list-style-type: none"> • While the method aims at reducing costs (e.g. by providing a structured procedure, examples, and interview guidelines), the method's configuration and application cause costs, too. • Is the effort caused by the method's configuration and application acceptable, or would any ad hoc approach serve the purpose as well? • Do the benefits of the method application exceed its costs?
Deployment perspective
<ul style="list-style-type: none"> • In order to foster communication between the stakeholders involved, the method should be understandable. It should correspond to concepts, the prospective method users are familiar with. • Are the concepts of the method (i.e. the meta-model, the procedure model) understandable? • Will the approach be accepted in a company? Is there a risk of a "not invented here" syndrome?
Engineering perspective
<ul style="list-style-type: none"> • Does the method fulfill the requirements to be taken into account? • Is the method's specification and structuring suited to supporting the intended purposes? • Does the method describe the considered aspects correctly and in a sufficient depth? • Is the method's structuring clear and understandable?
Epistemological perspective
<ul style="list-style-type: none"> • Is the method design process clear and understandable? Are there method elements (e.g. a role, an activity) without an obvious purpose? • Does the method design process generate credibility of the method, or does the method appear to be an ad hoc approach?

Table D-5: Guiding questions for evaluation

Success factors for effective data quality measurements		Method characteristics
SF1: Business impact	Well-defined causal chains between validation rules and business problems ensure monitoring of relevant data quality issues and thereby management support.	<ul style="list-style-type: none"> • Identification of causal chains between data defects and business problems is conducted prior to specification of data quality metrics (see Activity I.3).
SF2: Business user acceptance	Business knowledge is needed to identify data defects that are critical for business success. Therefore, business users should support the process of measuring data quality.	<ul style="list-style-type: none"> • Process owner and data user are defined roles of the method. • Interviews and workshops are preferred techniques for several for phase I activities.
SF3: Clear	To publish measurement results	Reporting and presentation of measur-

Success factors for effective data quality measurements		Method characteristics
presentation	geared to the target group (e.g. presenting results via an easy-to-use web interface) increases management attention and the amount of interested and participating business users.	ing results is not addressed by the method.
SF4: Comparability of measurements	Comparable measurements (e.g. for particular countries) increases motivation of local DQM teams.	<ul style="list-style-type: none"> Validation rules in a standardized form are proposed as a uniform measurement technique for all data quality metrics (see Activity II.2 and technique specification of validation rules). A uniform pattern for calculation formulas also supports comparability (see technique specification of calculation formulas).
SF5: Defined metric update procedure	To update the data quality measuring technique in order to provide effective measurements over time is important to sustain business user acceptance.	Maintenance processes for specified data quality metrics are not addressed by the method.
SF6: Management attention	Support by directors or senior management ensures business users aids and increases the visibility and acceptance of data quality measurements.	<ul style="list-style-type: none"> The involvement of a sponsor role is supposed to catch senior management attention. Causal chains between data defects and business problems provide a business perspective on data quality.
SF7: Plain metric system	A single, simple metric is more meaningful than a huge amount of abstract and interrelated metrics. Simplicity ensures comprehensible and accepted data quality measurements.	<ul style="list-style-type: none"> The techniques specification of validation rules and specification of calculation formulas enable a uniform measurement technique for data quality metrics which can easily be used to define a plain metric system.
SF8: Providing business support	To offer services for improving data quality in addition to measuring data quality can convert a negative stance on data quality measurements into acceptance and demand for DQM support.	Support processes for DQM are not addressed by the method.
SF9: Rule-based measuring technique	Validation rules that cover clearly defined data defects are an effective approach for measuring data quality.	<ul style="list-style-type: none"> Validation rules are proposed to specify the measurement technique of data quality metrics (see technique specification of validation rules).

Table D-6: Comparison of success factors and method characteristics

Anhang E Anhänge zu Publikation B-3

E.1 Entitätstypen des Metamodells der Methode

Tabelle E-1 beschreibt die Entitätstypen des Metamodells der Methode (vgl. Abschnitt 3.5.1) und ihre Beziehungen zu Entitätstypen des Metamodells des BE-Methodenkerns (vgl. Abschnitt 3.2.1). Das Metamodell des BE-Methodenkerns spezifiziert HÖNING [2009, S. 252-257].

Metaentitätstyp	Beschreibung
Datenobjektattribut (kurz: Attribut)	<p>Ein Datenobjektattribut (kurz: Attribut) ist der Typ eines Datenelements. Datenobjektattribute entsprechen Attributen in relationalen Datenmodellen. Atomaren Datenobjektattributen wird bei der Instanziierung ein atomarer Wert zugeordnet [Schmidt 2010, S. 222].</p> <p>Beziehungen Aus einem Datenobjektattribut können Datenelemente instanziiert werden. Ein Datenobjektattribut ist Bestandteil eines Datenobjekttyps.</p>
Datenobjekt	<p>Ein Datenobjekt ist eine Menge mehrerer Datenelementen, die zusammen die Ausprägung eines entsprechenden Datenobjekttyps bilden [Schmidt 2010, S. 222].</p> <p>Beziehungen Ein Datenobjekt setzt sich aus mehreren Datenelementen zusammen. Datenobjekte können hierarchisch strukturiert sein. Ein Datenobjekt repräsentiert ein Geschäftsobjekt auf Systemebene. Ein Datenobjekt ist Instanz eines Datenobjekttyps.</p>
Datenobjekttyp	<p>Ein Datenobjekttyp beschreibt ein Datenobjekt z. B. anhand seines Wertebereichs oder seiner Klassifizierung. Er vereinigt eine Menge zusammengehöriger Datenobjektattribute, mit deren Hilfe Datenobjekte (Instanzebene) abstrakt (z. B. in einem Modell) beschrieben werden können [Schmidt 2010, S. 222].</p> <p>Beziehungen Aus einem Datenobjekttyp können Datenobjekte instanziiert werden. Ein Datenobjekttyp besteht aus mehreren Datenobjektattributen. Ein Datenobjekttyp realisiert genau einen Geschäftsobjekttyp durch Auswahl bestimmter Objekteigenschaften und deren Abbildung auf Datenobjektattribute. Datenobjekttypen können hierarchisch strukturiert sein.</p>
Datenqualität	<p>Datenqualität ist ein durch einen Nutzer von Datenelementen wahrgenommenes Phänomen. Datenqualität bewertet die Erfüllung von an Datenelemente gestellte Anforderungen eines Datennutzers [Wang/Strong 1996, S. 6]. Ein Datennutzer kann bei der Ausführung verschiedener Aktivitäten unterschiedliche Anforderungen an ein (gleiches) Datenelement stellen (Kontextabhängigkeit).</p> <p>Beziehungen Datenqualität ist ein Bestandteil der Prozessleistung und wird von Unterstützungsprozessen (Datenmanagement, Datenqualitätsmanagement) erzeugt.</p>
DQ-Dimension	DQ-Dimensionen repräsentieren bestimmte Aspekte (z. B. Aktualität,

Metaentitätstyp	Beschreibung
	<p>Glaubwürdigkeit, Vollständigkeit) des Phänomens Datenqualität [Wang/Strong 1996, S. 6].</p> <p>Beziehungen Eine DQ-Dimension ist ein spezieller Erfolgsfaktor. DQ-Dimensionen konzeptualisieren das Phänomen Datenqualität. DQ-Dimensionen können hierarchisch strukturiert sein.</p>
DQ-Kennzahl	<p>Eine DQ-Kennzahl ist ein quantitatives Mass für Datenqualität. Ein Messsystem misst einen Messwert durch die Anwendung einer Messtechnik (z. B. Überprüfung von Datenelementen mit Validierungsregeln) an einem Messpunkt (z. B. ein Datenbehälter) zu einem durch eine Messfrequenz definierten Zeitpunkt, die Abbildung der Messung auf einer Messskala und die Zuweisung einer Masseinheit [IEEE 1998, DIN/EN/ISO 2000, Batini/Scannapieco 2006, S. 19].</p> <p>Beziehungen DQ-Kennzahlen operationalisieren DQ-Dimensionen. DQ-Kennzahlen können hierarchisch strukturiert sein.</p>
DQM-Massnahme	<p>Eine präventive DQM-Massnahme (z. B. Spezifikation und Nutzung von DQ-Kennzahlen) reduziert das Risiko von Datendefekten (Reduktion der Auftrittswahrscheinlichkeit und/oder der Wirkung eines Datendefekts). Eine reaktive DQM-Massnahme (z. B. Entfernen von Dubletten) behebt Datendefekte nach deren Auftreten [Otto et al. 2010].</p> <p>Beziehungen DQM-Massnahmen können hierarchisch strukturiert sein.</p>
Geschäftsobjektattribut	<p>Ein Geschäftsobjektattribut (z. B. Alter, Gewicht) kennzeichnet gemeinsame Eigenschaften von Geschäftsobjekttypen. Geschäftsobjektattribute werden durch fachliche Attributtypen und Multiplizitäten charakterisiert [Schmidt 2010, S. 223].</p> <p>Beziehungen Ein Geschäftsobjektattribut ist Bestandteil eines Geschäftsobjekttyps.</p>
Geschäftsobjekttyp	<p>Geschäftsobjekttypen (z. B. Kunden, Produkte) vereinigen eine Menge zusammengehöriger Geschäftsobjektattribute, mit deren Hilfe Geschäftsobjekte der Realwelt (Instanzebene) abstrakt (z. B. in einem Modell) beschrieben werden können [Schmidt 2010, S. 223].</p> <p>Beziehungen Ein Geschäftsobjekttyp ist eine Sammlung von Metadaten zu einem Geschäftsobjekt (Instanz). Aus einem Geschäftsobjekttyp können mehrere Geschäftsobjekte instanziiert werden. Ein Geschäftsobjekttyp besteht aus mehreren Geschäftsobjektattributen. Ein Geschäftsobjekttyp wird durch genau einen Datenobjekttypen realisiert.</p>
Metadaten	<p>Metadaten sind Datenelemente, die die Bedeutung sowie fachliche und technische Eigenschaften anderer Daten beschreiben [Weber 2009, S. 224]. Technische Metadaten (z. B. Datentypen, Felddlängen, Wertebereiche) spezifizieren Datenobjekte (und deren Attribute) eines IS, fachliche Metadaten (z. B. Verantwortlichkeit für Pflege, korrekte Erzeugung) beschreiben die Nutzung von Geschäftsobjekten (oder der sie repräsentierenden Datenobjekte) in Aktivitäten der Ablauforganisation eines</p>

Metaentitätstyp	Beschreibung
	<p>Unternehmens.</p> <p>Beziehungen Metadaten sind spezielle Datenelemente. Metadaten beschreiben Datenobjekttypen. Metadaten beschreiben Geschäftsobjekttypen.</p>
Stammdatenklasse	<p>Eine Stammdatenklasse gruppiert Stammdatenobjekte des gleichen Typs (z. B. Kundendaten, Produktdaten).</p> <p>Beziehungen Eine Stammdatenklasse ist ein spezieller Datenobjekttyp. Stammdatenklassen können hierarchisch strukturiert sein.</p>
Stammdatenobjekt	<p>Ein Stammdatenobjekt ist ein spezielles Datenobjekt, dessen Datenelemente sich im Vergleich zu anderen Datenobjekten (z. B. Bewegungsdatenobjekten) seltener ändern [Schmidt 2010, S. 225].</p> <p>Beziehungen Ein Stammdatenobjekt ist ein spezielles Datenobjekt.</p>

Tabelle E-1: Definition der Metaentitätstypen der Methode

E.2 Datenqualitätsdimensionen

E.2.1 Dimensionen nach Larry P. English

ENGLISH [1999, S. 87-118] unterscheidet die in Tabelle E-2 erläuterten DQ-Dimensionen. Dieser und die folgenden Anhänge verwenden englische Bezeichner der Originalquellen (jeweils angegeben in den Tabellenbeschriftungen), um die von den Autoren intendierte Differenzierung korrekt anzugeben. Die angegebenen Beschreibungen sind frei übersetzt und erläutern die verschiedenen Konzepte im Kontext der vorgestellten Methode.

Kategorie / Dimension	Beschreibung
Information Product Specification Quality	
Data Standards Quality	Bewertet Genauigkeit, Klarheit, Konsistenz, Verständlichkeit und Vollständigkeit von Metadaten, die zur Spezifikation von Daten genutzt werden.
Data Name Quality	Bewertet Genauigkeit, Klarheit, Konsistenz, Verständlichkeit und Vollständigkeit der Bezeichner (und Abkürzungen) von Geschäftsobjekten, Datenobjekttypen, Attributen und Datentypen.
Definition Quality	Bewertet Genauigkeit, Klarheit, Konsistenz, Verständlichkeit und Vollständigkeit von Metadaten, die die Bedeutung von Daten für fachliche und technische Nutzung erläutern.
Business Rule Quality	Bewertet Genauigkeit, Klarheit, Konsistenz, Verständlichkeit und Vollständigkeit von Geschäftsregeln, die Anforderungen von Geschäftsprozessen und somit auch an Daten beschreiben.
Information and Data Architecture Quality	Bewertet den Wiederverwendungsgrad von Datenmodellen, ihre Stabilität und Flexibilität, sowie Genauigkeit und Vollständigkeit der Repräsentation der Anforderungen des Unternehmens und ihrer Implementierung in IS.

Kategorie / Dimension	Beschreibung
Inherent Information Quality	
Definition Conformance	Bewertet die Konsistenz zwischen Datenelement und Datenspezifikation (Metadaten).
Completeness of Values	Bewertet die Vollständigkeit von Daten, d. h. ob Datenelemente vorhanden sind.
Validity, or Business Rule Conformance	Bewertet, ob Datenelemente innerhalb des spezifizierten Gültigkeitsbereichs (z. B. Toleranzbereiche, Wertelisten) liegen.
Accuracy of Surrogate Source	Bewertet die Genauigkeit eines Datenelements im Vergleich zu einem Datenelement des gleichen Attributs (zum selben Geschäftsobjekt) aus einem anderen, als „korrekt“ bewerteten Datenbehälter.
Accuracy to Reality	Bewertet die Genauigkeit eines Datenelements im Vergleich zu der durch das zugehörige Attribut repräsentierten Eigenschaft des Geschäftsobjekts.
Precision	Bewertet, wie detailliert oder exakt ein Datenelement angegeben ist (z. B. Umfang und Verständlichkeit bei Fliesstext-Beschreibungen, Anzahl nötiger Nachkommastellen).
Nonduplication	Bewertet die Anzahl von Dubletten, d. h. wie viele redundante Datenobjekte es zu einem Geschäftsobjekt gibt.
Equivalence of Redundant or Distributed Data	Bewertet die (semantische) Übereinstimmung verschiedener Datenelemente (z. B. in unterschiedlichen Datenbehältern) zu einer Eigenschaft (z. B. Geburtsdatum, Geschlecht) eines Geschäftsobjekts.
Concurrency of Redundant or Distributed Data	Bewertet die Zeitdauer zwischen der Änderung eines Datenelements und der Aktualisierung aller Datenelemente in anderen Datenbehältern, die die gleiche Eigenschaft des Geschäftsobjekts repräsentieren.
Pragmatic Information Quality	
Accessibility	Bewertet, ob (d. h. technisch möglich) und wie einfach (bewertet durch Datennutzer) auf ein Datenelement zugegriffen werden kann.
Timeliness	Bewertet, ob auf ein Datenelement innerhalb einer bestimmten Zeit zugegriffen werden kann.
Contextual Clarity	Bewertet, ob ein Datenelement in der Form, wie es einem Datennutzer präsentiert wird, verständlich ist.
Derivation Integrity	Bewertet, ob Datenelemente korrekt (d. h. entsprechend ihrer Bedeutung) zu einem neuen Datenelement aggregiert werden.
Usability	Bewertet, wie einfach (d. h. intuitiv verständlich, ohne weitere Bearbeitung) ein Datenelement für eine bestimmte Aufgabe genutzt werden kann.
Rightness of Fact Completeness	Bewertet, ob zur Bearbeitung einer Aufgabe alle benötigten Daten in der benötigten Qualität verfügbar sind.

Tabelle E-2: DQ-Dimensionen nach ENGLISH [1999]

E.2.2 Dimensionen nach David Loshin

Tabelle E-3 zeigt eine Liste mit 56 von LOSHIN [2001, S. 102-123] definierten DQ-Dimensionen, die durch fünf Kategorien strukturiert ist.

Kategorie / Dimension	Beschreibung
Data Quality of Data Models	
Clarity of Definition	Bewertet Genauigkeit und Verständlichkeit von Bezeichnern für Datenobjekttypen, Attribute und Relationen.
Comprehensiveness	Bewertet die Weitsicht (z. B. Berücksichtigung möglicher zukünftiger Applikationen, Berücksichtigung möglicher zukünftiger Nutzer) einer Datenspezifikation (Datenobjekttyp und Metadaten).
Flexibility	Bewertet, wie schnell und mit wie viel Änderungsaufwand ein Datenspezifikation an neue (nicht berücksichtigte, vgl. Comprehensiveness) Anforderungen angepasst werden kann.
Robustness	Bewertet, wie gut eine Datenspezifikation neue Anforderungen ohne Anpassungen der Spezifikation erfüllen kann.
Essentialness	Bewertet den Anteil nicht benötigter, redundanter und überladener Elemente einer Datenspezifikation (z. B. Attribute, die für verschiedene Zwecke genutzt werden können).
Attribute Granularity	Bewertet den Detaillierungsgrad einer Datenspezifikation (z. B. Anzahl der Attribute eines Datenobjekttyps).
Precision of Domains	Bewertet den Detaillierungsgrad einer Datenspezifikation (z. B. Anzahl der Datenelemente einer Auswahlliste).
Homogeneity	Bewertet die Nutzung von Datenelementen zur Klassifikation von Geschäftsobjekten (z. B. die Einteilung von Kunden nach Höhe des jährlichen Umsatzes). Solche impliziten Klassifikationsstrukturen können Geschäftsregeln repräsentieren.
Naturalness	Bewertet die Nähe zwischen Datenspezifikation und repräsentiertem Geschäftsobjekt.
Identifiability	Bewertet, ob Datenobjekte eindeutig einem Geschäftsobjekt zugeordnet werden können oder ob es Dubletten gibt.
Obtainability	Bewertet, ob benötigte (d. h. durch die Datenspezifikation vorgesehene) Datenelemente verfügbar sind (d. h. gesammelt und eingegeben werden können).
Relevance	Bewertet die Relevanz spezifizierter Attribute (d. h. ob ein Attribut benötigt wird oder voraussichtlich benötigt werden wird).
Simplicity	Bewertet die Verständlichkeit einer Datenspezifikation (d. h. klare Struktur mit geringer Hierarchie und wenigen Sonderfällen).
Semantic Consistency	Bewertet die Konsistenz von Bezeichnern und Beschreibungen einer Datenspezifikation (z. B. gleiche Bedeutung für gleich benannte Attribute, Wiederverwendung von Beschreibungen).
Structural Consistency	Bewertet die Konsistenz der technischen Datenspezifikation (d. h. gleiche Modellierung gleicher Strukturen, z. B. gleiche Datentypen für Abmessungen, gleicher Aufbau von Bezeichnern).
Data Quality of Data Values	
Accuracy	Bewertet, wie exakt ein Datenelement mit einem anderen Datenelement übereinstimmt, das als korrekt eingestuft ist.
Null Values	Bewertet die Verwendung von NULL-Werten, (z. B. ob bestimmte Einträge wie 13.12.2099 als NULL-Synonym verwendet werden).
Completeness	Bewertet die Vollständigkeit von Daten, d. h. ob Datenelemente vorhanden sind oder nicht.
Consistency	Bewertet die Konsistenz von Datenelementen d. h. ob das gleiche Datenelement für die gleiche Eigenschaft eines Geschäftsobjekts (u. U. in einem anderen Datenobjekt oder einem anderen Datenbehälter) verwendet

Kategorie / Dimension	Beschreibung
	wird.
Currency/Timeliness	Bewertet die Aktualität eines Datenelements, d. h. die Übereinstimmung des Datenelements mit der aktuellen repräsentierten Eigenschaft eines Geschäftsobjekts.
Data Quality of Data Domains	
Enterprise Agreement of Usage	Bewertet die Akzeptanz einer unternehmensweit gültigen Datendomäne (d. h. einer Menge eindeutig spezifizierter Datenelemente, z. B. Adressdaten, Produktklassifikation).
Stewardship	Bewertet die Organisation von Verantwortlichkeiten für eine Datendomäne.
Ubiquity	Bewertet die Nutzung (z. B. Anzahl an Abteilungen) einer Datendomäne.
Data Quality of Data Presentation	
Appropriateness	Bewertet die Bereitstellung (z. B. Format, Präsentationsform) eines Datenelements.
Correct Interpretation	Bewertet die Verständlichkeit eines Datenelements.
Flexibility	Bewertet, wie einfach und schnell die Präsentation eines Datenelements an neue Anforderungen (z. B. neue Anforderungen eines neuen Nutzers, veränderte Struktur der Daten durch Änderungen der Datenspezifikation) angepasst werden kann.
Format Precision	Bewertet die Präzision der Darstellung eines Datenelements (z. B. Anzeige von Nachkommastellen).
Portability	Bewertet, wie einfach und schnell die Präsentation eines Datenelements an eine andere Umgebung (z. B. neue Präsentationsapplikation, neue Präsentationssprache) angepasst werden kann.
Representation Consistency	Bewertet die Konsistenz der Präsentation eines Datenelements im Vergleich zu anderen Informationen und Unternehmensstandards (z. B. Datumsformate, Sprache, Präsentationsfarben).
Representation of Null Values	Bewertet die Darstellung von NULL-Werten (z. B. die Verschattung fehlender oder die Hervorhebung bewusst nicht angegebener Datenelemente).
Use of Storage	Bewertet die Beanspruchung von Datenspeicher-Ressourcen (z. B. die Nutzung oder bewusste Auflösung von Normalformen bei der Spezifikation von Speicherstrukturen).
Data Quality of Information Policy	
Accessibility	Bewertet, ob (d. h. technisch möglich) und wie einfach (bewertet durch Datennutzer) auf ein Datenelement zugegriffen werden kann und wie einfach es ist, ein verfügbares Datenelement im bereitgestellten Format zu nutzen. Die Dimension betrachtet nur Datenelemente, für die keine nutzerspezifischen Zugriffsberechtigungen spezifiziert sind (vgl. Privacy, Security).
Metadata	Bewertet die Zufriedenheit mit der Bereitstellung von Metadaten.
Privacy	Bewertet, wie effektiv Datenelemente vor dem Zugriff unberechtigter Nutzer geschützt werden können (z. B. Berechtigungskonzepte, Flexibilität von Nutzerschnittstellen).
Redundancy	Bewertet die Zufriedenheit mit redundanten Daten, die z. B. Fehlertoleranz aber auch Synchronisationsprobleme erzeugen.
Security	Bewertet, wie effektiv Datenelemente vor ungewünschten Modifikationen (z. B. mutwillige Zerstörung, Spionage) geschützt werden können (z. B. Verschlüsselung).

Kategorie / Dimension	Beschreibung
Unit of Costs	Bestimmt die Kosten für Datenverwaltung und -pflege (z. B. Bereitstellung von Daten, Verbesserung von Datenqualität).

Tabelle E-3: DQ-Dimensionen nach LOSHIN [2001]

E.2.3 Qualitätsdimensionen korrekter Daten nach Jack E. Olson

OLSON [2003, S. 29-32] differenziert die DQ-Dimension *Richtigkeit* (org. *Accuracy*) durch die sechs in Tabelle E-4 erläuterten Dimension.

Dimension	Erläuterung
Value Representation Consistency	Bewertet, ob Datenelemente mit gleicher Bedeutung die gleiche Form haben (z. B. St. Gallen und Saint Gallen oder USA und United States of America).
Changed-Induced Inconsistencies	Bewertet Inkonsistenzen zwischen Datenelementen, deren Ursache nicht eine falsche Eingabe, sondern eine Änderung der Datenspezifikation ohne anschließende Korrektur von bereits vorhandenen Datenelementen ist (z. B. durch die Differenzierung möglicher Datenelemente – erst nur rot / blau / gelb erlaubt, später 256 verschiedene Farbwerte).
Valid Values	Bewertet, ob ein Datenelement das Potential hat korrekt zu sein, d. h. ob es innerhalb des erlaubten Wertebereichs liegt und eine konsistente, unmissverständliche Form hat.
Missing Values	Bewertet, ob ein nicht vorhandenes Datenelement ein Defekt ist oder eine bewusst nicht angegebene (oder nicht vorhandene) Eigenschaft repräsentiert.
Object-Level Accuracy	Bewertet, ob ganze Datenobjekte zu Geschäftsobjekten fehlen (oder vorhanden sind, obwohl sie nicht vorhanden sein sollten).
Object-Level Inconsistencies	Bewertet die Konsistenz nicht von Datenelementen, sondern von Datenobjekten (z. B. die Anzahl von Kundendatenobjekten in verschiedenen Systemen).

Tabelle E-4: Differenzierung der DQ-Dimension *Richtigkeit* nach OLSON [2003]

E.2.4 Dimensionen nach Thomas C. Redman

REDMAN [1996, S. 245-266] differenziert die in Tabelle E-5 beschriebenen DQ-Dimensionen und Kategorien.

Kategorie / Dimension	Erläuterung
Quality Dimensions of Conceptual View	
Content	Bewertet die Relevanz von Attributen, die Verfügbarkeit von Datenelementen und die Verständlichkeit der fachlichen Datenspezifikation.
Scope	Bewertet, ob der durch die Daten repräsentierte Ausschnitt der Realwelt (z. B. die repräsentierten Eigenschaften eines Geschäftsobjekts) zur Bearbeitung bestimmter Aufgaben (subjektive Bewertung) ausreicht.
Level of Detail	Bewertet, ob Datenelemente ein Geschäftsobjekt ausreichend detailliert repräsentieren.
Composition	Bewertet, ob Datenelemente ein Geschäftsobjekt adäquat (d. h. möglichst exakte Abbildung der Realwelt) repräsentieren.
View Consistency	Bewertet die Konsistenz (Inhalt und Struktur) verschiedener Datenelemente als Repräsentation eines Geschäftsobjekts.
Reaction to Change	Bewertet, wie flexibel und Robust Datenobjekttypen auf Veränderungen (z. B. neue Anwendungen, veränderte Schnittstellen) reagieren,

Kategorie / Dimension	Erläuterung
	d. h. ob Anpassungen möglich sind oder die veränderten Anforderungen keine Anpassungen erfordern.
Quality Dimensions of Data Values	
Accuracy	Bewertet die Richtigkeit von Datenelementen im Sinne einer möglichst exakten Repräsentation von Eigenschaften eines Geschäftsobjekts.
Completeness	Bewertet die Vollständigkeit von Datenelementen, d. h. den Anteil vorhandener Datenelemente für spezifizierte Attribute.
Currency	Bewertet die Aktualität von Datenelementen im Sinne einer möglichst exakten Repräsentation veränderter Eigenschaften eines Geschäftsobjekts.
Value Consistency	Bewertet die Gleichheit von Datenelementen, die die gleiche Eigenschaft eines Geschäftsobjekts repräsentieren.
Quality Dimensions of Data Representation	
Appropriateness	Bewertet die Art der Datenrepräsentation für eine bestimmte Aufgabe.
Interpretability	Bewertet, ob Datenelemente korrekt interpretiert werden können.
Portability	Bewertet, wie flexibel eine Datenrepräsentation in verschiedenen Kontexten (z. B. in unterschiedliche Abteilungen oder Anwendungen) genutzt werden kann.
Format Precision	Bewertet, ob ein Datennutzer Datenelemente in ihrer Darstellung ausreichend unterscheiden kann (z. B. ob genügend Nachkommastellen angezeigt werden).
Format Flexibility	Bewertet die Flexibilität der Datenrepräsentation (z. B. die Möglichkeit, Daten als Diagramm oder Zahlen darstellen zu können).
Ability to Represent Null Values	Bewertet die Darstellung von NULL-Werten (z. B. ob diese ausreichend hervorgehoben werden).
Efficient Usage of Recording Media	Bewertet die sparsame Nutzung von Datenspeicher-Ressourcen (z. B. die Nutzung von Text statt grafischer Symbole).
Representation Consistency	Bewertet, ob gleiche Datenelemente auch gleich dargestellt werden.

Tabelle E-5: DQ-Dimensionen nach REDMAN [1996]

E.2.5 Dimensionen nach Wang/Strong

WANG/STRONG [1996] haben die 15 in Tabelle E-6 beschriebenen DQ-Dimensionen und eine Gliederung durch vier Kategorien in einer zweistufigen Befragung (Identifikation möglicher Dimensionen, Bewertung der gesammelten Dimensionen) von Datennutzern identifiziert.

Kategorie / Dimension	Beschreibung
Intrinsic Data Quality	
Believability	Bewertet die Vertrauenswürdigkeit von Datenelementen.
Accuracy	Bewertet die Übereinstimmung von Datenelementen und der (z. B. durch einen Standard definierten) Attributspezifikation.
Objectivity	Bewertet, wie unverfälscht Datenelemente und wie unvoreingenommen oder unparteiisch Datenbehälter sind.
Reputation	Bewertet die Vertrauenswürdigkeit eines Datenbehälters.
Contextual Data Quality	

Kategorie / Dimension	Beschreibung
Value-Added	Bewertet den Nutzen von Daten für die Bearbeitung einer bestimmten Aufgabe.
Appropriate Amount of Data	Bewertet, ob der Umfang bereitgestellter Daten für die Bearbeitung einer bestimmten Aufgabe ausreicht.
Completeness	Bewertet die Vollständigkeit von Datenelementen (d. h. ob Datenelemente fehlen) für die Bearbeitung einer bestimmten Aufgabe.
Relevancy	Bewertet, ob Datenelemente für die Bearbeitung einer bestimmten Aufgabe verwendet werden können.
Timeliness	Bewertet die Aktualität von Datenelemente für die Bearbeitung einer bestimmten Aufgabe.
Representational Data Quality	
Interpretability	Bewertet, ob Sprache, Zeichen und Symbole von Datenelementen verständlich sind.
Ease of Understanding	Bewertet die Verständlichkeit (d. h. Klarheit der Bedeutung) von Datenelemente.
Representational Consistency	Bewertet die Konsistenz der Darstellung von Datenelementen.
Concise Representation	Bewertet die Präzision der Darstellung von Datenelementen für die Bearbeitung einer bestimmten Aufgabe.
Accessibility Data Quality	
Access Security	Bewertet die Zugriffsverwaltung von Datenelementen, d. h. wie gut Datenelemente vor unberechtigtem Zugriff geschützt sind.
Accessibility	Bewertet die Verfügbarkeit von Datenelementen, d. h. ob und wie einfach Datenelemente zur Verfügung stehen.

Tabelle E-6: DQ-Dimensionen nach WANG/STRONG [1996]

E.3 Kostenarten defekter Daten

E.3.1 Kostenarten nach Larry P. English

Für die Messung von Auswirkungen schlechter Datenqualität (org. *Nonquality Information Costs*) definiert ENGLISH [1999, S. 209-212] die in Tabelle E-7 erläuterten Kostenarten defekter Daten.

Kostenart	Beschreibung
Process Failure Costs	
Irrecoverable costs	Kosten sinnloser Aktivitäten (z. B. mehrfacher Rechnungs- oder Katalogversand an gleiche Person, Versand von Werbung an falsche Zielgruppe).
Liability and exposure costs	Kosten gesetzlicher Verstöße (z. B. Strafen, unnötige Zahlungen wie zu hohe Abfindungen oder Schadensersatzzahlungen aufgrund falscher Daten oder Unwissenheit).
Recovery costs of unhappy customers	Kosten zur Wiederherstellung von Kundenzufriedenheit (z. B. Kulanzzahlungen).
Information Scrap and Rework	
Redundant data handling and support costs	Kosten für Datenaufbereitung (z. B. Bereinigung, Formatierung), falls verfügbare Datenelemente in der bereitgestellten Form nicht brauchbar sind.

Kostenart	Beschreibung
Costs of hunting or chasing missing information	Kosten für Datenbeschaffung (z. B. Zeit für Recherche, Kommunikationsaufwand), wenn benötigte Datenelemente nicht verfügbar sind.
Business rework costs	Kosten für Nachbesserungen oder Neuproduktion (z. B. bei fehlenden Gefahrgutkennzeichen oder falschen Massen).
Workaround costs and decreased productivity	Mehrkosten alternativer Aktivitäten, falls der vorgesehene Arbeitsablauf nicht möglich ist (z. B. bei fehlenden Berechtigungen).
Data verification costs	Kosten für Datenüberprüfung, falls verfügbaren Datenelementen aufgrund bekannter möglicher Defekte nicht vertraut wird.
Software rewrite costs	Kosten zur Neuentwicklung oder Korrektur von Software (insbesondere Schnittstellen), falls diese basierend auf unvollständigen, ungenauen oder veralteten Datenobjekttypen entwickelt wurde.
Data cleansing and correction costs	Kosten zur Datenbereinigung, falls fehlerhafte Datenerfassung oder -eingabe die Ursache der Defekte ist (Bereinigungsmassnahmen wie z. B. die Aktualisierung veralteter Adressdaten sind nicht vermeidbarer Mehraufwand, sondern präventive DQM-Massnahmen).
Data cleansing software costs	Kosten für Datenbereinigungs-Software, falls diese zur reaktiven Datenbereinigung genutzt wird (s. Data cleansing and correction costs).
Lost and Missed Opportunity Costs	
Lost opportunity costs	Wert (im Sinne von Umsatzpotential) eines Kunden, falls dieser (z. B. aufgrund falscher Anrede oder falscher Schreibweise des Namens) einen Vertrag kündigt und zu einem Mitbewerber wechselt.
Missed opportunity costs	Wert (im Sinne von Umsatz- oder Einsparpotential) von Massnahmen, falls diese aufgrund defekter Daten scheitern oder nicht durchgeführt werden können.
Lost shareholder value	Verringerter Marktwert eines Unternehmens, z. B. aufgrund defekter Bilanzdaten oder geringeren Gewinns (s. andere Kostenarten).

Tabelle E-7: Kostenarten defekter Daten nach ENGLISH [1999]

E.3.2 Kostenarten nach David Loshin

LOSHIN [2001, S. 87-93] beschreibt verschiedene Kostenarten defekter Daten, (s. Tabelle E-8), um die Identifikation von Geschäftsproblemen und damit verbundener Kosten als Effekt von Datendefekten [S. 93-99] zu unterstützen.

Kategorie / Auswirkung	Beschreibung
Operational Impacts	
Detection	Kosten zur Identifikation von System- und Prozessfehlern (z. B. unerwartete Zustände von Anwendungen, erhöhte Durchlaufzeiten von Prozessaktivitäten), insbesondere Zeitaufwand, aber auch Kosten für Beschaffung und Pflege unterstützender Software.
Correction	Kosten (insbesondere Zeitaufwand) zur Korrektur von System- und Prozessfehlern (z. B. Neustart von Prozessen, Anpassung von Datenmodellen oder Anwendungsschnittstellen).
Rollback	Kosten zur Wiederherstellung von alten (d. h. vor dem Auftreten eines Fehlers) Systemzuständen.
Rework	Kosten für Prozessaktivitäten, die aufgrund eines Fehlers erneut ausgeführt wer-

Kategorie / Auswirkung	Beschreibung
	den müssen (z. B. Dateneingabe, Produktion, Rechnungsversand).
Prevention	Kosten zur Gestaltung, Implementierung und Integration präventiver DQM-Prozesse und -Werkzeuge (z. B. Spezifikation und Nutzung von DQ-Kennzahlen).
Warranty	Kosten zur Korrektur und Kompensation (z. B. Kulanzleistungen) von Beeinträchtigungen von Kunden (z. B. Rechnungsversand an falsche Adresse und damit verbundene Datenschutzverletzung).
Spin	Kosten für Leistungen, die aufgrund der Auswirkung defekter Daten nötig werden (z. B. erhöhte Auslastung des Kundendienst, vermehrte Nutzung kostenfreier Service-Telefonnummern).
Reduction	Geringerer Umsatz (z. B. Kündigung von Verträgen).
Attrition	Geringerer Umsatz, da bestimmte Kunden gar keine Leistung mehr beziehen (und ggf. sogar zu Mitbewerbern wechseln).
Blockading	Geringerer Umsatz aufgrund defekter Daten, die zu Unzufriedenheit von Kunden führen und dazu, dass diese unzufriedenen Kunden andere Kunden beeinflussen (z. B. durch Leserbriefe).
Tactical and Strategic Impacts	
Delay	Kosten verspäteter Entscheidungen aufgrund nicht oder verspätet verfügbarer Daten, die z. B. zu verspäteter Produktion, verspäteter Produkteinführung oder der Produktion nicht benötigter Produkte führen.
Preemption	Kosten verspäteter Entscheidungen aufgrund unglaubwürdiger oder inkonsistenter Daten, die zur Aufschiebung von Entscheidungen führen.
Idling	Kosten verspäteter Entscheidungen, die zu Verzögerungen in Projekten führen.
Increased Difficulty	Kosten erhöhten Aufwands (z. B. bei der Umgestaltung von Prozessen oder der Integration neuer Anwendungssysteme) aufgrund fehlender oder falscher Datenelemente.
Lost Opportunities	Kosten strategische Effekte unzufriedener Kunden, die weniger Produkte kaufen oder zu Mitbewerbern wechseln (z. B. Verringerung des Marktanteils).
Organizational Mistrust	Kosten redundanter und „privater“ Datenhaltung (d. h. Umgehung der vorgesehenen Datennutzungsprozesse), die zu inkonsistenten Datenelementen und falschen Entscheidungen führen kann.
Misalignment	Kosten für Abstimmungsaufwand (z. B. Identifikation korrekter Prozesskennzahlen-Messwerte) aufgrund redundanter und „privater“ Datenhaltung (s. o.) einzelner Personen oder Abteilungen.
Acquisition Overhead	Kosten für Analysen, Korrekturen und Formatanpassungen eingehender Daten (z. B. bei der Übermittlung von Produktdaten von Hersteller zu Einzelhändler).
Decay	Kosten zur Pflege nicht benötigter Daten (z. B. inaktiver Produkte).
Infrastructure	Kosten zur Gestaltung, Implementierung und Pflege eines unternehmensweit genutzten Datenbehälters.

Tabelle E-8: Kostenarten defekter Daten nach LOSHIN [2001]

E.3.3 Kostenarten nach Jack E. Olson

OLSON [2003, S. 106-107] differenziert die drei in Tabelle E-9 beschriebenen Kostenarten defekter Daten.

Kostenart	Beschreibung
Typical Costs	Kosten für Ereignisse, die in der Vergangenheit aufgrund von Datendefekten entstanden sind (z. B. Kosten für Nachbesserungen, verlorene Kunden, verspätetes Berichtswesen, falsche Entscheidungen aufgrund unzureichender Informationen) und deren Wiederauftreten wahrscheinlich ist, falls die ursächlichen Datendefekte nicht behoben werden.
Wasted Project Costs	Kosten für Projekte (z. B. Einführung neuer Software für Customer Relationship Management (CRM), Decision Support Systems (DSS) oder Enterprise Resource Planning (ERP)), die aufgrund defekter Daten scheitern oder sich deutlich verzögern.
Costs of Slow Response to New Needs	Wert (im Sinne von Umsatz- oder Einsparpotential) von Massnahmen (z. B. Einkaufszentralisierung, Konsolidierung der Applikationslandschaft), die aufgrund von Datendefekten (z. B. unvollständige und ungenaue Lieferantendaten, unvollständige Dokumentation der Applikationslandschaft) nicht umgesetzt werden können.

Tabelle E-9: Kostenarten defekter Daten nach OLSON [2003]

E.4 Identifikation von Kausalketten

E.4.1 Exemplarische Kausalketten

Tabelle E-10 zeigt in Fallstudien und Fokusgruppeninterviews identifizierte Kausalketten. Der Kontext einer Kausalkette beschreibt dabei neben dem strategischen Ziel jeweils auch kurz das Unternehmen und den Betrachtungsbereich.

Kausalkette	Beschreibung
Befestigungstechnik / Kundendaten	
Kontext	Das Unternehmen verkauft seine Produkte überwiegend im Direktvertrieb unmittelbar bei Kunden (z. B. Bauunternehmer, Handwerker). Dies ermöglicht einen guten Einblick in aktuelle Kundenbedarfe und somit einen kundenorientierten Innovationsprozess. Es erfordert aber auch eine flexible und weltweit gut vernetzte Vertriebsorganisation, da im Vergleich zu alternativen Vertriebskanälen weniger Kunden betreut werden können. Strategische Ziele des Unternehmens sind daher hohe Kundenzufriedenheit und hohes Vertrauen in die angebotenen Dienstleistungen und Produkte (erreicht durch möglichst konstante Zuordnung von Kunden und Vertriebsmitarbeitern) und eine effiziente Vertriebsorganisation (d. h. gute Betreuung möglichst vieler Kunden durch einen Vertriebsmitarbeiter).
Geschäftsproblem	Neben allgemeinen Geschäftsproblemen, wie Fehlern bei der Rechnungserstellung, falsch zugestellten Rechnungen oder falsch geplanten zentralen Marketinginitiativen, verursachen Defekte der Kundendaten im Direktvertrieb spezielle Probleme. So können Defekte der Kundenklassifizierung dazu führen, dass einem Kunden für ihn eigentlich interessante Produkte gar nicht angeboten werden, da der (u. U. neue) Vertriebsmitarbeiter den Kundenbedarf nicht vollständig kennt. Ausserdem wirken Defekte der Adress- und Kommunikationsdaten unmittelbar auf die (wichtige, s. o.) Effektivität eines Vertriebsmitarbeiters aus, da dieser seine Vertriebsroute nicht optimal planen kann und korrekte Daten erst recherchieren muss.
Datendefekt	Fehlende, ungenaue oder veraltete Kundendaten und Kundenklassifizierung (d. h. Zuordnung von Kunden zu potentiell für ihn interessanten Produktgruppen).
Chemische Industrie / Materialdaten	

Kausalkette	Beschreibung
Kontext	Das Unternehmen hat Systeme aus Gesellschaften in mehr als 120 Ländern in drei regionalen ERP-Systemen konsolidiert. Ein laufendes Projekt hat das Ziel, die drei Regionalsysteme in einem global genutzten ERP-System zu konsolidieren. Ein zentraler Stammdaten-Server verteilt Kunden-, Lieferanten- und Materialdaten in die Regionalsysteme und zukünftig in das zentrale ERP-System. Ein strategisches Ziel des Unternehmens ist es, Materialbedarfe konzernweit zu konsolidieren, um in einem zentralen Einkauf bessere Einkaufskonditionen aushandeln zu können.
Geschäftsproblem	In der Produktionsplanung verhindern Defekte in der Produkthierarchie eine korrekte Konsolidierung von Bedarfen und dadurch eine optimale Bedarfsplanung. Im Berichtswesen ist die korrekte Zuordnung des Absatzes zu den Produkten und den zuständigen Geschäftsbereichen nicht möglich.
Datendefekt	Falsche oder unvollständige Klassifizierung von Produkten mit der Produkthierarchie (Gruppierung von Produkten in einer Baumstruktur), Inkonsistente Klassifizierung verpackter und unverpackter Materialien.
Energieversorgung / Kundendaten	
Kontext	Strom AG, vgl. Abschnitt 3.5.2. Ein strategisches Ziel des Unternehmens ist hohe Kundenzufriedenheit, um Kunden nicht an Mitbewerber zu verlieren (verstärkte Konkurrenz durch Liberalisierung des Energiemarkts).
Geschäftsproblem	Fehlende Anreden führen zu unvollständigen Anschreiben, z. B. in Rechnungen oder Marketing-Informationen. Ausserdem werden vollständige Anreden benötigt, um personalisierte Anreden (z. B. für Begrüssungsschreiben, Marketing-Anforderung) zu realisieren.
Datendefekt	Fehlende Anreden.
Geschäftsproblem	Abrechnungssperren blockieren u. a. den Rechnungsversand, wenn beispielsweise ein Empfänger für längere Zeit im Ausland ist. Abrechnungssperren sollten aber ohne triftigen Grund nicht zu lange gesetzt sein, da sie den Abrechnungsprozess verzögern und Mehraufwand verursachen.
Datendefekt	Veraltete Abrechnungssperren.
Geschäftsproblem	Rechnungen und Mahnungen werden an falsche Personen verschickt. Falls Kunden unterschiedliche Adressaten wünschen, wird dies explizit in einer Prüftabelle vermerkt.
Datendefekt	Inkonsistente Adressdaten für Mahnungs- und Rechnungsempfänger.
Geschäftsproblem	Ein falsches Datenelement führt zur Verwendung falscher Rechnungsformulare und somit zu fehlerhaften Rechnungsausdrucken.
Datendefekt	Falsche Datenelemente für Auswahl des Rechnungsformulars (Gross- und Tarifkunden).
Geschäftsproblem	Kunden erhalten eine Rechnung, obwohl sie bereits (mehrfach) Bankeinzug als gewünschten Zahlweg angegeben haben.
Datendefekt	Veralteter Zahlweg. Meldet ein Kunde einen Umzug, d. h. ändert sich seine Adresse, bleibt zwar der Kunde als Datenobjekt (Geschäftspartner) erhalten, es wird aber ein neues Datenobjekt für Abrechnungsdaten (Vertragskonto) angelegt. Dabei wird in Einzelfällen der vom Kunden gewünschte Zahlweg (z. B. Bankeinzug, Rechnung) nicht aus dem alten Vertragskonto übernommen.
Konsumgüterindustrie / Produktdaten	
Kontext	Konsumgüter AG, vgl. Abschnitt 3.5.2. Strategische Ziele des Unternehmens sind die Einhaltung gesetzlicher Vorga-

Kausalkette	Beschreibung
	ben, Kostenreduktion und ein hohes Service Level.
Geschäftsproblem	Defekte logistischer Daten führen in Logistikprozessen zu Mehraufwand und Verzögerungen. So können Paletten, deren Gewicht über dem zulässigen Höchstgewicht eines Lagers liegt, nicht eingelagert werden. Dies erfordert dann ggf. ein Umpacken der Paletten, was weitere Verzögerungen und ggf. fehlerhafte Lieferungen verursacht.
Datendefekt	Fehlende oder ungenaue logistische Produktdaten (z. B. Abmessung, Gewicht, Stückzahlen).
Geschäftsproblem	Inkonsistenzen zwischen GTINs auf Produkten oder logistische Einheiten (z. B. Paletten) und den an einen Daten-Pool gesendeten GTINs führen zu Lieferungen, die bei Kunden nicht problemlos verarbeitet werden können. Werden zu Inkonsistenzen vor der Warenlieferung entdeckt, verursachen sie Mehraufwand und ggf. Kosten für die Produktion neuer Artworks oder neuer Produkte.
Datendefekt	Inkonsistente GTINs.
Geschäftsproblem	Die Freigabe einer Stückliste ist erst möglich, wenn alle erforderlichen Stücklistenpositionen angegeben sind. Die verspätete Freigabe einer Stückliste führt zu einem verspäteten Produktionsbeginn und zu nicht eingehaltenen Lieferterminen führen.
Datendefekt	Unvollständige Stücklistenpositionen.
Telekommunikationsindustrie / Netzinfrastrukturdaten	
Kontext	Telefon AG, vgl. Abschnitt 3.5.2. Ein strategisches Unternehmensziel ist die
Geschäftsproblem	Defekte der Netzinfrastrukturdaten führen zu falschen Antworten bei der Prüfung, ob bestimmte Produkte, die die Verfügbarkeit von Breitband-Internet benötigen, an einer bestimmten Adresse angeboten werden können. Falsch-positive Antworten führen zu Unzufriedenheit bei Kunden (bestellte Produkte sind letztendlich nicht oder in nicht ausreichender Qualität verfügbar), falsch-negative Antworten führen zu geringerem Umsatz (eigentlich verfügbare Produkte werden nicht angeboten).
Datendefekt	Inkonsistente Netzinfrastrukturdaten. Zwar sind eigentlich Genauigkeit und Vollständigkeit die Erfolgsfaktoren für eine fehlerfreie Verfügbarkeitsprüfung, aufgrund der Komplexität (Umfang, notwendiger Detaillierungsgrad) der benötigten Dokumentation ist die Erreichung dieses Ziels aber unrealistisch.
Uhrwerkherstellung / Produktdaten	
Kontext	Das Unternehmen produziert Uhrwerke mit einem hohen Anteil manueller Fertigung. Aufgrund hoher Nachfrage und begrenzter Produktionskapazitäten ist das Unternehmen auf zügige und möglichst fehlerfreie Produktion besonders stark angewiesen. Strategische Ziele des Unternehmens sind daher geringe Abweichungen des Produktionsbeginns (org. Start of Production, Kennzahl der Produktionsplanung), schnelle Durchlaufzeiten (Produktionsdauer eines Uhrwerks) und geringer Anteil fehlerhafter Uhrwerke (Ausschuss).
Geschäftsproblem	Fehlen bei Produktionsbeginn Konstruktionszeichnungen (exakte Spezifikation einzelner Bauteile eines Uhrwerks) oder sind diese offensichtlich veraltet, müssen diese bei Konstrukteuren erfragt und nachträglich manuell in die Stücklisten eingepflegt werden. Dies verzögert den Produktionsbeginn. Fehlende oder veraltete Konstruktionszeichnungen können auch dazu führen, dass in der Fertigung auf Zeichnungen ähnlicher Produkte zurückgegriffen wird, was zur Produktion von Ausschuss führen kann.

Kausalkette	Beschreibung
Datendefekt	Fehlende oder veraltete Verknüpfung von Konstruktionszeichnungen in Stücklisten.
Zulieferung Kraftfahrzeugindustrie / Materialdaten	
Kontext	Das Unternehmen fertigt gleiche Produktteile in verschiedenen Werken mit hoch spezialisierten Produktionsanlagen. Ersatzteile für diese Maschinen sind meist teuer und müssen bei Bedarf erst angefertigt werden, so dass u. U. Produktionsverzögerungen entstehen. Aufgrund ähnlicher Produktionsanlagen in verschiedenen Werken, könnten vorrätige Ersatzteile theoretisch zwischen Werken ausgetauscht werden. Ein strategisches Ziel des Unternehmens ist es, den Austausch von Ersatzteilen für Produktionsanlagen zwischen Werken zu ermöglichen, um die Dauer von Wartungsarbeiten zu verkürzen und Lagerkosten zu senken.
Geschäftsproblem	Der Austausch passender Ersatzteile ist nur möglich, wenn diese korrekt durch technische Daten beschrieben und klassifiziert sind. Datendefekte führen zur Identifikation falscher Ersatzteile oder dazu, dass ein eigentlich vorrätiges Ersatzteil nicht gefunden wird.
Datendefekt	Fehlende oder ungenaue technische Daten (z. B. Abmessungen, Gewicht, Klassifikation, Material) von Ersatzteilen für Produktionsanlagen.

Tabelle E-10: Exemplarische Kausalketten

E.4.2 Leitfragen zur Identifikation kritischer Datendefekte

Die Leitfragen aus Tabelle E-11 unterstützen Interviews zur Identifikation geschäftskritischer Datendefekte. Sie sind durch die 15 von WANG/STRONG [1996] definierten DQ-Dimensionen strukturiert.

DQ-Dimension	Möglicher Datendefekt	Leitfrage
Access security	Datenelemente können nicht geändert werden, für die eigentlich Änderungsberechtigungen gegeben sein sollten. Datenelemente können geändert werden, für die eigentlich keine Änderungsberechtigungen gegeben sein sollten.	Gibt es Daten, die Sie ändern möchten und eigentlich auch ändern können sollten, aber dies nicht dürfen? Ist es möglich, fehlende oder falsche Daten unmittelbar zu ergänzen oder zu korrigieren? Nehmen Sie Änderungen an Daten vor, die Sie eigentlich nicht vornehmen können sollten? Gibt es Daten, die Sie benötigen, auf die Sie aber keinen Zugriff haben?
Accessibility	Die Beschaffung der benötigten Datenelemente dauert zu lange, weil die Systemleistung unzureichend oder der Beschaffungsweg zu lang ist.	Sind die benötigten Daten über den vorgesehenen Weg (z. B. verwendete Anwendung, vorgegebene Anzeigenmasken) erreichbar? Können die benötigten Daten über andere Quellen (z. B. Internet, Telefon, für die Prozessaktivität nicht vorgesehene Anwendung) bezogen werden? Wie oft ist dies nötig?
Accuracy	Eingabeformulare enthalten viele Pflichtfelder, für die aus mangelnder Kenntnis der Auswirkungen „Dummy“-Werte (d. h. NULL-Werte) eingetragen werden. Die verfügbaren Datenelemente	Kommt es oft vor, dass Daten falsch sind? (z. B. Adressdaten, falsche Telefonnummern)? Woran erkennen Sie, dass die Daten falsch sind? Gibt es Eingabeformulare, die viele Felder enthalten, die Sie ausfüllen müssen aber nicht

DQ-Dimension	Möglicher Datendefekt	Leitfrage
	<p>sind nicht ausreichend, um zwei Datenobjekte zweifelsfrei voneinander zu unterscheiden (Dubletten).</p> <p>Verschiedene Systeme liefern zu einem bestimmten Sachverhalt verschiedene Datenelemente (Inkonsistenz).</p>	<p>ausfüllen können (z. B. aus Zeitmangel oder Unwissenheit)?</p> <p>Passiert es, dass Daten mehrfach vorkommen, obwohl sie nur einmal im System sein dürfen? (z. B. verschiedene Adressen eines Kunden?)</p> <p>Gibt es Dateninkonsistenzen zwischen verschiedenen Systemen?</p>
Appropriate Amount of Data	<p>Es werden so viele Datenelemente angeboten, dass ein Mitarbeiter die relevanten Datenelemente nur schwer finden kann.</p> <p>Es ist notwendig, Datenelemente über verschiedene Wege (z. B. Systeme, Bildschirmmasken, Dokumente) zu beziehen, um einen einzelnen Arbeitsschritt zu absolvieren.</p> <p>Es können nicht alle nötigen Datenelemente erfasst werden. Daher werden z. B. Freitextfelder zweckentfremdet.</p>	<p>Gibt es Eingabeformulare oder Berichte (Reports), die so umfangreiche Daten enthalten, dass Sie Schwierigkeiten haben, relevante Daten zu finden?</p> <p>Gibt es Arbeitsschritte zu denen Sie Daten aus mehreren Anwendungen einholen oder verknüpfen müssen?</p> <p>Fehlen Ihnen in bestimmten Arbeitsschritten Eingabefelder, um für die Bearbeitung notwendige Daten einzutragen? Wenn ja: wo genau bringen Sie Ihre Daten unter?</p> <p>Gibt es Eingabeformulare, auf denen Sie viele Informationen eintragen müssen, die für den Geschäftsfall nicht relevant erscheinen? Wie gehen Sie damit um?</p>
Believability	<p>Da in vielen Fällen Datenelemente falsch oder unzureichend sind, wird von vorneherein schlechte Datenqualität unterstellt. Daher werden Datenelemente erfragt, die eigentlich vorliegen.</p>	<p>Gehen Sie von der Richtigkeit verfügbarer Daten aus, oder nehmen Sie von vorneherein Fehler an?</p> <p>Werden bestimmte Daten mit anderen Daten noch einmal validiert (z. B. Rückfrage gegen Kunden oder Recherche in alternativen Datenbehältern)?</p> <p>Welche Daten erscheinen Ihnen unglaubwürdig? Wie erkennen Sie dies?</p>
Completeness	<p>Die verfügbaren Datenelemente sind nicht ausreichend, um einen bestimmten Arbeitsschritt zu absolvieren. Es ist nicht möglich, die fehlenden Daten über andere Quellen (z. B. Telefon, andere Software) zu beziehen.</p> <p>Es fehlen Datenelemente.</p>	<p>Fehlen in bestimmten Bildschirmmasken oder Formularen Datenelemente, die eigentlich vorhanden sein sollten?</p> <p>Gibt es Datenelemente, die unverständlich oder zu stark verkürzt dargestellt sind?</p>
Concise Representation	<p>Die Darstellung von Datenelementen ist für die Bearbeitung einer bestimmten Aufgabe unvorteilhaft und könnte präziser sein.</p>	<p>Gibt es Fälle, in denen Ihnen Datenelemente angeboten werden, für deren Verständnis sie erst „lange nachdenken“ oder zusätzliche Informationsquellen verwenden müssen?</p>
Ease of Understanding	<p>Datenelemente sind lesbar (d. h. die Bedeutung „für sich“ ist klar), es ist aber unklar, wie Sie zu verwenden sind oder warum Sie angeboten werden.</p>	<p>Gibt es Daten die Sie verstehen, deren Bedeutung bzw. Auswirkung Ihnen aber nicht klar ist?</p>
Interpretability	<p>Datenelemente sind aufgrund einer nicht bekannten Sprache oder einer nicht korrekt eingesetz-</p>	<p>Gibt es Daten, die in einer für Sie fremden Sprache angezeigt werden?</p> <p>Gibt es Daten, in denen Sonderzeichen (z.B.</p>

DQ-Dimension	Möglicher Datendefekt	Leitfrage
	ten Codierung nicht verständlich.	„\$%&?#“) angezeigt werden, wo dies nicht der Fall sein sollte? Gibt es Felder, in denen unverständliche Werte („Kauderwelsch“) angezeigt wird?
Objectivity	Datenelemente liefern die Meinung oder Einschätzung einzelner Personen und keine objektive Information.	Zweifeln Sie an der Richtigkeit von Daten, weil Sie eine subjektive Meinung oder Einschätzung vermuten, aber eigentlich objektive Informationen erwarten?
Relevancy	Für einen Geschäftsfall werden so viele Datenelemente angeboten, dass die Auswahl der für den Geschäftsfall benötigten Datenelemente mit erheblichem Aufwand verbunden ist.	Werden Sie mit Daten „überflutet“, d. h. Ihnen fällt die Auswahl der zur Bearbeitung einer Aufgabe benötigten Daten schwer?
Representational consistency	Kein einheitliches Datenformat, z. B. von Adressdaten in verschiedenen Anwendungen. Dies kann zu Missverständnissen oder technischen Problemen beim Datenaustausch führen. Unterschiedliche Systeme greifen auf verschiedene Datenbehälter zurück und stellen Datenelemente im gleichen Kontext unterschiedlich dar. Datenelemente wurden teilweise nicht in den dafür vorgesehenen Eingabefeldern erfasst, daher werden sie auch in der Darstellung an falscher Stelle präsentiert.	Gibt es Daten, die in verschiedenen Anwendungen auf unterschiedliche Art erfasst oder dargestellt werden? Gibt es Fälle, in denen zu einem Kunden in unterschiedlichen Systemen unterschiedliche Daten angezeigt werden? Gibt es Bildschirmmasken, in denen Informationen „verdreht“ dargestellt sind? (z. B. Nachname steht im Feld für Vorname)?
Reputation	Das Vertrauen in verfügbare Daten (d. h. in die datenbereitstellende Anwendung oder Abteilung) besteht nicht.	Zweifeln Sie an der Richtigkeit von Daten aufgrund der Herkunft der Daten? Prüfen Sie regelmässig bestimmte Daten gegen vergleichbare Informationen aus anderen Anwendungen oder Abteilungen?
Timeliness	Datenelemente sind veraltet, obwohl aktuellere Datenelemente verfügbar sind oder sogar bereits eingegeben wurden. Nachbearbeitung oder Speichern von Datenelementen dauert so lange, dass der Eindruck entsteht, dass die Eingabe noch nicht getätigt wurden. Daher werden Daten mehrfach eingegeben (u. U. Ursache von Dubletten).	Sind eingegebene Daten für Folgeaufgaben rechtzeitig verfügbar oder aktualisiert? Gibt es Fälle, in denen Sie Eingaben mehrfach machen, da nur zeitverzögert oder gar nicht sichtbar ist, dass diese schon getätigt wurden?
Value-Added	Datenelemente sind zwar korrekt, aber nicht hilfreich für die Bearbeitung der jeweiligen Aufgabe.	Gibt es Daten, die Sie verstehen, die Ihnen aber bei der Bearbeitung der jeweiligen Aufgabe nicht weiterhelfen?

Tabelle E-11: Leitfragen zur Identifikation kritischer Datendefekte

E.5 Anforderungen an Datenqualitätskennzahlen

E.5.1 Anforderungen der DAMA

Die DAMA [2009, S. 299] beschreibt die in Tabelle E-12 beschriebenen Anforderungen an DQ-Kennzahlen.

Anforderung	Beschreibung
Measurability	Eine DQ-Kennzahl sollte messbar sein, d. h. es sollte möglich sein, Messwerte für die Kennzahl zu messen und in Beziehung mit Messwerten von Prozesskennzahlen oder mit Geschäftsproblemen zu setzen.
Business Relevance	Eine DQ-Kennzahl sollte einen Geschäftsbezug haben, d. h. es sollte beschrieben sein, welche Auswirkungen bei der Veränderung der Messwerte einer Kennzahl zu erwarten sind.
Acceptability	Für eine DQ-Kennzahl sollten ein Bereich akzeptierter Messwerte und Massnahmen, die bei Messwerten ausserhalb dieses Bereichs zu ergreifen sind, definiert sein.
Accountability / Stewardship	Einer DQ-Kennzahl sollte eine für die Kennzahl verantwortliche Rolle zugeordnet sein, die z. B. bei Messwerten ausserhalb des akzeptierten Bereichs informiert wird.
Controllability	Durch eine DQ-Kennzahl überwachte Dateneigenschaften sollten kontrollierbar, d. h. durch DQM-Massnahmen beeinflussbar, sein.
Trackability	Messwerte von DQ-Kennzahlen sollten über einen längeren Zeitraum messbar sein, um z. B. den Effekt von DQM-Massnahmen nachzuweisen.

Tabelle E-12: Anforderungen an DQ-Kennzahlen der DAMA [2009]

E.5.2 Anforderungen nach Kaiser et al.

KAISER ET AL. [2007, S. 3-4] definieren aufbauend auf der Arbeit von EVEN/SHANKARANARAYANAN [2007] und PIPINO ET AL. [2002] die in Tabelle E-13 beschriebenen Anforderungen an DQ-Kennzahlen.

Anforderung	Beschreibung
Normalization	Um die Vergleichbarkeit verschiedener Messwerte zu gewährleisten, sollten Messwerte einer DQ-Kennzahl normierbar sein, z. B. mit der gleichen Berechnungsvorschrift aggregiert werden.
Interval scale	Um Effekte von DQM-Massnahmen und Auswirkungen von Datendefekten auf Prozessleistung (überwacht durch Prozesskennzahlen) aufzeigen und vergleichen zu können, sollten Messwerte einer DQ-Kennzahl auf einer Verhältnisskala (z. B. Prozentskala) abgebildet werden.
Interpretability	Für Messwerte einer DQ-Kennzahl sollte immer ein Bezug zu konkreten Geschäftsproblemen herstellbar sein, um DQ-Messungen durch die Auswirkung der überwachten Datendefekte interpretieren zu können.
Aggregation	DQ-Kennzahlen sollten auf verschiedenen Ebenen (d. h. Attribute, Datenelemente, Datenobjekte) anwendbar sein mit gleicher Bedeutung der jeweiligen Messwerte. Zudem sollten Messwerte einer bestimmten Ebene (z. B. Datenelement, Datenobjekt) aus Messwerten der darunter liegenden Ebene (z. B. Datenelemente) zu berechnen sein.
Adaptivity	DQ-Kennzahlen sollten für konkrete Anwendungsfälle adaptierbar sein, d. h. nicht zu spezifisch beschrieben sein.

Anforderung	Beschreibung
Feasibility	DQ-Kennzahlen sollten so spezifiziert sein, dass Messungen möglichst maschinell (z. B. durch die zu bestimmten Zeitpunkten automatisch durchgeführte Prüfung bestimmter Datenobjekte durch Validierungsregeln) ohne zu hohen Aufwand möglich sind.

Tabelle E-13: Anforderungen an DQ-Kennzahlen nach KAISER ET AL. [2007]

E.5.3 Anforderungen identifiziert in einem Fokusgruppeninterview

In Einem Fokusgruppeninterview (11. Juni 2008, 90 Minuten, 18 Teilnehmer aus 10 verschiedenen Unternehmen) sind die in Tabelle E-14 beschriebenen Anforderungen an DQ-Kennzahlen identifiziert worden.

Anforderung	Beschreibung
Aufwand / Nutzen	Der Aufwand, der zur Definition und Erhebung der Werte einer DQ-Kennzahl erforderlich ist, sollte in einem positivem (< 1) Verhältnis zum Nutzen (kontrolliertes Fehlerpotential) stehen.
Definition Messfrequenz	Für eine DQ-Kennzahl sollte definiert sein, zu welchen Zeitpunkten Werte für sie erhoben werden.
Definition Messpunkt	Für eine DQ-Kennzahl sollte definiert sein, an welcher Stelle (z. B. Datenbehälter, Prozess, Abteilung) sie erhoben wird.
Definition Messverfahren	Für eine DQ-Kennzahl sollte definiert sein, mit welchem Instrument (z. B. Fragebogen, Software) ein Wert für die Kennzahl ermittelt wird.
Definition Skala	Für eine DQ-Kennzahl sollte eine Skala (z. B. Prozent, Schulnoten, Zeit) definiert werden.
Einschränkung der zu prüfenden Daten	Für eine DQ-Kennzahl sollte definiert sein, welche Daten (z. B. Materialstamm, europäische Kunden) betrachtet werden.
Eskalationsprozess	Für eine DQ-Kennzahl sollten Massnahmen in Abhängigkeit bestimmter Schwellwerte definiert sein, d. h. ab welchem Wert welche Massnahme durchzuführen ist.
Gültigkeitsbereich	Für eine DQ-Kennzahl sollte ein Bereich definiert sein, in dem ihre Werte gültig sind.
SMART-Prinzipien	Eine DQ-Kennzahl sollte die SMART-Ziele (spezifisch, messbar, anspruchsvoll/attraktiv, realistisch, terminiert) erfüllen.
Störvariablen	Die Metadaten einer DQ-Kennzahl sollten Angaben zu möglichen Störvariablen enthalten, d. h. mögliche Ereignisse oder Einflüsse beschreiben, durch die Werte der DQ-Kennzahl verfälscht werden könnten.
Verantwortlichkeit	Für eine DQ-Kennzahl sollten klare Verantwortlichkeiten definiert sein, d. h. es sollte definiert sein, an wen welche Werte der DQ-Kennzahl berichtet werden, wer für Wartung/Pflege der Kennzahl (z. B. Aktualität/Sinn der Definition, Implementierung des Messverfahrens) zuständig ist.
Vergleichbarkeit	Eine DQ-Kennzahl sollte so definiert sein, dass ihre Werte mit denen andere Kennzahlen (DQ-Kennzahlen und Prozesskennzahlen) verglichen werden können.
Verständlichkeit	Für eine DQ-Kennzahl sollten Metadaten verfügbar sein, die ihren Verwendungszweck und die korrekte Interpretation ihrer Werte erläutern.

Anforderung	Beschreibung
Verwendung in SLAs	DQ-Kennzahlen sollten in Service Level Agreements verwendet werden können.
Visualisierung	Die Werte einer DQ-Kennzahl sollten visualisiert werden können (z. B. Zeitreihen, Diagramme).
Wiederholbarkeit	Werte für eine DQ-Kennzahl sollten nicht nur einmalig, sondern mehrfach erhoben werden können.
Zielwert	Für eine DQ-Kennzahl sollte ein Zielwert definiert sein.
Zuordnung DQ-Dimension	Eine DQ-Kennzahl sollte einer oder mehreren DQ-Dimensionen zugeordnet werden können.
Zuordnung Geschäftsproblem	Eine DQ-Kennzahl sollte einem konkreten (unternehmensspezifischen) Geschäftsproblem zugeordnet werden können.
Zuordnung Prozesskennzahlen	Eine DQ-Kennzahl soll einer oder mehreren Prozesskennzahlen zugeordnet werden können.
Zuordnung Unternehmensstrategie	Eine DQ-Kennzahl sollte einem oder mehreren strategischen Zielen des Unternehmens zugeordnet werden können.

Tabelle E-14: Anforderungen an DQ-Kennzahlen

E.6 Messverfahren

E.6.1 Bewertung von Datenqualität nach Huang et al.

HUANG ET AL. [1999, S. 87f] stellen einen Fragebogen mit 69 Fragen (s. Tabelle E-15) zur Messung subjektiver Datenqualität vor (Bewertung auf einer elfstufigen Likert-Skala).

Bewertung subjektiver DQ-Dimensionen
This information is easy to manipulate to meet our needs.
It is easy to interpret what this information means.
This information is consistently presented in the same format.
This information includes all necessary values.
This information is easily retrievable.
This information is formatted compactly.
This information is protected against unauthorized access.
This information is incomplete.
This information is not presented consistently.
This information has a poor reputation for quality.
This information is complete.
This information is presented concisely.
This information is easy to understand.
This information is believable.

Bewertung subjektiver DQ-Dimensionen
This information is easy to aggregate.
This information is of sufficient volume for our needs.
This information is correct.
This information is useful to our work.
This information provides a major benefit to our work.
This information is easily accessible.
This information has a good reputation.
This information is sufficiently current for our work.
This information is difficult to interpret.
This information is not protected with adequate security.
This information is of doubtful credibility.
The amount of information does not match our needs.
This information is difficult to manipulate to meet our needs.
This information is not sufficiently timely.

Bewertung subjektiver DQ-Dimensionen	Bewertung subjektiver DQ-Dimensionen
This information is difficult to aggregate.	This information is appropriate for our work.
The amount of information is not sufficient for our needs.	The meaning of this information is easy to understand.
This information is incorrect.	This information is credible.
This information does not add value to our work.	This information covers the needs of our tasks.
This information was objectively collected.	Representation of this information is compact and concise.
It is difficult to interpret the coded information.	This information adds value to our tasks.
The meaning of this information is difficult to understand.	The measurement units for this information are clear.
This information is not sufficiently current for our work. I C	This information is objective.
This information is easily interpretable.	Information can only be accessed by people who should see it.
The amount of information is neither too much nor too little.	This information is sufficiently timely.
This information is accurate.	This information is easy to combine with other information.
Access to this information is sufficiently restricted.	This information is represented in a consistent format.
This information is presented consistently.	This information is easily obtainable.
This information has a reputation for quality.	This information comes from good sources.
This information is easy to comprehend.	This information is quickly accessible when needed.
This information is based on facts.	This information has sufficient breadth and depth for tasks.
This information is sufficiently complete for our needs.	This information presents an impartial view.
This information is trustworthy.	This information is applicable to our work.
This information is relevant to our work.	This information is sufficiently up to date for our work.
Using this information increases the value of our work.	This information is reliable.
This information is presented in a compact form.	

Tabelle E-15: Fragen zur Bewertung von Datenqualität nach HUANG ET AL. [1999]

E.6.2 Bewertung von Datenqualität nach Lee et al.

LEE ET AL. [2002, S. 143f] haben 65 Aussagen (s. Tabelle E-16) zur Bewertung von Datenqualität definiert (Bewertung auf einer elfstufigen Likert-Skala).

Dimension	Zu bewertende Aussagen
Accessibility	This information is easily retrievable. This information is easily accessible. This information is easily obtainable. This information is quickly accessible when needed.
Appropriate Amount	This information is of sufficient volume for our needs. The amount of information does not match our needs. The amount of information is not sufficient for our needs. The amount of information is neither too much nor too little.
Believability	This information is believable.

Dimension	Zu bewertende Aussagen
	<p>This information is of doubtful credibility.</p> <p>This information is trustworthy.</p> <p>This information is credible.</p>
Completeness	<p>This information includes all necessary values.</p> <p>This information is incomplete.</p> <p>This information is complete.</p> <p>This information is sufficiently complete for our needs.</p> <p>This information covers the needs of our tasks.</p> <p>This information has sufficient breadth and depth for our task.</p>
Concise Representation	<p>This information is formatted compactly.</p> <p>This information is presented concisely.</p> <p>This information is presented in a compact form.</p> <p>The representation of this information is compact and concise.</p>
Consistent Representation	<p>This information is consistently presented in the same format.</p> <p>This information is not presented consistently.</p> <p>This information is presented consistently.</p> <p>This information is represented in a consistent format.</p>
Free of Error	<p>This information is correct.</p> <p>This information is incorrect.</p> <p>This information is accurate.</p> <p>This information is reliable.</p>
Interpretability	<p>It is easy to interpret what this information means.</p> <p>This information is difficult to interpret.</p> <p>It is difficult to interpret the coded information.</p> <p>This information is easily interpretable.</p> <p>The measurement units for this information are clear.</p>
Objectivity	<p>This information was objectively collected.</p> <p>This information is based on facts.</p> <p>This information is objective.</p> <p>This information presents an impartial view.</p>
Relevancy	<p>This information is useful to our work.</p> <p>This information is relevant to our work.</p> <p>This information is appropriate for our work.</p> <p>This information is applicable to our work.</p>
Reputation	<p>This information has a poor reputation for quality.</p> <p>This information has a good reputation.</p> <p>This information has a reputation for quality.</p> <p>This information comes from good sources.</p>
Security	<p>This information is protected against unauthorized access.</p> <p>This information is not protected with adequate security.</p> <p>Access to this information is sufficiently restricted.</p> <p>This information can only be accessed by people who should see it.</p>
Timeliness	<p>This information is sufficiently current for our work.</p> <p>This information is not sufficiently timely.</p> <p>This information is not sufficiently current for our work.</p> <p>This information is sufficiently timely.</p> <p>This information is sufficiently up-to-date for our work.</p>
Understandability	<p>This information is easy to understand.</p> <p>The meaning of this information is difficult to understand.</p> <p>This information is easy to comprehend.</p> <p>The meaning of this information is easy to understand.</p>

Dimension	Zu bewertende Aussagen
Value-Added	This information is easy to manipulate to meet our needs. This information is easy to aggregate. This information is difficult to manipulate to meet our needs. This information is difficult to aggregate. This information is easy to combine with other information.

Tabelle E-16: Aussagen zur Bewertung von Datenqualität nach LEE ET AL. [2002]

E.6.3 Bewertung von Datenqualität nach Price et al.

PRICE ET AL. [2008, S. 73f] identifizieren 34 Aussagen (s. Tabelle E-17) für eine subjektive Bewertung von Datenqualität (Bewertung auf einer fünfstufigen Likert-Skala). Die Differenzierung objektiver und subjektiver Datenqualitätsaspekte basiert auf einer semiotischen Definition von Datenqualität [Price/Shanks 2005].

Bewertung subjektiver DQ-Dimensionen	Bewertung subjektiver DQ-Dimensionen
The currency (age or date) of the information is suitable for your needs.	required.
The units of measurement used for retrieved information can be easily changed as needed.	The level of detail presented can be easily changed as needed
The information is missing required values.	Information is not presented at the appropriate level of detail or precision.
The format of retrieved information can easily be changed as needed.	It is easy to modify the types of fields displayed.
Information is appropriately protected from damage.	Information is easily accessible.
You can get information quickly.	The display of information can easily be changed as needed.
Measurement units can be customized.	It is not easy to change the measurement units used to display information.
It is easy to find explanations of terminology, abbreviations, codes, or formatting conventions used in presenting information	Information can easily be collated.
Information is not protected with adequate security.	Information is easy to find and retrieve.
Information is easily obtainable	Some types of information that you need are missing. (Ex. tutorial duration not given)
Information output can be customized as required	The level of detail or precision for information can be modified to suit your needs.
The types of information presented are suitable for your needs.	Information is presented in an appropriate format.
There is duplicate or inconsistent information.	The precision of numeric fields cannot be customized.
Information can only be modified by people who should be able to modify it.	The types of information presented can be easily changed as needed.
Information is quick to retrieve.	Information is quickly accessible when needed.
The display of retrieved information is suitable for your needs.	Information is easy to comprehend.
Information is protected against unauthorized access.	It is easy to interpret what this information means.
Information layout can easily be modified as	

Tabelle E-17: Aussagen zur Bewertung von Datenqualität nach PRICE ET AL. [2008]

E.6.4 Validierungsregeln nach David Loshin

Für DQ-Messungen beschreibt LOSHIN [2001, S. 171-197] die neun in Tabelle E-18 erläuterten Kategorien für Validierungsregeln (org. data quality rules).

Kategorie	Beschreibung
Null Value Rules	Prüfen, ob nur Attribute, die NULL-Werte enthalten dürfen, auch NULL-Werte enthalten.
Value Rules	Prüfen, ob Datenelemente der überprüften Attribute zulässig sind (z. B. durch die Spezifikation erlaubter Wertelisten, Definitionsbereiche oder reguläre Ausdrücke).
Domain Membership Rules	Prüfen, ob Datenelemente einer spezifizierten Domäne angehören (z. B. spezifiziert durch Wertelisten, Definitionsbereiche oder reguläre Ausdrücke).
Domain Mappings	Prüfen unter Verwendung von Mappings (d. h. Zuordnung bestimmter Datenelemente zu Datenelementen einer Liste, z. B. Währungsabkürzungen, Länder-Codes), ob Datenelemente einer spezifizierten Domäne angehören.
Relation Rules	Vergleichen die Datenelemente von zwei oder mehreren Attributen und überprüfen die Existenz einer bestimmten Beziehung (z. B. Gleichheit, gleiche Domäne, Ordnungsrelation).
Table, Cross-table, and Cross-message Assertions	Prüfen spezifizierte Eigenschaften des Datenmodells (z. B. funktionale Abhängigkeiten von Attributen, Schlüssel- und Fremdschlüsselbeziehungen).
In-process Rules	Prüfen Datenelemente in Abhängigkeit von Prozesszuständen (z. B. die Vollständigkeit von Stücklistenpositionen nach der Freigabe einer Stückliste).
Operational Rules	Prüfen die korrekte Transformation von Datenelementen (z. B. Nach Datenübertragung zwischen zwei Datenbehältern) oder die Aktualisierung von Datenelementen zu spezifizierten Zeitpunkten.
Other Rules	Prüfen z. B. Konformität zu unscharfen (org. fuzzy) Datenspezifikationen oder die Nutzung vorgesehener Navigationspfade auf Websites.

Tabelle E-18: Validierungsregeln nach LOSHIN [2001]

E.6.5 Validierungsregeln nach Jack E. Olson

OLSON [2003, S. 143-254] beschreibt fünf verschiedene Verfahren für regelbasierte DQ-Messungen und liefert dadurch die in Tabelle E-19 erläuterte Kategorisierung von Validierungsregeln.

Verfahren / Kategorie	Beschreibung
Column Property Analysis	
Business Meaning	Prüft, ob ein Attribut für den eigentlich gedachten Zweck verwendet wird, d. h. ob die Bedeutung der Datenelemente der fachlichen Attributspezifikation entspricht.
Storage Properties	Prüft, ob ein Datenelement der technischen Attributspezifikation entspricht, d. h. ob z. B. die vorgesehene Anzahl von Zeichen nicht überschritten, die vorgesehenen Datentypen verwendet und die vorgesehene Genauigkeit eingehalten werden.

Verfahren / Kategorie	Beschreibung
Valid Value Property	Prüft, ob ein Datenelement innerhalb des erlaubten Wertebereichs liegt (z. B. Toleranzen, Auswahllisten, Ausschlusslisten).
Empty Condition Rules	Prüft, ob ein Datenelement leer (oder mit einem äquivalenten Wert wie NULL, BLANK, ?, * belegt) ist, d. h. nur angegeben wurde, um z. B. in einem Formular einen erforderlichen Wert anzugeben.
Structure Analysis	
Functional Dependencies	Prüft, ob funktionale Abhängigkeiten, die für Datenobjekttypen spezifiziert sind, von den Datenobjekten erfüllt werden. Eine funktionale Abhängigkeit besteht dann zwischen zwei Attributen (oder Attributgruppen), wenn durch ein Datenelement des einen Attributs das Datenelement des anderen Attributs eindeutig bestimmt werden kann.
Keys	Prüft, ob Datenelemente von Schlüsselattributen und Fremdschlüsselbeziehungen eindeutig sind.
Normal Forms of Data	Prüft, ob die für eine Datenobjekttypen spezifizierte Normalform erfüllt ist.
Synonyms	Prüft, ob als synonym spezifizierte Attribute synonyme Datenelemente enthalten.
Simple Data Rule Analysis	
Dates	Prüft, ob Datums-Datenelemente sinnvoll sind, z. B. sollte ein Bestelldatum früher sein als ein Lieferdatum oder ein Einstellungsdatum später als ein Geburtsdatum.
Durations	Prüft, ob Zeitspannen sinnvoll sind, z. B. sollten zwischen Geburts- und Einstellungsdatum mehr als 16 Jahre liegen.
Object Subgrouping Columns	Prüft eine Menge mit Datenelementen in Abhängigkeit anderer Datenelemente, die die erste Menge in disjunkte Gruppen aufteilt (z. B. sollte für einen männlichen Patienten kein Datum der letzten Geburt vorliegen).
Work Flow	Prüft Datenelemente in Abhängigkeit des Fortschritts eines bestimmten Prozesses (z. B. sollte ein Produkt mit einer noch nicht freigegebenen Stückliste noch nicht auf Lager sein).
Derived-Value Rules	Prüft Datenelemente in Abhängigkeit anderer Datenelemente, für die Richtlinien spezifiziert sind (z. B. sollte bei vollem Lager und wenigen Aufträgen eine Produktionssperre gesetzt sein).
Rules Involving Multiple Rows of Same Column	Prüft Datenelemente in Abhängigkeit von Datenelementen anderer Datenobjekte (z. B. ist für Datenobjekte mit Auftragspositionen ein Produktbezeichner nicht eindeutig, eingeschränkt auf die Positionen eines einzelnen Auftrags aber schon).
Complex Data Rule Analysis	
Dates and Time	Prüft Datums-Datenelemente geschäftsobjektübergreifend in Abhängigkeit anderer Datenelemente (z. B. sollte ein Gegenstand zu einem bestimmten Zeitpunkt nicht an mehr als eine Person ausgeliehen sein).
Location	Prüft Orts-Datenelemente geschäftsobjektübergreifend in Abhängigkeit anderer Datenelemente (z. B. sollten an einem Lagerplatz nicht mehr als ein Lagerobjekt gelagert werden).
Other Types of Exclusivity	Prüft Datenelemente geschäftsobjektübergreifend auf Exklusivität (z. B. sollte eine Frau nicht für zwei Männer als Ehefrau angegeben werden).
Aggregations	Prüft bestimmte Eigenschaften (z. B. Summe, Mittelwert) einer Gruppe von Datenelementen eines Attributs (z. B. sollte die Summe der Arbeitsstunden eines Mitarbeiters pro Woche nicht grösser 50 sein).

Verfahren / Kategorie	Beschreibung
Lookup	Prüft Datenelemente gegen andere Datenelemente, die die gleichen Daten enthalten sollten (z. B. Vergleich konkreter Werte, Korrelationsanalysen).
Value Rule Analysis	
Cardinality	Vergleicht die Kardinalität eines Attributs (d. h. die Anzahl unterschiedlicher Werte) mit der Anzahl der Datenobjekte, wobei eine hohe Abweichung ein Indikator für Dubletten ist.
Frequency Distributions	Prüft, wie oft bestimmte Datenelemente für ein Attribut vorkommen, wobei sowohl häufige Werte (z. B. Platzhalter wie blank oder NULL) als auch seltene Werte (z. B. Tippfehler) Defekte aufzeigen können.
Extreme Values	Identifiziert minimale und maximale Datenelemente eines Attributs (z. B. 0 oder 999 als Altersangabe) als mögliche Datendefekte.

Tabelle E-19: Validierungsregeln nach OLSON [2003]

E.6.6 Exemplarische Validierungsregeln aus partizipativen Fallstudien

Die in Tabelle E-20 zeigt beschreibt einen Auszug der Validierungsregeln, die durch Anwendung der Technik Spezifikation von Validierungsregeln (vgl. Abschnitt 3.5.2.5.1) in partizipativen Fallstudien mit der Konsumgüter AG und der Strom AG definiert worden sind. Es sind jeweils nur die für ein Datenobjekt *d* durchzuführenden Prüfungen beschrieben. die Auswahl von für die jeweilige Prüfung relevanten (d. h. potentiell defekter) Datenobjekte (vgl. Abschnitte 3.5.2.2 und 3.5.2.5) sowie eine diese ausgewählten Datenobjekte iterierende Schleife (vgl. Abbildung B3-15) wird vorausgesetzt.

DQ-Dimension / Regel	Beschreibung
Aktualität / Rechtzeitigkeit	
<pre>if (CurrentTime() + threshold) > FirstDemand(d) if d.BoM.Status != 'released' : return 0 end if return 1</pre> <p><i>CurrentTime()</i> gibt die die aktuelle Zeit zurück. <i>FirstDemand(d)</i> gibt den Zeitpunkt zurück, zu dem das Datenelement <i>d</i> zum ersten Mal gebraucht wird (z. B. Produktionsbeginn).</p>	Prüft, ob eine Stückliste (bill of material, BoM) rechtzeitig (definiert durch die Zeitspanne <i>threshold</i>) freigegeben worden ist.
<pre>if (CurrentTime() + threshold) > FirstDemand(d) if d.CommodityCode == NULL : return 0 if d.CountryOfOrigin == NULL : return 0 end if return 1</pre>	Prüft, ob für die Zollabwicklung benötigte Daten rechtzeitig vorhanden sind.
<pre>app1 = LastChange(A1.d.BoM.Status, 'released') app2 = LastChange(A2.d.BoM.Status, 'released') if app1 > app2 + threshold : return 0 return 1</pre> <p><i>LastChange(a,e)</i> gibt den Zeitpunkt der letzten Belegung des Attributs <i>a</i> mit dem Datenelement <i>e</i> zurück.</p>	Prüft, ob die Freigabe einer Stückliste in Anwendung A1 rechtzeitig auch in Anwendung A2 vorgenommen wurde (eigentlich Konsistenzprüfung, aber aufgrund der Prüfung des Zeitpunkts hier eingeordnet).

DQ-Dimension / Regel	Beschreibung
Änderungshäufigkeit	
<pre>if d.BoM.Status == 'released' for each position p in d.BoM if NumberOfChanges(p) > 1 : return 0 end for end if return 1</pre> <p>NumberOfChanges(a) gibt die Anzahl der seit der ersten Eingabe eines Datenelements vorgenommenen Änderungen des Datenelements des Attributs a zurück.</p>	<p>Prüft, wie oft die Positionen einer Stückliste vor deren Freigabe geändert worden sind. Häufige Änderungen sind ein Indiz für Probleme im Datenpflegeprozess und daher ein Indiz für Datendefekte.</p>
<pre>if NumberOfChanges(d.BoM.Status) > 1 : return 0 return 1</pre>	<p>Prüft, wie oft der Status einer Stückliste geändert wurde. Häufiges Zurücksetzen eines bereits freigegebenen Artikels (systemseitige Repräsentation eines Produkts) ist ein Indiz für Probleme im Datenpflegeprozess und daher ein Indiz für Datendefekte.</p>
Genauigkeit	
<pre>for each data object d2 if d.GTIN == d2.GTIN if GS1(d,d2) == FALSE : return 0 end if end for return 1</pre> <p>GS1(d1,d2) gibt TRUE zurück, wenn die Gewichte von zwei Datenobjekten d1 und d2 nicht mehr als 20% abweichen, sonst FALSE.</p>	<p>Prüft, ob zwei Datenobjekte mit gleicher GTIN die von GS1 geforderten Eigenschaften keine Gewichtsabweichung von mehr als 20% bei gleicher GTIN haben.</p>
<pre>for each data object d2 if d.GTIN == d2.GTIN if d.SH.Pieces != d2.SH.Pieces : return 0 if d.PU.Pieces != d2.PU.Pieces : return 0 if d.PA.Pieces != d2.PA.Pieces : return 0 end if end for return 1</pre>	<p>Prüft, ob zwei Datenobjekte mit gleicher GTIN die von GS1 geforderten Eigenschaften gleiche Stückzahl auf allen logistischen Ebenen bei gleicher GTIN haben (SH: Shrink, PU, Packungseinheit (engl. packaging unit), PA, Palette).</p>
<pre>for each data object d2 if EditDist(d,d2) > threshold : return 0 return 1</pre> <p>EditDist(s1,s2) gibt die minimale Editierdistanz von zwei Zeichenketten s1 und s2 zurück.</p>	<p>Prüft, ob sich zwei Datenobjekte so sehr ähneln, dass eine Dublette zu vermuten ist. Die Prüfung durch die minimale Editierdistanz erfordert Zeichenfolgen als Eingabe und ist daher nur für bestimmte Datenobjekttypen (z. B. Adressdaten) geeignet.</p>
Konsistenz	
<pre>if d.GTIN != ArtworkGTIN(d) : return 0</pre>	<p>Prüft die Konsistenz der auf</p>

DQ-Dimension / Regel	Beschreibung
<pre>return 1 ArtworkGTIN(d) gibt die in ein Produkt-Artwork eingefügte GTIN zurück (ggf. durch Anwendung eines Texterkennungsalgorithmus').</pre>	<p>einem Produkt-Artwork abgebildeten GTIN und der systemseitig gespeicherten GTIN.</p>
<pre>for each local system l if d.GTIN != l.d.GTIN : return 0 if d.SH.GTIN != l.d.SH.GTIN : return 0 if d.VE.GTIN != l.d.PU.GTIN : return 0 if d.PA.GTIN != l.d.PA.GTIN : return 0 end for return 1</pre>	<p>Prüft die Konsistenz logistischer Daten, die von einem zentralen System in lokale Systeme verteilt, und von dort in einen GS1-konformen Daten-Pool übertragen werden.</p>
<pre>for each business partner b pm = NULL // payment method for each account a of b.Accounts if pm == NULL : pm = a.PaymentMethod else if pm != a.PaymentMethod : return 0 end for end for return 1</pre>	<p>Prüft, ob für alle Vertragskonten eines Kunden der gleiche Zahlweg angegeben ist. Einem Kunden ist immer genau ein Vertragskonto zugeordnet. Mehrere Konten für einen Kunden entstehen nur bei einem Umzug (Adressänderung), wenn das alte Konto deaktiviert (aber nicht gelöscht) wird. Falls der vom Kunden präferierte Zahlweg nicht übertragen wird, enthält das deaktivierte Vertragskonto den aktuellen Zahlweg.</p>
Vollständigkeit	
<pre>if d.GrossWeight == NULL : return 0 if d.Height == NULL : return 0 if d.Lenght == NULL : return 0 if d.Width == NULL : return 0 if d.Volume == NULL : return 0 return 1</pre>	<p>Prüft die Vollständigkeit logistischer Daten für ein Produkt.</p>
<pre>if d.Prenamel == NULL : return 0 if d.Name1 == NULL : return 0 if d.Salutation == 'Mr and Mr' OR 'Mr and Ms' OR 'Ms and Ms' if d.Prenamel2 == NULL : return 0 if d.Name2 == NULL : return 0 end if return 1</pre>	<p>Prüft, ob für einen Geschäftspartner Vor- und Nachname angegeben ist und – falls das als Anrede gespeicherte Datenelement dies fordert – auch ein zweiter Vor- und Nachname angegeben ist.</p>

Tabelle E-20: Exemplarische Validierungsregeln

E.7 Dokumentationsvorlagen

E.7.1 Vorlagen für Aktivität I.1

Methode zur Spezifikation geschäftsorientierte Datenqualitätskennzahlen

Aktivität I.1: Betrachtungsbereich definieren Dokumentationsvorlage *Sponsor-Interview*

Rollen

Konzern-Datensteward	
	Name, Kontaktdaten
Auftraggeber	
	Name, Kontaktdaten

Abgeschlossene Projekte

Projekte, in denen Datendefekte oder DQM-Massnahmen diskutiert worden sind

Aktuell laufende Projekte

Projekte, in denen Datendefekte oder DQM-Massnahmen aktuell diskutiert werden

Geplante Projekte

Geplante Projekte, die Datendefekte oder DQM-Massnahmen adressieren

Bekannte Geschäftsprobleme

Bekannte Geschäftsprobleme, die durch Datendefekte verursacht werden

Copyright © CC CDQ2 (<http://cdq.iwi.unisg.ch>), St. Gallen 2010

Universität St. Gallen, Institut für Wirtschaftsinformatik
Lehrstuhl Prof. Dr. Hubert Osterle

Abbildung E-1: Dokumentationsvorlage Auftraggeber-Interview

Methode zur Spezifikation geschäftsorientierte Datenqualitätskennzahlen

Aktivität I.1: Betrachtungsbereich definieren Dokumentationsvorlage *Betrachtungsbereiche*

Rollen

Konzern-Datensteward	
	<small>Name, Kontaktdaten</small>
Prozesseigentümer	
	<small>Name, Kontaktdaten</small>

Geschäftsprozess oder -aktivität

Bezeichner	
	<small>Name, ID</small>
Beschreibung	
	<small>Erzeugte Prozessleistung (z. B. Produkte, Service), strategische Bedeutung für das Unternehmen</small>

Prozesskennzahl

Bezeichner	
	<small>Name, ID</small>
Verantwortlichkeit	
	<small>Name, Kontaktdaten</small>
Beschreibung	
	<small>Geschäftsbezug, Skalenniveau, Masseinheit, Messfrequenz, Zielwert</small>

Geschäftsproblem

Bezeichner	
	<small>Name, ID</small>
Beschreibung	
	<small>Häufigkeit des Problems, verursachte Kosten oder andere Effekte, beeinträchtigte Unternehmensziele</small>
Datendefekte als mögliche Ursache	
	<small>Datendefekte ,die als Ursache des Geschäftsproblems in Frage kommen</small>

Copyright © CC CDQ2 (<http://cdq.iwi.unisg.ch>), St. Gallen 2010

Universität St. Gallen, Institut für Wirtschaftsinformatik
Lehrstuhl Prof. Dr. Hubert Osterle

Abbildung E-2: Dokumentationsvorlage für Betrachtungsbereich

E.7.2 Vorlagen für Aktivität I.2

Methode zur Spezifikation geschäftorientierte Datenqualitätskennzahlen

Aktivität I.2: Daten und IT-Systeme identifizieren Dokumentationsvorlage *Datenbehälter und -einschränkung*

Rollen

Konzern-Datensteward	Name, Kontaktdaten
Prozesseigentümer	Name, Kontaktdaten

Dateneinschränkung

Bezeichner	Name, ID
Geschäftsproblem	Bezeichner, s. Dokumentationsvorlage <i>Betrachtungsbereich</i>
Datenobjekttyp	Name, z. B. Kundendaten, Produktdaten
Datenbehälter	Bezeichner
Verantwortlichkeit	Name, Kontaktdaten
Einschränkungen	<ul style="list-style-type: none"> • • • • • • • • • •

Nur aktive Datenobjekte, bestimmte Produkte, Kunden mit bestimmtem Alter, ...

Copyright © CC CDQ2 (<http://cdq.iwi.unisg.ch>), St. Gallen 2010

Universität St. Gallen, Institut für Wirtschaftsinformatik
Lehrstuhl Prof. Dr. Hubert Osterle

Abbildung E-3: Dokumentationsvorlage für Datennutzung und -einschränkung

Methode zur Spezifikation geschäftsorientierte Datenqualitätskennzahlen

Aktivität I.2: Daten und IT-Systeme identifizieren Dokumentationsvorlage *Leitfragen*

Rollen

Konzern-Datensteward	Name, Kontaktdaten
Prozesseigentümer	Name, Kontaktdaten

Leitfragen

DQ-Dimension	Aktualität, Änderungshäufigkeit, Genauigkeit, Konsistenz, Objektivität, Rechtzeitigkeit, Vollständigkeit, ...
Leitfragen	<ul style="list-style-type: none"> • •
DQ-Dimension	Aktualität, Änderungshäufigkeit, Genauigkeit, Konsistenz, Objektivität, Rechtzeitigkeit, Vollständigkeit, ...
Leitfragen	<ul style="list-style-type: none"> • •
DQ-Dimension	Aktualität, Änderungshäufigkeit, Genauigkeit, Konsistenz, Objektivität, Rechtzeitigkeit, Vollständigkeit, ...
Leitfragen	<ul style="list-style-type: none"> • •
DQ-Dimension	Aktualität, Änderungshäufigkeit, Genauigkeit, Konsistenz, Objektivität, Rechtzeitigkeit, Vollständigkeit, ...
Leitfragen	<ul style="list-style-type: none"> • •

Copyright © CC CDQ2 (<http://cdq.iwi.unisg.ch>), St. Gallen 2010

Universität St. Gallen, Institut für Wirtschaftsinformatik
Lehrstuhl Prof. Dr. Hubert Osterle

Abbildung E-4: Dokumentationsvorlage für Leitfragen zu DQ-Dimensionen

E.7.3 Vorlagen für Aktivität I.3

Methode zur Spezifikation geschäftsorientierte Datenqualitätskennzahlen

Aktivität I.3: Kausalketten identifizieren Dokumentationsvorlage *Kausalketten*

Rollen

Konzern-Datensteward	
	<small>Name, Kontaktdaten</small>
Fachliche(r) Datensteward(s)	
	<small>Name, Kontaktdaten</small>
Datennutzer	
	<small>Name, Kontaktdaten</small>
Technische(r) Datensteward(s)	
	<small>Name, Kontaktdaten</small>
Prozesseigentümer	
	<small>Name, Kontaktdaten</small>

Geschäftsproblem

Bezeichner	
	<small>Name, ID</small>
Prozessaktivität	
	<small>Bezeichner</small>
Beeinträchtigung	
	<small>Welche Unternehmensziele werden wie beeinträchtigt? Häufigkeit, Auswirkung (z. B. Mehraufwand, fehlerhafte Produkte)</small>

Datendefekt

Bezeichner	
	<small>Name, ID</small>
Datenobjekte	
	<small>Bezeichner (s. Dateneinschränkung, Aktivität I.2)</small>
DQ-Dimension	
	<small>s. Dokumentationsvorlage <i>Leitfragen</i></small>
Verursachung	
	<small>Wie verursacht der Datendefekt das Geschäftsproblem? Immer? Nur unter bestimmten Umständen?</small>

Copyright © CC CDQ2 (<http://cdq.iwi.unisg.ch>), St. Gallen 2010

Universität St. Gallen, Institut für Wirtschaftsinformatik
Lehrstuhl Prof. Dr. Hubert Osterle

Abbildung E-5: Dokumentationsvorlage für Kausalkette

E.7.4 Vorlagen für Aktivität II.1

Methode zur Spezifikation geschäftsorientierte Datenqualitätskennzahlen

Aktivität II.1: Anforderungen definieren und priorisieren Dokumentationsvorlage *Anforderungen*

Rollen

Konzern-Datensteward	
	<small>Name, Kontaktdaten</small>
Fachliche(r) Datensteward(s)	
	<small>Name, Kontaktdaten</small>
Auftraggeber	
	<small>Name, Kontaktdaten</small>
Technische(r) Datensteward(s)	
	<small>Name, Kontaktdaten</small>

Anforderungen

Bezeichner		Priorität	
	<small>Name, ID</small>		
Beschreibung			
Bezeichner		Priorität	
	<small>Name, ID</small>		
Beschreibung			
Bezeichner		Priorität	
	<small>Name, ID</small>		
Beschreibung			
Bezeichner		Priorität	
	<small>Name, ID</small>		
Beschreibung			

Copyright © CC CDQ2 (<http://cdq.iwi.unisg.ch>), St. Gallen 2010

Universität St. Gallen, Institut für Wirtschaftsinformatik
Lehrstuhl Prof. Dr. Hubert Osterle

Abbildung E-6: Dokumentationsvorlage für Anforderungen

E.7.5 Vorlagen für Aktivität II.2

Methode zur Spezifikation geschäftsorientierte Datenqualitätskennzahlen

Aktivität II.2: Datenqualitätskennzahlen spezifizieren Dokumentationsvorlage *Validierungsregeln*

Rollen

Konzern-Datensteward	
	<small>Name, Kontaktdaten</small>
Fachliche(r) Datensteward(s)	
	<small>Name, Kontaktdaten</small>
Technische(r) Datensteward(s)	
	<small>Name, Kontaktdaten</small>

Validierungsregel

Bezeichner	
	<small>Name, ID</small>
Geschäftsproblem	
	<small>Bezeichner, s. Dokumentationsvorlage <i>Kausalketten</i></small>
Datendefekt	
	<small>Bezeichner, s. Dokumentationsvorlage <i>Kausalketten</i></small>
Attribute	<ul style="list-style-type: none"> • • •
mögliche Prüfungen	<ul style="list-style-type: none"> • • • • • • •

Welche Eigenschaften der Attribute (z. B. bestimmte Datenelemente, Formate, Vergleiche von Datenelementen) sind ein Defekt?

Copyright © CC CDQ2 (<http://cdq.iwi.unisg.ch>), St. Gallen 2010

Universität St. Gallen, Institut für Wirtschaftsinformatik
Lehrstuhl Prof. Dr. Hubert Osterle

Abbildung E-7: Dokumentationsvorlage für Validierungsregeln

Methode zur Spezifikation geschäftsorientierte Datenqualitätskennzahlen

Aktivität II.2: Datenqualitätskennzahlen spezifizieren Dokumentationsvorlage *DQ-Kennzahl*

Rollen

Konzern-Datensteward	
	Name, Kontaktdaten
Fachliche(r) Datensteward(s)	
	Name, Kontaktdaten
Technische(r) Datensteward(s)	
	Name, Kontaktdaten

Datenqualitätskennzahl

Bezeichner	
	Name, ID
DQ-Dimensionen	s. Dokumentationsvorlagen <i>Leitfragen, Kausalketten</i>
Geschäftsprobleme	<ul style="list-style-type: none"> • • •
	Bezeichner, s. Dokumentationsvorlage <i>Kausalketten</i>
Messpunkt	z. B. Datenbehälter, s. Dokumentationsvorlage <i>Datenbehälter und -einschränkung</i>
Berechnungsvorschrift	Abbildung der Ergebnisse einer Messung mit Validierungsregeln auf einer Skala, Zuweisung einer Masseinheit
Validierungsregeln	<ul style="list-style-type: none"> • • • • •
	Bezeichner, s. Dokumentationsvorlage <i>Validierungsregeln</i>

Copyright © CC CDQ2 (<http://cdq.iwi.unisg.ch>), St. Gallen 2010

Universität St. Gallen, Institut für Wirtschaftsinformatik
Lehrstuhl Prof. Dr. Hubert Osterle

Abbildung E-8: Dokumentationsvorlage für DQ-Kennzahlen

E.7.6 Vorlagen für Aktivität III.1

Methode zur Spezifikation geschäftsorientierte Datenqualitätskennzahlen	
Aktivität III.1: Erfüllung der Anforderungen prüfen	Dokumentationsvorlage <i>Anforderungsprüfung</i>
Rollen	
Konzern-Datensteward	<input style="width: 80%;" type="text"/>
	Name, Kontaktdaten
Fachliche(r) Datensteward(s)	<input style="width: 80%;" type="text"/>
	Name, Kontaktdaten
Technische(r) Datensteward(s)	<input style="width: 80%;" type="text"/>
	Name, Kontaktdaten
Anforderungsprüfung	
Bezeichner	<input style="width: 60%;" type="text"/> Erfüllt? <input style="width: 20%;" type="checkbox"/>
	s. Dokumentationsvorlage <i>Anforderungen</i>
Begründung	<input style="width: 80%;" type="text"/>
	Begründen, falls Anforderung nicht erfüllt, die Spezifikation der DQ-Kennzahlen aber trotzdem in Ordnung ist.
Bezeichner	<input style="width: 60%;" type="text"/> Erfüllt? <input style="width: 20%;" type="checkbox"/>
	s. Dokumentationsvorlage <i>Anforderungen</i>
Begründung	<input style="width: 80%;" type="text"/>
	Begründen, falls Anforderung nicht erfüllt, die Spezifikation der DQ-Kennzahlen aber trotzdem in Ordnung ist.
Bezeichner	<input style="width: 60%;" type="text"/> Erfüllt? <input style="width: 20%;" type="checkbox"/>
	s. Dokumentationsvorlage <i>Anforderungen</i>
Begründung	<input style="width: 80%;" type="text"/>
	Begründen, falls Anforderung nicht erfüllt, die Spezifikation der DQ-Kennzahlen aber trotzdem in Ordnung ist.
Bezeichner	<input style="width: 60%;" type="text"/> Erfüllt? <input style="width: 20%;" type="checkbox"/>
	s. Dokumentationsvorlage <i>Anforderungen</i>
Begründung	<input style="width: 80%;" type="text"/>
	Begründen, falls Anforderung nicht erfüllt, die Spezifikation der DQ-Kennzahlen aber trotzdem in Ordnung ist.
<hr/> Copyright © CC CDQ2 (http://cdq.iwi.unisg.ch), St. Gallen 2010 Universität St. Gallen, Institut für Wirtschaftsinformatik Lehrstuhl Prof. Dr. Hubert Osterle	

Abbildung E-9: Dokumentationsvorlage für die Prüfung von Anforderungen

E.7.7 Vorlagen für Aktivität III.2

Methode zur Spezifikation geschäftsorientierte Datenqualitätskennzahlen

Aktivität III.2: Spezifikation dokumentierenMögliche Dokumentationsgliederung

Einleitung

- Bekannte Datendefekte und durch sie verursachte Geschäftsprobleme der Vergangenheit.
- Welche Potenziale (z. B. bessere Bedarfsplanung, zentraler Einkauf, bessere Entscheidungsunterstützung) könnten mit besserer Datenqualität realisiert werden?

Grundlagen

Datenqualitätsmanagement

- Unterscheidung präventiver und reaktiver DQM-Massnahmen.
- Erläuterung der Nutzung von DQ-Kennzahlen als Werkzeug für präventives DQM.

Datenqualitätsmessung

- Notwendigkeit der Ausrichtung von DQ-Messungen an Unternehmenszielen und Prozessleistung.
- Erläuterung der Konzepte *Kausalkette* und *geschäftsorientierte DQ-Kennzahl*.

Ergebnis Kausalkettenidentifikation

- Liste identifizierter Geschäftsprobleme und Datendefekte, strukturiert als Kausalketten.
- Visualisierung der Bewertung von Geschäftsproblemen (z. B. Heatmap).
- Benennung der durch die DQ-Kennzahlen adressierten Geschäftsprobleme.

Datenqualitätskennzahlen

Kennzahl 1

- Kurze Beschreibung mit Bezeichner und adressiertem Geschäftsproblem
- Adressierte Datendefekte.
- Messpunkt und Dateneinschränkung.
- Validierungsregeln.

Kennzahl 2

...

Struktur Kennzahlensystem

- Erläuterung der ausgewählten Berechnungsvorschrift.
- Zuordnung Validierungsregeln zu Kennzahlen.

Datenqualitätssimulation

- Quantifizierung der Risiken der durch die DQ-Kennzahlen überwachten Datendefekte.
- Simulation verschiedener Szenarien und Erläuterung der Auswirkung von Datendefekten auf Messwerte der DQ-Kennzahlen.

Skizze für Umsetzung

- Priorisierung der spezifizierten DQ-Kennzahlen unter Berücksichtigung des erwarteten Implementierungsaufwands.
- Skizze eines Messsystems (auch Cockpit, Scorecard, Dashboard) zur Visualisierung von Messwerten.

Copyright © CC CDQ2 (<http://cdq.iwi.unisg.ch>), St. Gallen 2010

Universität St. Gallen, Institut für Wirtschaftsinformatik
Lehrstuhl Prof. Dr. Hubert Österle

Abbildung E-10: Gliederungsvorschlag für Dokumentation

E.8 Fokusgruppe und Interview-Leitfaden

Am 9. Februar 2010 haben elf Fachexperten aus zehn Unternehmen im Rahmen eines zweitägigen Workshops des CC CDQ in einem 90-minütigen Fokusgruppeninterview die Methode zur Spezifikation geschäftsorientierter Datenqualitätskennzahlen diskutiert und bewertet. Tabelle E-21 zeigt die Teilnehmer, Tabelle E-22 die Leitfragen (auf Englisch aufgrund englischsprachiger Teilnehmer) des Fokusgruppeninterviews.

Unternehmen	Funktion
Bayer CropScience AG	Integration Manager Enterprise Master Data Management
Beiersdorf AG	Team-Leiter Beiersdorf Shared Services (BSS) Master Data
Beiersdorf AG	Leiter Supply Chain Data Process Management
DB Netz AG	Strategisches Infrastrukturdatenmanagement, geographische Schieneninfrastrukturdaten
Festo AG & Co. KG	Leiter Product Lifecycle Management
Grundfos A/S	Supply Chain Coordinator
Henkel AG & Co. KGaA	Projektmanager Datenqualitäts-Strategie
Metro Group AG	Leiter Business Intelligence Support
Rewe-Zentral AG	Project Manager Data Quality
Siemens Enterprise Communications GmbH & Co. KG	Vice President Master Data
Syngenta Crop Protection AG	Lead Steward Material Master Data

Tabelle E-21: Teilnehmer des Fokusgruppeninterviews am 9. Februar 2010

Relevance perspective
<ul style="list-style-type: none"> • Does the method's objective (i.e. identifying causalities between data defects and business problems, and specifying data quality metrics based on these causalities) address a relevant task from a company's point of view? • Does the method produce a relevant result? • Do the method's output (i.e. the metrics) serve the intended purpose?
Economic perspective
<p>While the method aims at reducing costs (e.g. by providing a structured procedure, examples, and inter-view guidelines), the method's configuration and application cause costs, too.</p> <ul style="list-style-type: none"> • Is the effort caused by the method's configuration and application acceptable, or would any ad-hoc approach serve the purpose as well? • Do the benefits of the method application exceed its costs?
Deployment perspective
<p>In order to foster communication between the stakeholders involved, the method should be understandable. It should correspond to concepts, the prospective method users are familiar with.</p> <ul style="list-style-type: none"> • Are the concepts of the method (i.e. the meta-model, the procedure model) understandable? • Will the approach be accepted in a company? Is there a risk of a —not invented here syndrome?
Engineering perspective
<ul style="list-style-type: none"> • Does the method fulfill the requirements to be taken into account?

<ul style="list-style-type: none">• Is the method's specification and structuring suited for supporting the intended purposes?• Does the method describe the considered aspects correct and in a sufficient depth?• Is the method's structuring clear and understandable?
Epistemological perspective
<ul style="list-style-type: none">• Is the method design process clear and understandable? Are there method elements (e.g. a role, an activity) without an obvious purpose?• Does the method design process generate credibility of the method, or does the method appear to be an ad hoc approach?

Tabelle E-22: Leitfragen des Fokusgruppeninterviews am 9. Februar 2010

Anhang F Anhänge zu Publikation B-4

Metrics and validation rules	Explanation
Metric 'Bill of material' (BoM)	
Validation rule VR01 <pre>if (CurrentTime() + threshold) > FirstDemand(d) if d.BoM.Status != 'released' : return 0 end if return 1</pre>	The BoM (i.e. its position) should only be modified until <i>threshold</i> days before the first demand in order to guarantee the availability of ingredients.
VR02 <pre>for each position p in d.BoM.Positions if d.BoM.Status == 'ready' if NumberOfChanges(p) > 1 : return 0 end if end for return 1</pre>	Modifications of BoM positions should be avoided in order to ensure performance of processes using the BoM (e.g. production in plants).
VR03 <pre>if NumberOfChanges(d.BoM.Status, 'released') > 1 : return 0 return 1</pre>	Modifications of a BoM after its release should be avoided.
Metric 'Customs relevant data'	
VR04 <pre>if (CurrentTime() + threshold) > FirstDemand(d) if d.CommodityCode == BLANK : return 0 if d.CountryOfOrigin == BLANK : return 0 end if return 1</pre>	In order to avoid logistic delays, customs relevant data (i.e. the commodity code and the country of origin) should be present <i>threshold</i> days before the first demand.
Metric 'Documents'	
VR05 <pre>if (CurrentTime() + threshold) > FirstDemand(d) if d.Artwork.Status != 'released' : return 0 if d.Artwork.Document == BLANK : return 0 end if return 1</pre>	A product's artwork should be present (and released) <i>threshold</i> days before the first demand.
VR06 <pre>if d.PackagingFamily.Status == 'released' if d.TechnicalDrawing.Status != 'released' : return 0 end if return 1</pre>	For each specification of a product's packaging material a technical drawing should be available.
VR07 <pre>plm = LastChange(PLM.d.BoM.Status, 'released') mdw = LastChange(MDW.d.BoM.Status, 'released') if plm > mdw + threshold : return 0 return 1</pre>	If a BoM is released in the MDW system it should be released in the PLM system as well (within <i>threshold</i> days).
Metric 'Dangerous-goods indicators' (DGI)	
VR08 <pre>if d.BoM.Status == 'released' if d.DGI == 'no' for each position p in d.BoM.Positions if p.d.DGI == 'yes' : return 0 end for end if end if return 1</pre>	If a product contains dangerous goods, its DGI flag must be set.
Metric 'GTIN'	
VR09 – VR12 (check for GS1 compliance for products (see below) and for logistic units shrink, packaging unit and pallet) <pre>for each data object d2 if d.GTIN == d2.GTIN if GS1(d, d2) == FALSE : return 0 end if end for</pre>	GS1 defines a gross weight tolerance for products (and logistic units) with the same GTIN of 20%.

Metrics and validation rules	Explanation
return 1	
VR13 for each data object d2 if d.GTIN == d2.GTIN if d.Shrink.Item != d2.Shrink.Item : return 0 if d.PackagingUnit.Item != d2.PackagingUnit.Item : return 0 if d.Pallet.Item != d2.Pallet.Item : return 0 end if end for return 1	Two products with the same GTIN are not allowed to have a different number of items on secondary logistic levels.
VR14 if d.GTIN != ArtworkGTIN(d) : return 0 return 1	The data values of a product's GTIN stored in the PLM system and printed on the product's artwork should be equal.
VR15 – VR17 (for product level (see below), and logistic units shrink and packaging unit) if d.GTIN == BLANK : return 0 return 1	GTINs should be defined for all products.
VR18 for each data object d2 if d.GTIN == d2.Shrink.GTIN : return 0 if d.GTIN == d2.VE.GTIN : return 0 if d.GTIN == d2.PackagingUnit.GTIN : return 0 if d.Shrink.GTIN == d2.PackagingUnit.GTIN : return 0 if d.Shrink.GTIN == d2.Pallet.GTIN : return 0 if d.PackagingUnit.GTIN == d2.Pallet.GTIN : return 0 end for return 1	GTINs of all products and logistic levels must be unique. The PLM system already guarantees unique GTINs on product level.
VR19 for each localSAP sap if d.GTIN != sap.d.GTIN : return 0 if d.SH.GTIN != sap.d.Shrink.GTIN : return 0 if d.VE.GTIN != sap.d.PackagingUnit.GTIN : return 0 if d.P.GTIN != sap.d.Pallet.GTIN : return 0 end for return 1	GTIN information is transferred to GS1 not from the central PLM system but from local SAP systems. Therefore, GTINs from PLM should be consistent with GTINs from local SAP systems.
VR20 – VR23 (for product level (see below), and logistic units shrink, packaging unit and pallet) if NumberOfChanges(d.GTIN) > 1 : return 0 return 1	A GTIN change causes effort in various processes (e.g. artwork production) and should be avoided.
Metric 'Logistics data'	
VR24 – VR27 (for product level (see below), and logistic units packaging unit and pallet) if d.GrossWeight == BLANK : return 0 if d.Height == BLANK : return 0 if d.Lenght == BLANK : return 0 if d.Width == BLANK : return 0 if d.Volume == BLANK : return 0 return 1	Logistic data should be defined for all products.
VR28 – VR30 (for logistic units shrink (see below), packaging unit and pallet) if NumberOfChanges(d.Shrink.Items) > 1 : return 0 return 1	Pieces per shrink (and further logistic levels) should not be changed after release.
Metric 'Product marking format' (PMF)	
VR31 if d.PMF == 'no requirements' : return 0 else : return 1	Missing formats of product marking might cause legal complaints.
VR32 if NumberOfChanges(d.PMF) > 1 : return 0 return 1	The format of product marking should be defined only once as corresponding legal

Metrics and validation rules	Explanation
	regulations rarely change.

Table F-1: Data quality metrics and validation rules

The validation rules listed in Table F-1 partially contain functions that provide metadata for data fields or that execute special validation procedures:

- *CurrentTime*. Provides the current time.
- *FirstDemand*. Provides the timestamp a series of products is produced for the first time.
- *NumberOfChanges*. Provides the number of changes of a value of a data attribute.
- *GSI*. Checks if the difference in the gross weights of two products is more than 20 per cent (if this is the case, the products need to be assigned with different GTINs).
- *ArtworkGTIN*. Provides the GTIN indicated on a product's artwork (application of barcode image processing algorithms).

Anhang G Anhänge zu Publikation B-8

G.1 Focus group participants

On November 16, 2007, subject matter experts collected and discussed requirements to be met by a business metadata repository during a 45-minute meeting as part of a two-day workshop. Table G-1 lists the functions of the participants in the respective companies at the time of the workshop.

On December 3, 2008, subject matter experts identified and discussed strong points and weak points of the wiki described in this paper during a 30-minute meeting as part of a two-day workshop. Table G-2 lists the functions of the participants in the respective companies at the time of the workshop.

Company	Function
alfabet AG	Vice President
Bayer CropScience AG	Head of Master Data Services
Daimler AG	Project Manager IT Management
DB Netz AG	Head of Infrastructure Data Management
DB Netz AG	Head of Infrastructure Data Management Strategy
Deutsche Telekom AG	Project Manager Master Data Management
Deutsche Telekom AG	Head of Data Governance
E.ON AG	Head of Master Data Management
ETA SA	Project Manager Organization & IT
ETA SA	Project Manager Organization & IT
Robert Bosch GmbH	Project Manager Master Data Management
simple fact AG	Consultant
ZF Friedrichshafen AG	Head of Organizational Services
ZF Friedrichshafen AG	Project Manager Organizational Services

Table G-1. Participants of focus group specifying requirements to be met by a business metadata repository

Company	Function
Alstom (Schweiz) AG	Director Performance Improvement
B. Braun Melsungen AG	Head of Central Material Master Agency
DB Netz AG	Infrastructure Data Management Strategy
DB Netz AG	Infrastructure Data Management Team
Deutsche Telekom AG	Data Governance Team
Raiffeisen Schweiz Genossenschaft	Quality Manager IT
Raiffeisen Schweiz Genossenschaft	Head of IT Risk Controlling
Syngenta AG	Global Supply IS Manager
Tchibo GmbH	Head of Master Data Management

Table G-2. Participants of focus group evaluating the business metadata repository prototype

G.2 Interview guideline

- A1. Function and responsibilities at Bayer CropScience (department, business processes)?
- A2. Introduction of the business metadata repository (training, documentation, informal activities)?
- A3. Reaction and activities when functions of the business metadata repository are not understood?
- A4. Tasks the business metadata repository is used for (business process, activity, information required)?
- B1. Actual business metadata repository use (incident triggering the use of the handbook, search and navigation, metadata usage, going back to the task to be executed)?
- B2. Metadata needed for a given task (business object, process, structure, guideline, rule)?
- B3. Assessment of metadata found (is it relevant, up to date, well displayed)?
- B4. Frequency of business metadata repository usage?
- B5. Amount of (useful) metadata found? Possible reasons for unsatisfactory search results?
- B6. Reaction and activities when desired information cannot be found in the business metadata repository?
- B7. Weak points of the business metadata repository? Reasons why it is not used more frequently (is metadata not displayed well, is it missing)?
- B8. Suggestions for improvement (to increase usage of the business metadata repository)?
- C1. Clarity of metadata structure? Is all the metadata needed quickly available?
- C2. Which options for navigation are used and how are they used (suggestions for improvement)?
- C3. Which search options are used and how are they used (suggestions for improvement)?
- C4. Which information tabs are used and how often are they used (suggestions for improvement)?

Anhang H Anhänge zu Publikation B-9

Table H-1 lists the participants of a focus group interview of the MDM working group of the German-speaking SAP User Group, taking place on November 25, 2008, in St. Leon-Rot, Germany. The focus group interview started at 9:45 and ended at 12:15. After an introduction of 45 minutes about the context, motivation and current work progress, the first version of the reference model was presented to the 17 focus group participants. The presentation comprised a model overview analogous to Figure B9-5 as well as a detailed view of all six function groups. The participants were asked for their feedback with regard to completeness, consistency, usefulness, and correctness of the model. The focus group interview was recorded and transcribed.

Company	Type	Participant's Function in the Organization
ABeam Consulting (Europe) B.V.	Consulting	Senior Manager
Adolf Würth GmbH & Co. KG	User	Head of Order Processing and Processes
aseaco AG	Consulting	Managing Consultant
cbs Corporate Business Solutions Unternehmensberatung GmbH	Consulting	Senior Project Manager Corporate Data Management
Deloitte Consulting GmbH	Consulting	Senior Manager
gicom GmbH	Consulting	Managing Director
IBSolution GmbH	Consulting	Business Development Manager SOA, Director SOA
IMG AG	Consulting	Consulting Manager
ISO Software Systeme GmbH	Software vendor	Head of Division DQM Sales SAP
Karl Storz GmbH & Co. KG	User	Head of International Master Data Management
Koch BNC AG	Consulting	Consulting Manager
Mieschke Hofmann & Partner	Consulting	Head of Competence Center Strategy, Architecture & Methods
Roche Diagnostics GmbH	User	Head of Global Material Master Management
SAP Deutschland AG & Co. KG	Software vendor	MDM Consultant
Steria Mummert Consulting AG	Consulting	Principal Consultant
T-Systems Enterprise Services GmbH	User	Project Manager and Consultant
Voith Paper Holding GmbH & Co. KG	User	Business Processes & Information Technology, Head of Master Data Management

Table H-1: Participants of focus group A

Table H-2 shows the participants of focus group B, which was part of a regular workshop of the Research Program. The focus group took place on December 3, 2008, in Munich, Germany, and involved 12 participants. The focus group was not specifically

set up on the topic of MDQM functionalities, but was aiming at short-term and medium-term research activities in the field of corporate data management. The participants were asked to name and explain three short-term (for the first 18 months) and one mid-term (18 to 24 months) research priorities. After that a moderated discussion identified seven focus topics, of which one was the Functional Reference Model for MDQM.

Company	Type	Participant's Function in the Organization
ALSTOM Power	User	Chief Data Officer
B. Braun Melsungen AG	User	Head of Central Material Master Agency
BMW AG	User	Manager IT System Landscape
DB Netz AG	User	2 Team members Infrastructure Data Management
Deutsche Post Adress GmbH & Co. KG	User	Sales Manager
Deutsche Telekom AG	User	Team member Master Data Management
EFQM	Standards organization	Chief Architect
IBM Deutschland GmbH	Software vendor	Manager Enterprise Architecture Management Solutions
Nestlé S.A.	User	Manager Business Data Excellence
Raiffeisen Schweiz	User	Manager and team member Data Management
Syngenta Crop Protection AG	User	Team member Master Data Management
Tchibo GmbH	User	Manager Master Data Management

Table H-2: Participants of focus group B

Table H-3 shows the participants of focus group C, which was part of an “in-depth workshop” on corporate data quality management. The focus group took place from 15:45 to 16:45 on February 9, 2009, in Cologne, Germany. The focus group was set up and conducted analogous to focus group A. The results were documented on a flip-chart.

Company	Type	Participant's Function in the Organization
Client Vela GmbH	Consulting	Partner
Corning Cable Systems GmbH & Co. KG	User	Global Data Management Organization Leader
Helsana Versicherungen AG	User	Head of Quality Management PK, Solution Designer (technical/business) DWH
Just.dot GmbH	Consulting	Managing Director
Mars Service GmbH	User	Business Data Manager Europe, Supply Chain Development

Table H-3: Participants of focus group C

Table H-4 shows the participants of focus group D, which was part of a regular workshop of the research project which formed the context of this work. It took place between 10:30 and 11:30 on February 18, 2009, in (blinded for review). The focus group was set up and conducted analogous to focus groups A and C. The results were documented on flip-charts.

Company	Type	Participant's Function in the Organization
ABB Information Systems Ltd.	User	Master Data Consultant and Headquarter IS Architect
Aesculap AG	User	SAP Inhouse Consultant Development and Master Data Management
B. Braun Melsungen AG	User	Head of Central Material Master Agency
Bayer CropScience AG	User	Integration Manager Enterprise Master Data Management
Beiersdorf AG	User	Head of Supply Chain Data Process Management
Beiersdorf Shared Services GmbH	User	Head of Team BSS Master Data
DB Netz AG	User	Head of Strategic Infrastructure Data Management, Railway Geographical Data, Strategic Infrastructure Data Management, Railway Geographical Data
Geberit International AG	User	Head of Group Product Data Management
SBB Cargo AG	User	Head of Business Process Management

Table H-4: Participants of focus group D

The case study at Beiersdorf uses interviews for data collection. Table H-5 gives details on the six interviews conducted.

Date	Organizational unit	Meeting type (and topic)	Number of participants	Duration (hours)
23 Feb. 2010	BSS	Project meeting (conference call, project planning)	1 (BSS)	1
30 Mar 2010	BSS, SC DPM	Assessment workshop (onsite)	2 (BSS), 3 (SC DPM)	4
7 Apr 2010	BSS, SC DPM	Assessment workshop (onsite)	2 (BSS), 3 (SC DPM)	4
31 May 2010	BSS	Project meeting (conference call, discussion assessment analysis)	1 (BSS)	2
1 Sep 2010	BSS	Project meeting (conference call, knowledge sharing planning)	2 (BSS)	1
21 Oct to 8 Nov 2010	BSS, 7 manufacturing and consumer goods companies	Survey (sent by email)	1 (BSS), representatives of 7 peer group companies	n/a

Table H-5: Case Study Interviews

Table H-6 describes all 72 functions of the reference model.

Functions	Description
Master Data Lifecycle Management	
Data Creation	
Conditional Entries	Allows modeling of relations between master data classes that change depending on values of the associated classes. Example: Discount rates for different purchase quantities.
Bulk Editing	Allows to create a number of similar data objects (or particular attributes) at a single time, i.e. the creation process does not have to be executed for single data objects individually.
Plausibility Check	Ensures that no invalid data are entered in enterprise systems. Could use reference lists that contain correct addresses, correct names etc.
Data Maintenance	
Check-out	Prevents data objects from being edited by other users.
Bulk Editing	Allows to edit a number of data objects (or particular attributes) at a single time, i.e. an editing process does not have to be executed for single data objects individually.
Plausibility Check	Ensures that no invalid data are entered in enterprise systems. Could use reference lists that contain correct addresses, correct names etc.
Data Deactivation	
Bulk Editing	Allows to delete / deactivate a number of data objects at a single time, i.e. an editing process does not have to be executed for single data objects individually.
Data Archiving	
Archiving	Allows to persistently store data objects for a defined period of time. Supports compliance with relevant legal provisions.
History Control	Allows to archive different versions of any piece of master data. Ensures that any data object can be reconstructed as it was at a certain point in time.
Metadata Management and Master Data Modeling	
Data Modeling	
Data Model Editing	Allows to modify and adapt classes of master data, in order to, for example, add new attributes or mark existing attributes as mandatory fields.
Graphical Modeling	Data models can be created using graphical symbols (e.g. connecting types of data by lines indicating a relation between these types).
Classification	Allows to group and categorize master data. Assigning data objects to certain categories must not necessarily be unambiguous.
Support of Business Standards	Allows to implement business standards or to take advantage of options offering integration (e.g. import of an XML based standard as a data class for customer data).
Data Model Version Control	Allows to archive different versions of any data model. Ensures that a data model can be reconstructed as it was at a certain point in time.
Model Analysis	

Functions	Description
Dependency Analysis	Verifies the effect of a change in the data structure of a certain data class (e.g. deletion of an attribute) on other data classes.
Data Type Recognition	Allows automatic recognition of data types of existing data objects and automated recognition of existing data models (e.g. when consolidating two customer data sets).
Primary and Secondary Key Recognition	Allows automatic identification of single attributes suited to work as the primary key. Checks the key integrity (e.g. the unambiguousness of a foreign key attribute).
Relationship Recognition	Relations between types of data are automatically recognized. Supports consolidation of different data inventories.
Metadata Management	
Business Rules Documentation	Supports the communication of business rules (may refer to strategic decisions or to system activities), in order to simplify their use and to keep their definitions up to date.
Glossary/Dictionary	Clearly defines a company's central business objects or other elements needed to ensure smooth business processes (e.g. SAP fields).
Metadata Import	Allows to consolidate metadata of different formats. Important as metadata are often stored in distributed systems and in heterogeneous, partially unstructured formats.
Metadata Transport	Allows automatic transfer of metadata from test systems to transaction systems. Important as data structures usually are created in test systems.
Mandatory Fields Administration	Allows central configuration and administration of input fields specified as mandatory fields (e.g. for SAP transactions).
Metadata Publication	Metadata (e.g. business rules or information about business objects) are made available for being used in enterprise systems, where they can be requested with minimum effort.
Metadata Visualization	Uses metadata (threshold values, business rules templates etc.) in order to graphically display complex phenomena in a simplified manner (e.g. diagrams, scorecards).
Data Quality Assurance	
Data Analysis	
Compliance Verification	Allows to verify master data against certain guidelines or legal provisions.
Graphical Analysis	Allows graphical representation of profiles created by means of Profiling (e.g. by illustrating the frequency distribution of values of an attribute).
Plausibility Lists	Provides a basis for other functions (Profiling, Plausibility Check). May contain reference data (e.g. company addresses) or data formats.
Profiling	Allows to analyze data and, based on predefined rules, to create a statistical profile regarding compliance with such rules (basis for e.g. Duplicate Record Recognition).
Data Enrichment	
External Reference Data	Allows to substitute missing data by external data (e.g. a company address register) or to match existing data with external data in order to detect errors in one's database.
Classification Schemes	Supports the use of classification systems (for example eClass, ETIM) for corporate master data management.
Measuring Units	Supports conversion of measuring units (e.g. attributes of dimension or

Functions	Description
	weight).
Multilingual Capability	Allows to make master data and metadata available in various languages at constant data consistency.
Management of Unstructured Data	Allows efficient administration of unstructured data (e.g. artwork) and their relations with master data objects as well as efficient provision of such data.
Data Cleansing	
Delta Import	Allows to import data created or modified since the previous import (the Delta). Identification of the Delta can be useful to search for duplicate records, for example.
Duplicate Recognition	Allows to search for duplicate records. Generates warnings during data entry indicating data duplication.
Pattern Recognition	Identifies certain patterns in data repositories. Patterns allow to define expected data structures or invalid entries.
Plausibility Check	Ensures that no invalid data are entered in enterprise systems. Could use reference lists that contain correct addresses, correct names etc.
Spelling Check	Corrects typical mistakes occurring during data entry. The function can be supported by reference lists used also for Plausibility Check.
Master Data Integration	
Data Import	
Delta Import	Allows to import data created or modified since the previous import (the Delta).
Import Formats	Ensures that only such data can be processed the format of which is understood or which are converted into a known format during the importing process.
Connectors	Allows to create new interfaces for, importing data of a format originally not supported (usually offered as transformation languages (e.g. XSLT) or APIs).
Virtual Integration	Allows to temporarily bring together data from different source systems without needing to copy them into a common database.
Data Transformation	
Field Split	Allows to split values of one field into several values, following predefined rules (e.g. by a separator ' _ ', or ' ; ').
Field Merge	Allows to merge values of several fields.
Data Type Conversion	Allows to consolidate data on the basis of a standard data type (e.g. texts 256 characters long or 64 bits wide).
Pivot Tables	Allows to restructure data classes structured in tables (e.g. by new ordering schemes or inserting rows and columns).
Data Export	
Search Based Data Selection	Allows the explicit selection of data objects to be exported from a list (result of a search query).
Delta Export	Allows to export data created or modified since the previous export (the Delta).
Export Formats	Provides the data formats supported for data export and ensures that data are transferred to transaction systems again after being processed in one

Functions	Description
	way or the other.
Connectors	Allows to create new interfaces for exporting data of a format originally not supported (usually offered as transformation languages (e.g. XSLT) or APIs).
Limitation	Allows to export only a certain data set, what might be helpful in the context of test, for example to estimate the result of a cleansing initiative.
Preview	Allows to view data to be exported as they will be provided.
Cross-Function	
Automation	
Automated Enrichment	Allows automation of enrichment functionality (e.g. comparing data with external reference data, converting measurement units).
Automated Export	Allows (together with Automated Import) to build a system for automated exchange of master data between a test system and transaction systems.
Automated Import	Allows (together with Automated Export) to build a system for automated exchange of master data between a test system and transaction systems.
Cross-Function Automation	Allows automated execution of various, linked functions in a certain sequence (e.g. workflows that do not require human involvement).
Push and Pull Mechanisms	Allows to apply both push-mechanisms and pull-mechanisms for automated data import and export.
Reports	
Data Quality Reports	Allows to illustrate the results of data analyses, e.g. by diagrams to be used in dashboards, or by preconfigured templates for management reports.
Usage Statistics	Allows to record in real-time who is using or requesting which data.
Job Monitoring	Allows to monitor automated functions and assess them by various indicators (e.g. processing time, error rate).
Audit Support	Helps create (e.g. by providing templates or preconfigured analyses) reports demanded by legal provisions.
Search	
Dynamic Value Search	Allows to search for and identify data objects by means of known attribute values. Supported by dynamic sorting and filtering mechanisms.
Free Search	Allows to make full-text queries across the entire database. Search results are provided in a ranking list starting with the result supposed to be of highest relevance.
Fuzzy Search	Provides an extension of Free Search in terms of including similar words and synonyms into the search process (e.g. the name Maier/Meier, or München/Munich).
Workflow Management	
Bundling of Activities	Allows to bundle several activities within a single MDM workflow.
Graphical Workflow Modeling	Allows to model workflows by means of graphical symbols.
Create/Maintain Workflows	Allows to manage sequences of activities across processes and departments. Along the entire data lifecycle numerous activities are executed by nu-

Functions	Description
	merous people.
Administration	
Data History Management	
Data Lineage	Allows to trace back the origin of pieces or sets of master data. Important if master data from various, distributed information systems are consolidated.
Last User	Allows to identify the person who did the last modification in a set or piece of master data or who used a set or piece of master data last.
User Management	
User Interface Design	Allows to adapt the graphical user interface to meet role specific requirements.
Roles and Rights	Allows to define roles and to assign entitlements to execute certain activities by such roles.

Table H-6: Complete reference model overview

Lebenslauf

Persönliche Daten

Geburtsort Siegen, Deutschland

Nationalität Deutsch

Ausbildung

- 2007 – 2010 Universität St. Gallen, Schweiz
Doktorandenstudium der Wirtschaftswissenschaften
Schwerpunkt: Business Innovation
- 2004 – 2005 Ecole Polytechnique Fédérale de Lausanne (EPFL), Schweiz
Austauschstudium mit Stipendium
- 2001 – 2007 Universität Karlsruhe (TH), Deutschland
Studium der Informatik mit Nebenfach Betriebswirtschaftslehre

Berufstätigkeit

- seit 2010 Business Engineering Institute (BEI) St. Gallen, Schweiz
Principal Corporate Data Quality (CDQ)
Unternehmensberatung
- 2007 – 2010 Institut für Wirtschaftsinformatik, Universität St. Gallen, Schweiz
Forschungsassistent am Lehrstuhl Prof. Österle
Kompetenzzentrum Corporate Data Quality (CDQ)
- 2005 – 2006 FIDUCIA AG, Karlsruhe, Deutschland
Studentischer Mitarbeiter
Software-Entwicklung
- 2003 – 2006 Universität Karlsruhe (TH), Deutschland
Studentischer Mitarbeiter
Software-Entwicklung, Unterstützung von Lehrveranstaltungen