



**Methode zur Spezifikation
geschäftsentrierter
Datenqualitätskennzahlen**

Kai M. Hüner

Bericht Nr.: BE HSG / CC CDQ / 13

Lehrstuhl: Prof. Dr. H. Österle

Version: 1.0

Datum: 9. Januar 2011

**Universität St. Gallen -
Hochschule für Wirtschafts-, Rechts-
und Sozialwissenschaften (HSG)**

Institut für Wirtschaftsinformatik
Müller-Friedberg-Strasse 8
CH-9000 St. Gallen
Tel.: +41 71 224 2420
Fax: +41 71 224 2777

Prof. Dr. A. Back
Prof. Dr. W. Brenner (geschäftsführend)
Prof. Dr. R. Jung
Prof. Dr. H. Österle
Prof. Dr. R. Winter

Inhaltsverzeichnis

1	Einleitung.....	9
2	Grundlagen.....	13
2.1	Business Engineering als Bezugsrahmen.....	13
2.2	Datenqualität und Datenqualitätsmanagement.....	15
3	Forschungsmethodik.....	16
3.1	Forschungsprozess	16
3.2	Methoden-Engineering.....	17
4	Stand der Forschung.....	20
4.1	Batini et al.	20
4.2	Caballero et al.....	21
4.3	DAMA.....	22
4.4	Larry P. English.....	22
4.5	Heinrich et al.	24
4.6	IBM	24
4.7	ISO / IEEE.....	25
4.8	David Loshin	25
4.9	Jack E. Olson.....	27
4.10	Price et al.	28
4.11	Thomas C. Redman	29
4.12	Wang et al.....	29
4.13	Vergleichende Bewertung vorgestellter Ansätze	32
5	Methodenentwurf und Anwendung der Methode	35
5.1	Metamodell.....	35
5.2	Methodenfragmente	38
5.2.1	Betrachtungsbereich definieren	41
5.2.2	Daten und IT-Systeme identifizieren	48
5.2.3	Kausalketten identifizieren	50

5.2.4	Anforderungen definieren und priorisieren	56
5.2.5	Datenqualitätskennzahlen spezifizieren	59
5.2.6	Erfüllung der Anforderungen prüfen	69
5.2.7	Spezifikation dokumentieren	69
6	Evaluation der Methode	73
7	Schlussfolgerungen und weiterer Forschungsbedarf	77
Anhang A	Entitätstypen des Metamodells der Methode	87
Anhang B	Datenqualitätsdimensionen	90
B.1	Dimensionen nach Larry P. English	90
B.2	Dimensionen nach David Loshin	91
B.3	Qualitätsdimensionen korrekter Daten nach Jack E. Olson	93
B.4	Dimensionen nach Thomas C. Redman	94
B.5	Dimensionen nach Wang/Strong	95
Anhang C	Kostenarten defekter Daten	97
C.1	Kostenarten nach Larry P. English	97
C.2	Kostenarten nach David Loshin	98
C.3	Kostenarten nach Jack E. Olson	99
Anhang D	Identifikation von Kausalketten	100
D.1	Exemplarische Kausalketten	100
D.2	Leitfragen zur Identifikation kritischer Datendefekte	103
Anhang E	Anforderungen an DQ-Kennzahlen	106
E.1	Anforderungen der DAMA	106
E.2	Anforderungen nach Kaiser et al.	106
E.3	Anforderungen identifiziert in einem Fokusgruppeninterview	107
Anhang F	Messverfahren	109
F.1	Bewertung von Datenqualität nach Huang et al.	109
F.2	Bewertung von Datenqualität nach Lee et al.	110
F.3	Bewertung von Datenqualität nach Price et al.	112
F.4	Validierungsregeln nach David Loshin	112

F.5	Validierungsregeln nach Jack E. Olson.....	113
F.6	Exemplarische Validierungsregeln aus partizipativen Fallstudien	115
Anhang G	Dokumentationsvorlagen	118
G.1	Vorlagen für Aktivität I.1.....	118
G.2	Vorlagen für Aktivität I.2.....	120
G.3	Vorlagen für Aktivität I.3.....	122
G.4	Vorlagen für Aktivität II.1	123
G.5	Vorlagen für Aktivität II.2	124
G.6	Vorlagen für Aktivität III.1	126
G.7	Vorlagen für Aktivität III.2	127
Anhang H	Fokusgruppe und Interview-Leitfaden	128

Abkürzungsverzeichnis

BE	Business Engineering
BE-Methodenkern	Methodenkern des Business Engineering
BoM	Stückliste (engl. bill of material)
bzw.	beziehungsweise
CC CDQ	Kompetenzzentrum (engl. competence center) Corporate Data Quality
CHF	Schweizer Franken
CRM	Customer Relationship Management
d. h.	das heisst
DAMA	Data Management Association
DQ-Dimension	DQ-Dimension
DQ-Kennzahl	DQ-Kennzahl
DQM	Datenqualitätsmanagement
DQ-Messung	Datenqualitätsmessung
DSS	Decision Support Systems
ERP	Enterprise Resource Planning
EUR	Euro
GS1	Global Standards One
GTIN	Global Trade Item Number
IEEE	Institute of Electrical and Electronics Engineers
IEC	International Electrotechnical Commission
IS	Informationssystem
ISO	International Organization for Standardization
org.	Im Original
PLM	Product Lifecycle Management
s.	siehe
TQdM	Total Quality data Management
u. a.	unter anderem
vgl.	vergleiche
z. B.	zum Beispiel

Abbildungsverzeichnis

Abbildung 2-1: Einordnung der Konzepte der Methode in den BE-Methodenkern.....	14
Abbildung 3-1: Metamodell des Methoden-Engineering [Gutzwiller 1994, S. 13]	18
Abbildung 4-1: Prozesse der TQdM-Methodik [English 1999, S. 70]	23
Abbildung 4-2: Prozess zur Messung von Datenqualität [English 1999, S. 72-76]	24
Abbildung 4-3: Messbare Datendefekte [Olson 2003, S. 35]	27
Abbildung 4-4: Regelbasierte Verfahren für DQ-Messungen [Olson 2003, S. 131].....	28
Abbildung 5-1: Kausalketten im Kontext einer Unternehmensarchitektur	36
Abbildung 5-2: Metamodell der Methode	37
Abbildung 5-3: Vorgehensmodell mit Beziehungen zu Meta- und Rollenmodell	39
Abbildung 5-4: Werkzeug zur Simulation von Datenqualitätsmanagement-Szenarien	47
Abbildung 5-5: Leitfragen für ausgewählte DQ-Dimensionen	49
Abbildung 5-6: Workshop-Leitfaden zur Identifikation von Kausalketten	53
Abbildung 5-7: Dokumentationsvorlage zur Bewertung von Kausalketten	54
Abbildung 5-8: Regelschemata mit Beispielen.....	62
Abbildung 5-9: Zuordnung von Validierungsregeln (VR) zu DQ-Kennzahlen	64
Abbildung 5-10: DQ-Messung mit unterschiedlichen Berechnungsvorschriften.....	68
Abbildung 5-11: Simulation von Datendefekten	71
Abbildung 5-12: Durch Datendefekte verursachte Geschäftsprobleme	71
Abbildung 5-13: Überwachung von Datendefekten mit einer DQ-Kennzahl	72
Abbildung G-1: Dokumentationsvorlage Auftraggeber-Interview.....	118
Abbildung G-2: Dokumentationsvorlage für Betrachtungsbereich	119
Abbildung G-3: Dokumentationsvorlage für Datennutzung und -einschränkung.....	120
Abbildung G-4: Dokumentationsvorlage für Leitfragen zu DQ-Dimensionen	121
Abbildung G-5: Dokumentationsvorlage für Kausalkette	122
Abbildung G-6: Dokumentationsvorlage für Anforderungen.....	123
Abbildung G-7: Dokumentationsvorlage für Validierungsregeln	124
Abbildung G-8: Dokumentationsvorlage für DQ-Kennzahlen.....	125
Abbildung G-9: Dokumentationsvorlage für die Prüfung von Anforderungen.....	126
Abbildung G-10: Gliederungsvorschlag für Dokumentation	127

Tabellenverzeichnis

Tabelle 4-1: Bewertung vorgestellter Ansätze zur Messung von Datenqualität.....	34
Tabelle 5-1: Anwendungsunternehmen (anonymisiert) und Projektkontext	40
Tabelle 5-2: Methodenfragment zu Aktivität I.1	41
Tabelle 5-3: Methodenfragment zu Aktivität I.2	48
Tabelle 5-4: Methodenfragment zu Aktivität I.3	50
Tabelle 5-5: Methodenfragment zu Aktivität II.1	57
Tabelle 5-6: Methodenfragment zu Aktivität II.2.....	59
Tabelle 5-7: Methodenfragment zu Aktivität III.1	69
Tabelle 5-8: Methodenfragment zu Aktivität III.2	70
Tabelle 5-9: Parameter (Mittelwert, Varianz) normalverteilter Defekthäufigkeiten	71
Tabelle 6-1: Bewertungen der Fachexperten für die Relevanz-Perspektive.....	73
Tabelle 6-2: Bewertungen der Fachexperten für die ökonomische Perspektive.....	74
Tabelle 6-3: Bewertungen der Fachexperten für die Einführungs-Perspektive.....	74
Tabelle 6-4: Bewertungen der Fachexperten für die Konstruktions-Perspektive.....	75
Tabelle A-1: Definition der Metaentitätstypen der Methode.....	89
Tabelle B-1: DQ-Dimensionen nach ENGLISH [1999].....	91
Tabelle B-2: DQ-Dimensionen nach LOSHIN [2001].....	93
Tabelle B-3: Differenzierung der DQ-Dimension Richtigkeit nach OLSON [2003].....	94
Tabelle B-4: DQ-Dimensionen nach REDMAN [1996].....	95
Tabelle B-5: DQ-Dimensionen nach WANG/STRONG [1996].....	96
Tabelle C-1: Kostenarten defekter Daten nach ENGLISH [1999]	98
Tabelle C-2: Kostenarten defekter Daten nach LOSHIN [2001]	99
Tabelle C-3: Kostenarten defekter Daten nach OLSON [2003]	99
Tabelle D-1: Exemplarische Kausalketten.....	103
Tabelle D-2: Leitfragen zur Identifikation kritischer Datendefekte	105
Tabelle E-1: Anforderungen an DQ-Kennzahlen der DAMA [2009]	106
Tabelle E-2: Anforderungen an DQ-Kennzahlen nach KAISER ET AL. [2007]	107
Tabelle E-3: Anforderungen an DQ-Kennzahlen	108
Tabelle F-1: Fragen zur Bewertung von Datenqualität nach HUANG ET AL. [1999].....	110
Tabelle F-2: Aussagen zur Bewertung von Datenqualität nach LEE ET AL. [2002]	111
Tabelle F-3: Aussagen zur Bewertung von Datenqualität nach PRICE ET AL. [2008].....	112
Tabelle F-4: Validierungsregeln nach LOSHIN [2001]	113
Tabelle F-5: Validierungsregeln nach OLSON [2003].....	115
Tabelle F-6: Exemplarische Validierungsregeln	117
Tabelle H-1: Teilnehmer des Fokusgruppeninterviews am 9. Februar 2010.....	128
Tabelle H-2: Leitfragen des Fokusgruppeninterviews am 9. Februar 2010	129

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 Defekte. 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 Datenqualitätskennzahlen, Datenqualitätsdimensionen, Kausalketten, Kostenarten defekter Daten und Validierungsregeln.

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 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 2 und einer kurzen Beschreibung des Forschungsprozesses in Abschnitt 3 stellt Abschnitt 4 ähnliche Arbeiten vor und

diskutiert deren Ergebnisse in Bezug auf das Forschungsziel dieser Arbeit. Abschnitt 5 stellt die Methode zur Spezifikation geschäftsorientierter DQ-Kennzahlen als Forschungsergebnis vor. Methodenelemente und Beispiele der Methoden Anwendung werden dabei gemeinsam beschrieben, um den Zweck der einzelnen Elemente besser illustrieren zu können. Abschnitt 6 diskutiert Einschränkungen der Methoden Anwendung sowie weiteren Forschungsbedarf.

2 Grundlagen

Die vorgestellte Methode ist Ergebnis gestaltungsorientierter Forschung, die das Metamodell des Methodenkerns des Business Engineering (BE-Methodenkern) als konzeptionellen Bezugsrahmen nutzt. Abschnitt 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 2.2 erläutert daher kurz das der Arbeit zugrunde liegende Verständnis von Datenqualität und DQM.

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 3.2) 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-Engineerings (z. B. Aktivitäten, Rollen, Techniken) und spezifiziert die Gestaltungsobjekte des Kerns in einem Metamodell [Österle et al. 2007, Weber 2009, S. 96-116, 252-257].

Abbildung 2-1 ordnet die Konzepte der vorgestellten Methode in das Metamodell des BE-Methodenkerns unter Berücksichtigung der vorgegebenen Erweiterungsregeln [Höning 2009, S. 219-227] ein. Die mit dem Präfix *BE::* gekennzeichneten Konzepte sind Metaentitätstypen und Relationen des BE-Methodenkerns, erweiterte Metaentitätstypen sind grau eingefärbt. Die Modellierung der erweiterten Konzepte orientiert sich an dem Metamodell der Methode zur Stammdatenintegration von SCHMIDT [2010, S. 101, 221-225].

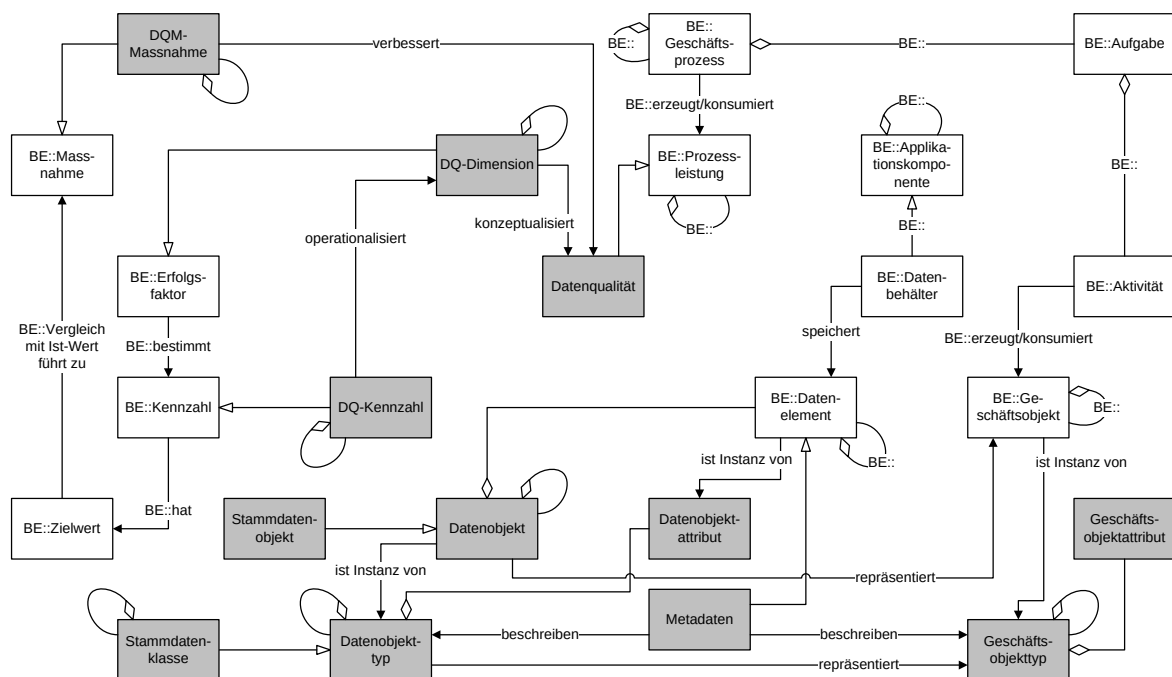


Abbildung 2-1: Einordnung der Konzepte der Methode in den BE-Methodenkern

Die Anwendung der Methode ist nicht auf Stammdaten beschränkt, sondern grundsätzlich für alle Datenelemente, Datenobjektattribute und Metadaten möglich, deren Defekte Geschäftsprobleme verursachen. Falls keine Differenzierung (z. B. bei bestimmten DQ-Dimensionen im Anhang oder konkreten Beispielen in Abschnitt 5.2) nötig ist, wird daher im Folgenden der allgemeine Bezeichner *Daten* verwendet. Die bisher durchgeführten Anwendungen der Methode haben jedoch überwiegend Datendefekte von Datenelementen identifiziert, die die Unternehmen als Stammdaten bezeichnen. Für die Differenzierung verschiedener Datenarten wie Bewegungsdaten, Metadaten und Stammdaten sei auf andere Arbeiten verwiesen [Bur-

nett et al. 1999, S. 184-194, Wedekind 2001, S. 72, Mertens 2007, S. 20, Loshin 2008, S. 8, Smith/McKeen 2008, S. 63].

Abschnitt 5.1 beschreibt das konkrete Metamodell der vorgestellten Methode als Spezifikation der Gestaltungsobjekte des Dokumentationsmodells und strukturiert die in Abbildung 2-1 dargestellten Konzepte durch die Ergebnisdokumente. Abbildung 2-1 zeigt lediglich die Beziehungen zentraler Konzepte der vorgestellten Methode zu Gestaltungsobjekten des BE-Methodenkerns. Anhang A beschreibt die erweiterten Metaentitätstypen, die Gestaltungsobjekttypen des BE-Methodenkerns spezifiziert HÖNING [2009, S. 252-257].

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, S. 5-34], 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, S. 95-105, English 1999, S. 69-81, Eppler/Helfert 2004, S. 311-325]. Eine Übersicht zu verschiedenen Ansätzen für DQM liefern BATINI ET AL. [2009].

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 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 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.

3.1 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 3.2) und Instanziierungen.

Den Forschungskontext des Gestaltungsprozesses der Methode bildet das Kompetenzzentrum Corporate Data Quality¹ (CC CDQ). Das CC CDQ ist Teil des Forschungsprogramms Busi-

¹ <http://cdq.iwi.unisg.ch>

ness Engineering² der Universität St. Gallen und entwickelt seit 2006 als Kooperation des Instituts für Wirtschaftsinformatik³ mit verschiedenen Unternehmen⁴ Lösungen zur Unterstützung des Qualitätsmanagements konzernweit genutzter Daten [Österle/Otto 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 5.2 einzelne Elemente der Methode. Abschnitt 6 erläutert die Evaluation der Methode in einem Fokusgruppeninterview.

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 instanziiieren

² <http://www.iwi.unisg/behsg/>

³ <http://www.iwi.unisg>

⁴ <http://cdq.iwi.unisg.ch/index.php?id=144>

sind und wer die für 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 3-1) gestaltet worden, das die Elemente der folgenden Teilmodelle einer Methode definiert. Eine vergleichbare Strukturierung beschreiben NUSELBEH ET AL. [1996].

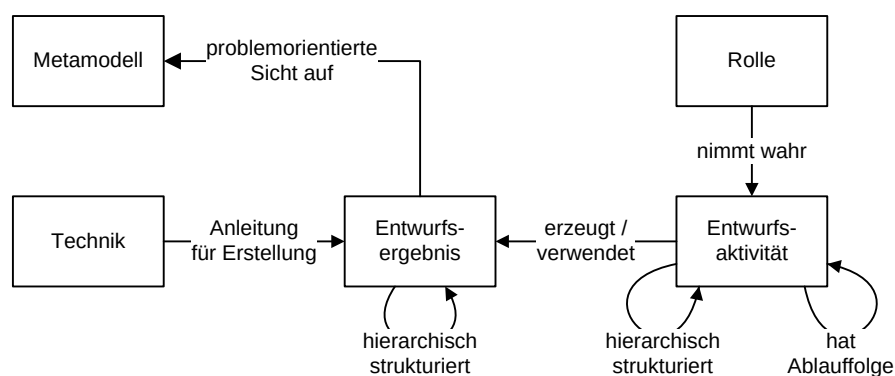


Abbildung 3-1: Metamodell des Methoden-Engineering [Gutzwiller 1994, S. 13]

- **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, dass festlegt, wann und in welcher Abfolge welche Ergebnisse zu gestalten sind, beschreiben Techniken, wie Ergebnisse gestaltet werden.

4 Stand der Forschung

Neben der Differenzierung verschiedener DQ-Dimensionen (vgl. Abschnitt 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 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].

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.

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 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.

- *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.

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.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.

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.

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 4-2, 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].

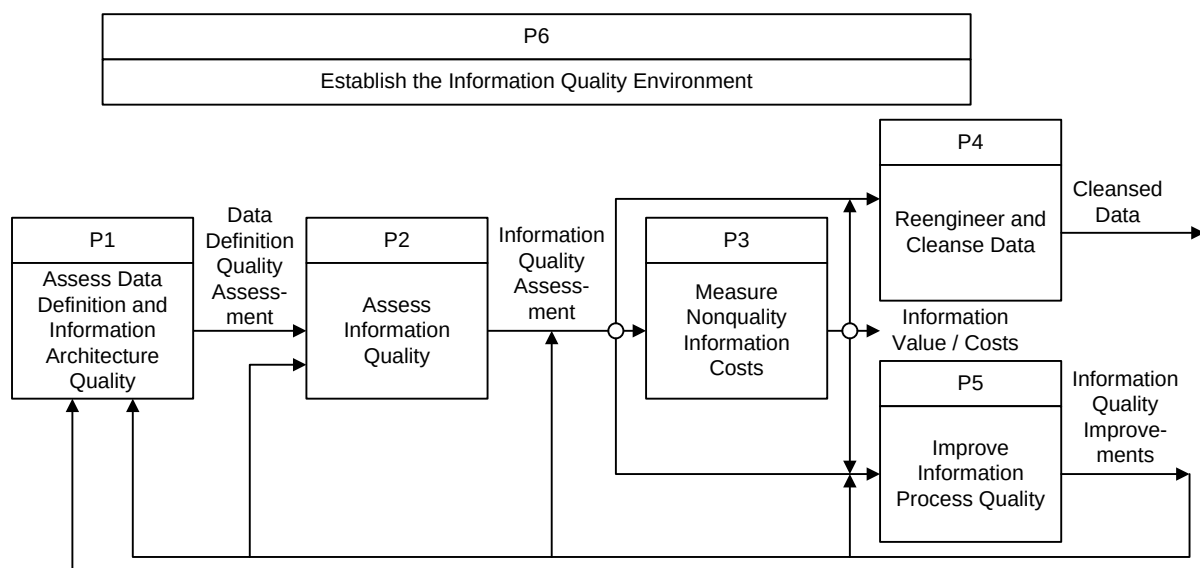


Abbildung 4-1: Prozesse der TQdM-Methodik [English 1999, S. 70]

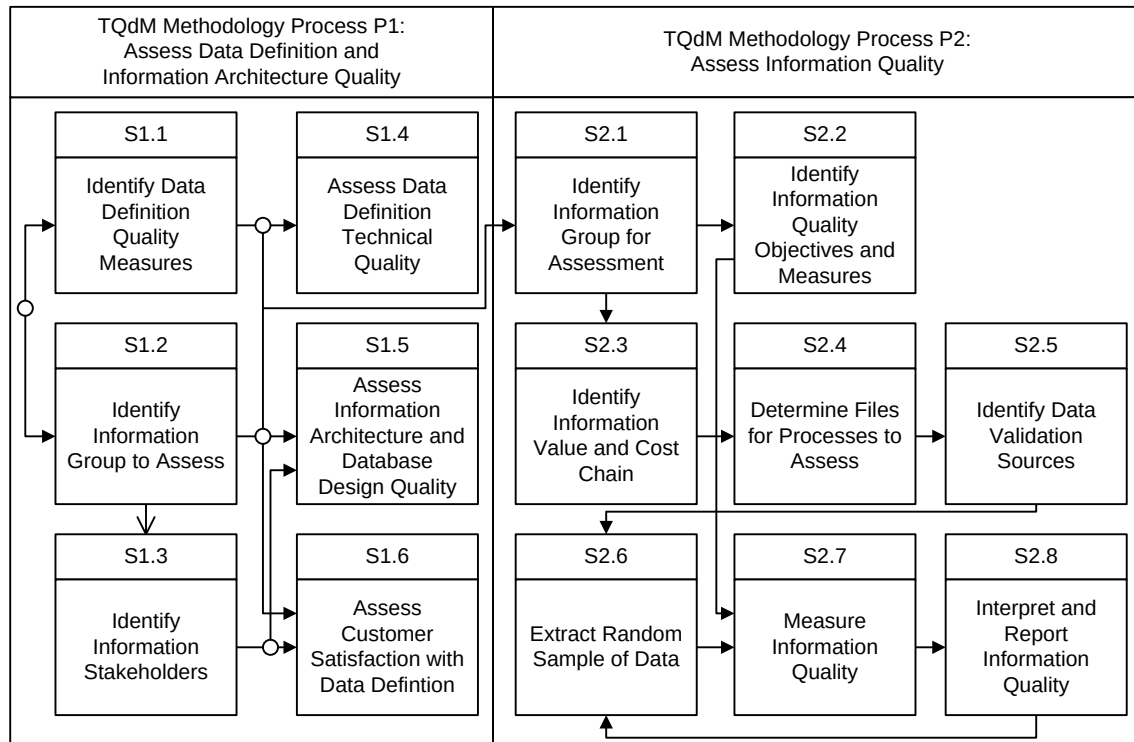


Abbildung 4-2: Prozess zur Messung von Datenqualität [English 1999, S. 72-76]

4.5 Heinrich et al.

KAISER ET AL. [2007, S. 3-4] definieren sechs Anforderungen an DQ-Kennzahlen (s. Anhang E.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].

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 Datentypkorrekturen, Primär- und Sekundärschlüssel-

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

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 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 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.

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 from Defects*) und *Anforderungserfüllung* (org. *Possesses Desirable Characteristics*) sowie verschiedene DQ-Dimensionen (s. Anhang B.2), die durch fünf Kategorien strukturiert sind [Loshin 2001, S. 102-123]. Für DQ-Messungen beschreibt LOSHIN [2001, S. 171-197] neun Kategorien mit Validierungsregeln (org. *Data Quality Rule*,

s. Anhang F.4) sowie verschiedene Beispiele (s. Anhang B.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 Aktivitäts/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 C.2).

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 B.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 4-3).

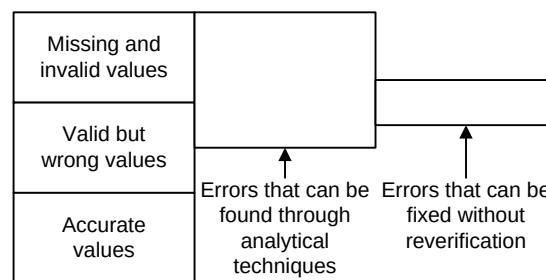


Abbildung 4-3: Messbare Datendefekte [Olson 2003, S. 35]

Zur Messung von Datenqualität (org. data profiling) beschreibt OLSON [2003, S. 121-254] fünf verschiedene Verfahren, die jeweils unterschiedliche Datendefekttypen adressieren (s. Anhang F.5). Alle Verfahren formulieren mögliche Datendefekte als Regeln und identifizieren bei einer Messung die Datenobjekte, die bestimmte Regeln nicht erfüllen (vgl. Abbildung 4-4).

Neben der ausführlichen Beschreibung regelbasierter Messverfahren beschreibt OLSON [2003, S. 103-118] verschiedene Kostenarten (s. Anhang C.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].

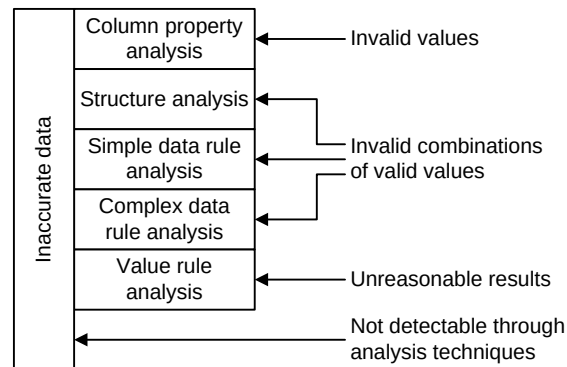


Abbildung 4-4: Regelbasierte Verfahren für DQ-Messungen [Olson 2003, S. 131]

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 F.3) und beschreiben deren Verwendung als Messverfahren.

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 B.4). 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].

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 B.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 2.2) prägen die Autoren das dieser Arbeit zugrunde liegende Datenqualitätsverständnis. Weitere Beiträge konkretisieren diese Perspektive [Strong et al. 1997, 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 F.1).

LEE ET AL. [2002, S. 143-144] definieren 65 Aussagen (vgl. Anhang F.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 11-stufige Likert-Skala von „stimme überhaupt nicht zu“ bis „stimme vollständig zu“).

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.

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 (vgl. Abschnitt 1). 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 3.2 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.

Beitrag	(K1) Vorgehen	(K2) Dokumentationsmodell	(K3) Metamodell	(K4) Rollen	(K5) Techniken	(K6) Definition und Darstellung	(K7) Geschäftsbezug	(K8) Implementierbarkeit
Alur et al. [2007]	◐	○	○	○	●	○	○	○
Ballou et al. [1998]	○	○	○	○	◐	○	○	◐
Batini et al. [2007]	◐	○	○	○	○	○	◐	○
Batini et al. [2009]	◐	○	○	○	○	○	◐	○
Batini/Scannapieco [2006]	○	○	○	○	◐	◐	◐	◐
Caballero et al. [2007]	○	○	◐	○	○	○	○	○
Caballero et al. [2008]	◐	◐	○	○	◐	○	○	○
DAMA [2009]	◐	○	○	○	○	◐	○	○
English [1999]	◐	○	○	◐	◐	◐	◐	○
Heinrich et al. [2009]	◐	○	○	○	○	○	○	◐
Huang et al. [1999]	○	○	○	○	◐	◐	◐	◐
EE [1990]	○	○	◐	○	○	○	○	○
IEEE [1998]	○	○	◐	○	○	○	○	○
ISO/IEC [2007]	○	○	◐	○	○	○	○	○
Kaiser et al. [2007]	○	○	○	○	◐	○	○	◐
Lee et al. [2002]	◐	○	○	○	◐	○	○	◐
Lee et al. [2006]	○	○	○	○	○	◐	○	◐
Loshin [2001]	◐	○	○	○	○	◐	◐	◐
Olson [2003]	◐	○	○	○	◐	◐	◐	◐
Pipino et al. [2002]	○	○	○	○	◐	◐	○	◐

Beitrag	(K1) Vorgehen	(K2) Dokumentationsmodell	(K3) Metamodell	(K4) Rollen	(K5) Techniken	(K6) Definition und Darstellung	(K7) Geschäftsbezug	(K8) Implementierbarkeit
Pipino et al. [2005]	○	○	○	○	◐	◐	○	◐
Price et al. [2008]	○	○	○	○	◑	○	○	◑
Price/Shanks [2005]	○	○	○	○	○	◑	○	○
Redman [1996]	○	○	○	○	○	◐	○	◐
Wand/Wang [1996]	○	○	○	○	○	◑	○	○
Wang/Strong [1996]	○	○	○	○	○	●	○	○
○: 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 4-1: Bewertung vorgestellter Ansätze zur Messung von Datenqualität

Tabelle 4-1 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.

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 5.2.1 bis 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.

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 5-1 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 5-1, links). Für die Identifikation konkreter Kausalketten (vgl. Abbildung 5-1, 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.

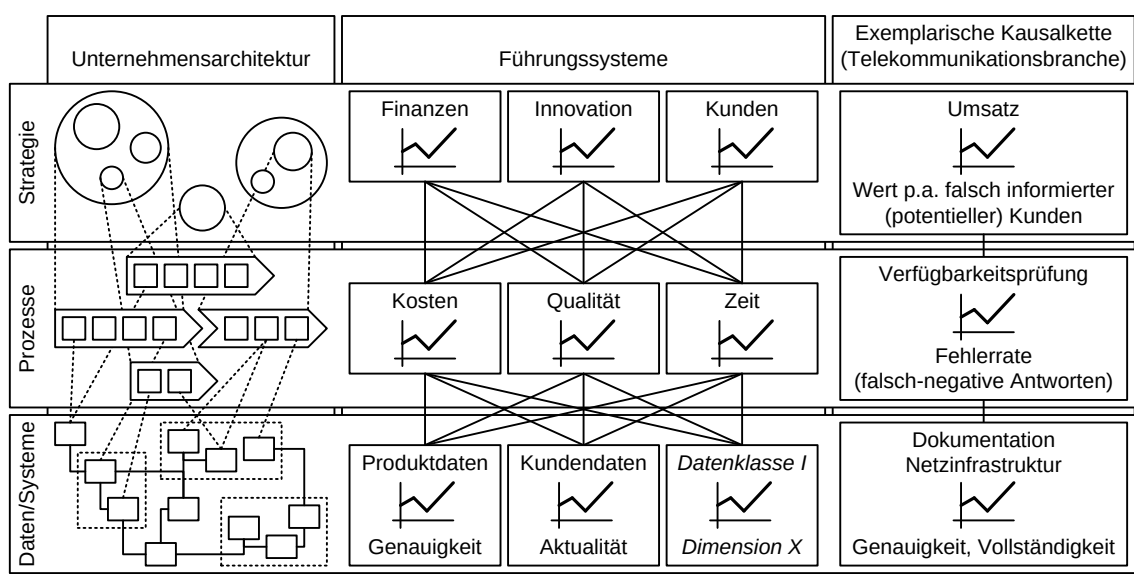


Abbildung 5-1: Kausalketten im Kontext einer Unternehmensarchitektur

Da strategische Erfolgsfaktoren meist durch Prozesskennzahlen oder strategische Führungssysteme (z. B. Balanced Scorecard [Kaplan/Norton 1992]) überwacht werden (vgl. Abbildung 5-1, Mitte), sind diese Überwachungsstrukturen sowohl bei der Identifikation der Kausalketten, als auch bei der Spezifikation von DQ-Kennzahlen zu berücksichtigen (vgl. Abschnitt 5.2.5).

Abbildung 5-2 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 5.2.1 – 5.2.7).

Die Definition des Konzepts *Datendefekt* basiert auf der in Abschnitt 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 Datenobjektyps 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.

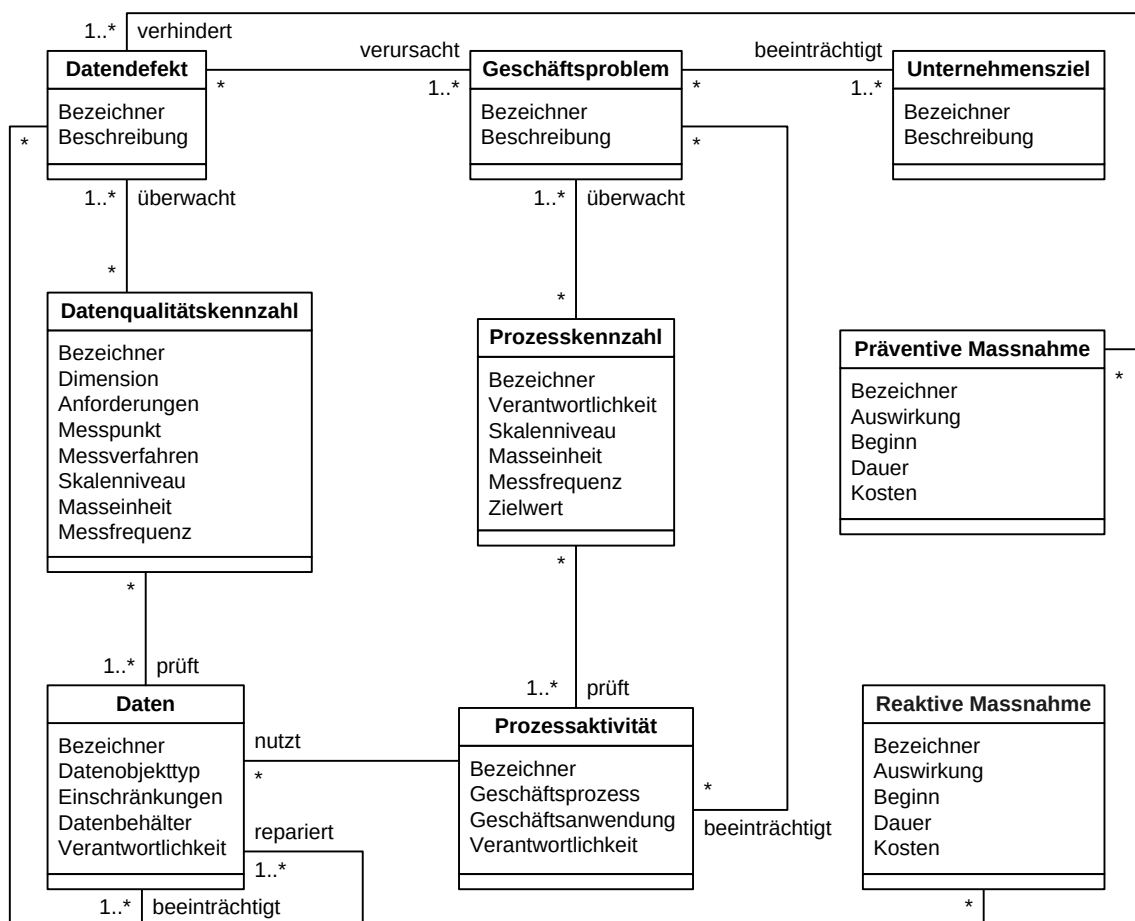


Abbildung 5-2: Metamodell der Methode

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 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.

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 5.2.1 bis 5.2.7 in Form von Methodenfragmenten (jeweils Entwurfsergebnis, erzeugende Aktivität, beteiligte Rollen und genutzte Techniken) beschrieben [Bucher 2009, S. 49]. Abbildung 5-3 zeigt eine Übersicht der Methode mit den Aktivitäten (linke Seite) und den jeweils beteiligten Rollen (rechte Seite). Die mittlere Spalte von Abbildung 5-3 zeigt die Attribute des Metamodells, die die angegebenen Rollen in der jeweiligen Aktivität instanziiieren und die somit das Ergebnis der einzelnen Aktivitäten bil-

den. Das „*-Symbol bedeutet, dass jeweils alle Attribute des angegebenen Metamodell-Entitätstyps instanziiert werden.

Aktivitäten	Ergebnisse	Rollen					
		Konzern-Datensteward	Auftraggeber	Prozess-eigentümer	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 5-3: Vorgehensmodell mit Beziehungen zu Meta- und Rollenmodell

Grundsätzlich gliedert sich die Methode in die folgenden drei Phasen:

- *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 5.2.1 bis 5.2.7 enthalten neben der Beschreibung der Methode Beispiele, die die Instanziierung bzw. Nutzung einzelner Methodenelemente in drei Unternehmen schildern.

Tabelle 5-1 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	Netzinfrastruktur-, Kundendaten

Tabelle 5-1: Anwendungsunternehmen (anonymisiert) und Projektkontext

5.2.1 Betrachtungsbereich definieren

Ziel von Aktivität I.1 ist die Definition möglicher Betrachtungsbereiche (vgl. Tabelle 5-2).

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 • Auftraggeber • Konzern-Datensteward • Prozesseigner	Techniken • Interview/Workshop • Datenqualitätssimulation
Checkliste ☐ Prozessaktivitäten und Geschäftsvorfälle mit bekannten oder vermuteten Geschäftsproblemen (potentiell verursacht durch Datendefekte) sind identifiziert. ☐ Prozesskennzahlen zur Leistungsüberwachung der identifizierten Prozessaktivitäten sind identifiziert. ☐ Geschäftsprobleme und Datendefekte als potentielle Ursache sind beschrieben. ☐ Kosten, die durch identifizierte Geschäftsprobleme entstehen, sind beschrieben.		

Tabelle 5-2: Methodenfragment zu Aktivität I.1

Ein Betrachtungsbereich fokussiert einen Ausschnitt der Unternehmensarchitektur (z. B. Prozessaktivitäten, Geschäftsobjekte, Datenobjekttypen) eines Unternehmens (Abbildung 5-1), 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, Um-

satzsteigerung). 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 Netzinfrasturdaten 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 vereinzelte 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.

5.2.1.1 Rollen

Die Rollen der Methode basieren auf einem Referenzmodell für Data Governance [Weber 2009, S. 106-114, Weber et al. 2009, S. 9-13]. Die zentrale Rolle ist der Konzern-Datensteward (vgl. Abbildung 5-3). 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 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.

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 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 C), exemplarische Kausalketten (s. Anhang D.1) und eine Dokumentationsvorlage (s. Anhang G.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 Anhang C, D.1 und G.1 können dabei unterstützen.

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 5-4 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 5-4). Ein Szenario um-

fasst 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 Simulationsschritt reduzieren. Die Simulation eines konkreten Szenarios für die Telefon AG wird in Abschnitt 5.2.7 beschrieben.

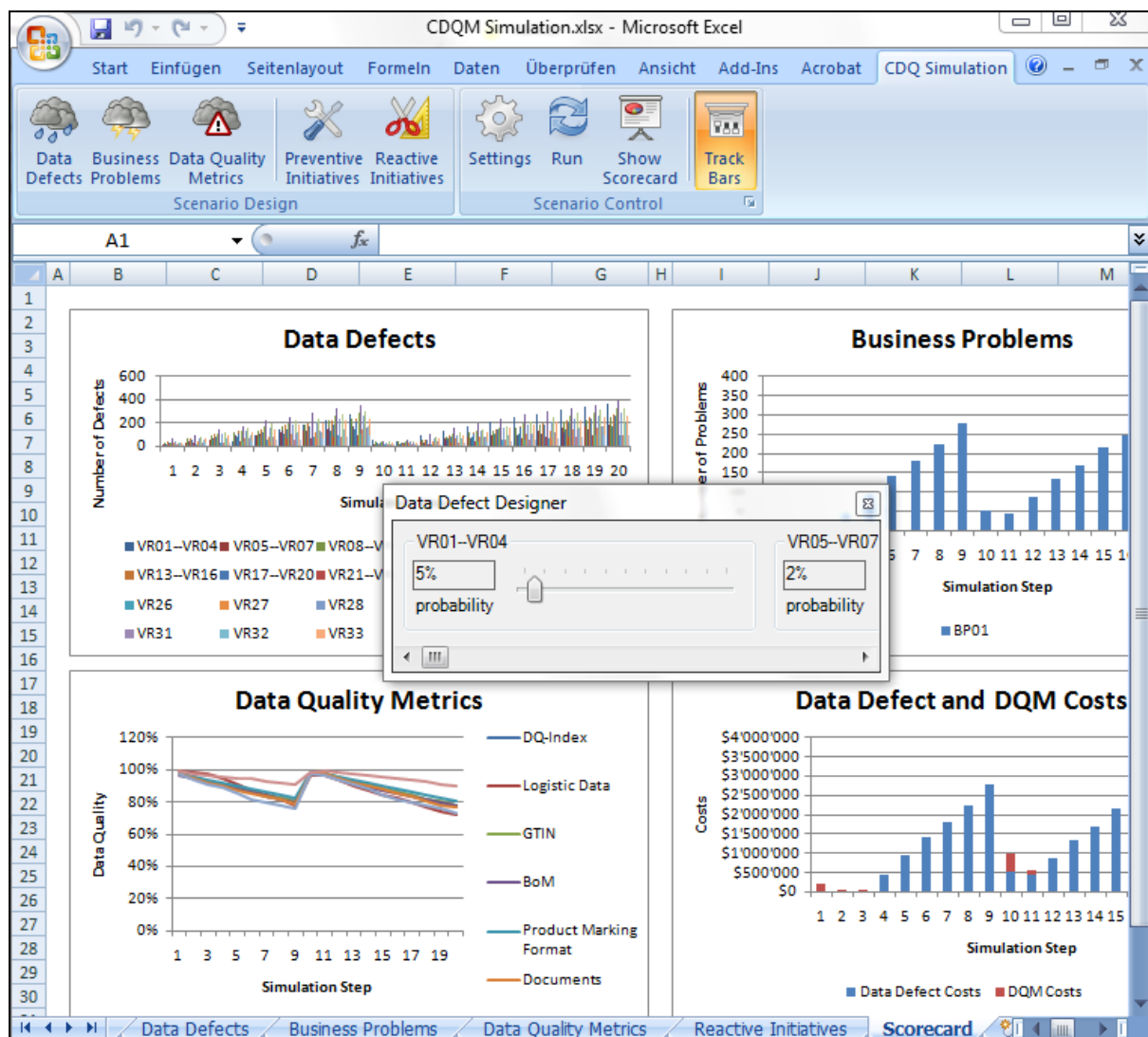


Abbildung 5-4: Werkzeug zur Simulation von Datenqualitätsmanagement-Szenarien

Bei der Definition möglicher Betrachtungsbereiche kann das in Abbildung 5-4 dargestellte Werkzeug zur Unterstützung der Interviews mit Auftraggeber und Prozesseigentümern ge-

nutzt werden, um diskutierte Zusammenhänge zu veranschaulichen und exemplarische Kausalketten zu erklären. Die Methode stellt für die in Anhang D.1 beschriebenen Kausalketten modellierte Szenarien bereit, die mit dem Werkzeug (s. Abbildung 5-4) angepasst und simuliert werden können. Für eine detailliertere Beschreibung der Technik und dem verwendeten Simulationsverfahren sei auf OTTO ET AL. [2010] verwiesen.

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 5-3). 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 G.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 - Konzern-Datensteward - Prozesseigentümer	Techniken Interview/Workshop
Checkliste ☐ Pro Betrachtungsbereich sind die betrachteten Daten möglichst exakt (d. h. Datenobjekttyp und Einschränkungen) beschrieben. ☐ Pro Betrachtungsbereich sind genutzte Datenbehälter und Geschäftsanwendungen identifiziert. ☐ Pro Betrachtungsbereich sind fokussierte DQ-Dimensionen ausgewählt.		

Tabelle 5-3: 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.

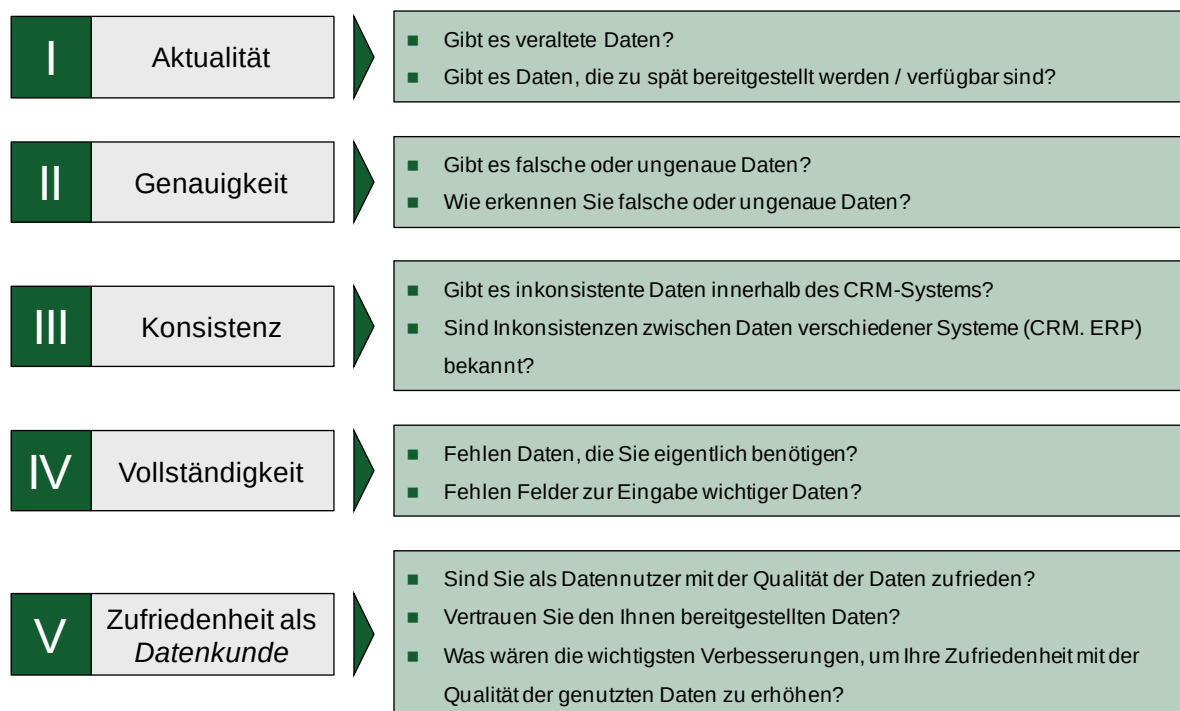


Abbildung 5-5: Leitfragen für ausgewählte DQ-Dimensionen

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 2 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 B.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.

Die Methode gibt keine Auswahl von DQ-Dimensionen vor, sondern stellt in Anhang A verschiedene Listen zur Orientierung bereit. Der Konzern-Datensteward sollte aus diesen Vorschlägen ca. vier bis acht Dimensionen auswählen und pro Dimension Leitfragen (vgl. Abbildung 5-5) zur Erläuterung formulieren. Die Anhänge F.1, F.2 und F.3 sowie eine Dokumentationsvorlage aus Anhang G.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.

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 5-4). Der Identifikationsprozess sollte *Top-Down* verlaufen, d. h. der Konzern-Datensteward sollte versuchen, ausgehend von bekannten Geschäftsproblemen verursachende Datendefekte zu identifizieren.

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 - Fachliche Datenstewards - Konzern-Datensteward - Datennutzer - Technische Datenstewards - Prozesseigentümer	Techniken - Datenqualitätsanalyse - Datenqualitätssimulation - Interview/Workshop
Checkliste ☐ Kausalketten sind komplett (d. h. Datendefekte, Geschäftsprobleme und Auswirkungen) beschrieben. ☐ Kausalitäten (Geschäftsproblem / Auswirkung auf strategisches Unternehmensziel, Datendefekt / Geschäftsproblem) sind exakt (d. h. möglichst quantifiziert) beschrieben.		

Tabelle 5-4: Methodenfragment zu Aktivität I.3

Die Technik *Interview/Workshop* unterstützt den Identifikationsprozess durch Interviews mit verschiedenen Experten des Unternehmens (vgl. Abschnitt 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 5.2.3.2) die Identifikation statistisch auffälliger Datenelemente.

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 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 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.

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 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).

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 ...	(0 – 5)	(0 – 5)	- Frauen werden als „Herr“ angeschrieben und umgekehrt. - Keine Anrede vorhanden. ...
Falsche Daten?	Beschreibung ...	(0 – 5)	(0 – 5)	
Inkonsistente Daten?	Beschreibung ...	(0 – 5)	(0 – 5)	
Fehlende Daten?	Beschreibung ...	(0 – 5)	(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 5-6: Workshop-Leitfaden zur Identifikation von Kausalketten

Die Strom AG hat einen Workshop mit insgesamt neun Mitarbeitern aus den Bereichen Abrechnung Grosskunden, Abrechnung Tarifykunden, 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 5-6).

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 5-7 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.

	A	B	C	D	E	F	G
				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?	Kurzbezeichnung (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 5-7: 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 eine 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 G.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 D.2 stellt dazu Leitfragen bereit.
- *Abteilungsspezifischer Workshop.* Workshop mit mehreren Mitarbeitern einer Fachabteilung (z. B. Finanzabteilung, Marketing, Produktentwicklung), einer Unterstützungsfunktion (z. B. Entscheidungsunterstützung, Qualitätsmanagement) 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 D.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 5-2 und Dokumentationsvorlage in Anhang G.5).

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 5-6). 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.

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 • Fachliche Datenstewards • Konzern-Datensteward • Auftraggeber • Technische Datenstewards	Techniken • Interview/Workshop
Checkliste ☐ Die definierten Anforderungen berücksichtigen alle Unternehmensrichtlinien (z. B. bereits existierende Kennzahlensysteme, Berichtsvorlagen, Normen). ☐ Der Anforderungskatalog enthält keine widersprüchlichen Anforderungen. ☐ Die Priorisierung repräsentiert die Wichtigkeit der einzelnen Anforderungen.		

Tabelle 5-5: Methodenfragment zu Aktivität II.1

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.

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). Diese Anforderungen können die Diskussion in einem Workshop (vgl. Technik *Interview/Workshop*, Abschnitt 5.2.1.2) zwischen Konzern-Datensteward und verschiedenen Data Stewards (die für die von den Kausalketten betroffenen Datenobjekttypen zuständigen Stewards) unterstützen. Anhang G.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 Att-

tribute 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 5.2.1 und 5.2.3), der Verwendung von Validierungsregeln als Messtechnik (vgl. Abschnitt 5.2.5.1) und der Abbildung von Messwerten auf einer Prozentskala (vgl. 5.2.5.2), gewährleistet grundsätzlich die Erfüllung verschiedener Anforderungen.

- *Aggregierbarkeit.* Die Verwendung einer einheitlichen Berechnungsvorschrift (vgl. *Technik Spezifikation von Berechnungsvorschriften*, Abschnitt 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 5.2.5.1) gewährleistet, dass Messwerte für die spezifizierten DQ-Kennzahlen ermittelt werden können.

- *Normierung.* Die Verwendung der in Abschnitt 5.2.5.2 erläuterten Berechnungsvorschriften gewährleistet die Interpretation von DQ-Messungen auf einer Verhältnisskala und somit die Vergleichbarkeit von Messwerten.

5.2.5 Datenqualitätskennzahlen spezifizieren

Ziel von Aktivität II.2 ist die Spezifikation von DQ-Kennzahlen zur Überwachung der zuvor identifizierten Datendefekte (vgl. Tabelle 5-6). 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.

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.		
Ergebnisdokument Liste mit Spezifikationen von DQ-Kennzahlen	Rollen • Fachliche Datenstewards • Konzern-Datensteward • Technische Datenstewards	Techniken • Datenqualitätssimulation • Interview/Workshop • Spezifikation von Berechnungsvorschriften • Spezifikation von Validierungsregeln
Checkliste ☐ 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. ☐ Die Spezifikation der Messverfahren berücksichtigt alle zuvor dokumentierten Kausalketten, soweit keine Begründung, warum die Kausalkette nicht berücksichtigt wird, angegeben ist. ☐ Alle spezifizierten Messverfahren können an den spezifizierten Messpunkten implementiert werden.		

Tabelle 5-6: Methodenfragment zu Aktivität II.2.

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 5-2) 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.

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 F.1, F.2 und F.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 G.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 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 5.2.5.2). So verwenden zudem alle Kennzahlen die gleiche Skala und Masseinheit. Zur Unterstützung des Konzern-Datenstewards stellt Anhang G.5 Dokumentationsvorlagen für die Spezifikation von Validierungsregeln und die Aggregation der Regeln durch eine Berechnungsvorschrift bereit.

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.

```

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.Attribute1 < d.Attribute2 : return 0
    if d.Attribute1 + 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.Attribute1.Type != d.Attribute2.Type : return 0

    return 1 // Datenobjekt ist korrekt
end for

```

Abbildung 5-8: Regelschemata mit Beispielen

Ein Datendefekt ist eine Eigenschaft eines Attributs oder eines Datenelements, die nicht den für den zugehörigen Datenobjektyp spezifizierten (kontextspezifischen) Anforderungen entspricht. Ein fehlendes Datenelement eines Attributs zur Repräsentation des Vornamens eines Kunden ist beispielsweise ein Datendefekt, wenn in einer Prozessaktivität Kunden personali-

sierte 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 5.2.5.2). Basierend auf einer Literaturrecherche (vgl. Abschnitt 3) und Projekterfahrung (vgl. Beispiele) können die folgenden Regelschemata unterschieden werden. Abbildung 5-9 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.

Die Anhänge F.4, F.5 und F.6 enthalten verschiedene Beispiele von Validierungsregeln und können den Konzern-Datensteward bei der Spezifikation unterstützen. Anhang G.5 stellt ausserdem eine Dokumentationsvorlage bereit, die ein einfaches Schema zur Dokumentation von zu prüfenden und zu vergleichenden Attributen enthält (vgl. Abbildung 5-9).

Der Konzern-Datensteward kann diese Vorlage in Interviews (vgl. Abschnitt 5.2.1.2) mit den für den jeweils fokussierten Datenobjekttyp zuständigen Data Stewards nutzen, um diskutier-

te Datendefekte direkt in einer Struktur zu dokumentieren, die der Prüf- und Vergleichslogik der Validierungsregeln ähnelt. Eine weitere Dokumentationsvorlage aus Anhang G.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.

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)

VR02: Anrede Rolle Organisation (Vollständigkeit)

VR03: Anrede (Genauigkeit)

VR04: Name (Vollständigkeit)

VR05: Anschrift (Vollständigkeit)

VR06: Anschrift (Genauigkeit, Dubletten)

VR07: Sprache (Konsistenz)

VR08: Abrechnungssperrgrund (Genauigkeit)

VR09: Rechnungs- / Mahnempfänger (Konsistenz)

VR10: Rechnungsformular (Genauigkeit)

VR11: Eingangs- / Ausgangszahlweg (Konsistenz)

Abbildung 5-9: Zuordnung von Validierungsregeln (VR) zu DQ-Kennzahlen

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.

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 5-8).

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

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 Datenobjekt-

typs 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.

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

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.

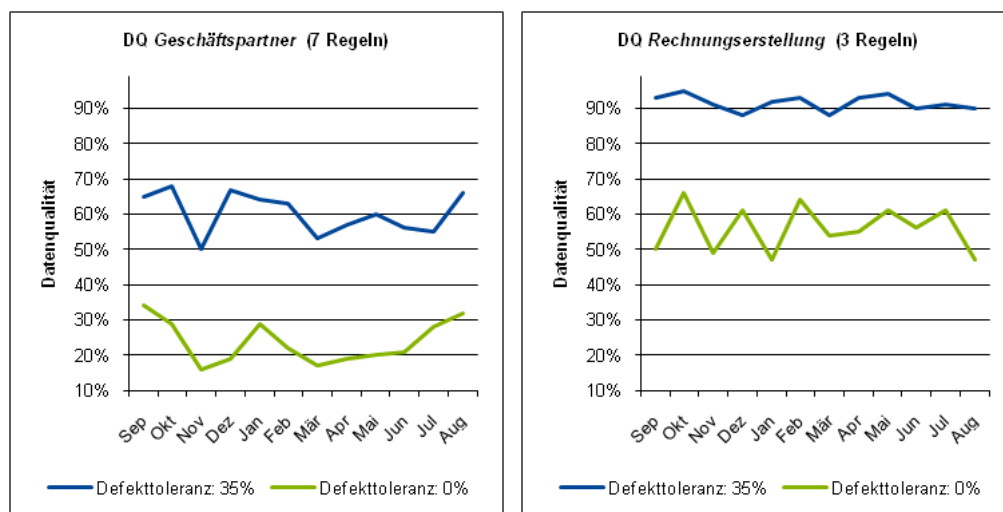


Abbildung 5-10: DQ-Messung mit unterschiedlichen Berechnungsvorschriften

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 5-10 zeigt simulierte Messwerte für zwei Kennzahlen, jeweils mit einem Schwellwert von 0% (grün) und 35% (blau).

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 5-7). 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 5.2.4 für Aktivität II.1 beschrieben, sollte der Konzern-Datensteward die Anforderungserfüllung gemeinsam mit den Data Stewards in einem Workshop (vgl. Abschnitt 5.2.1.2) prüfen. Die Dokumentationsvorlage aus Anhang G.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 - Fachliche Datenstewards - Konzern-Datensteward - Technische Datenstewards	Techniken Interview/Workshop
Checkliste ☐ Alle hoch priorisierten Anforderungen sind erfüllt. ☐ Für alle nicht erfüllten Anforderungen, ist pro Kennzahl, die die Anforderung nicht erfüllt, eine Begründung angegeben.		

Tabelle 5-7: Methodenfragment zu Aktivität III.1

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 5-8).

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 • Auftraggeber • Konzern-Datensteward	Techniken • Dokumentation • Datenqualitätssimulation
Checkliste ☐ Alle spezifizierten DQ-Kennzahlen sind dokumentiert, d. h. insbesondere Messverfahren und Skala für jede Kennzahl. ☐ Für jede DQ-Kennzahl ist eine Kausalkette dokumentiert, also die Auswirkung der durch sie überwachten Datendefekte.		

Tabelle 5-8: 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.

5.2.7.1 Technik *Dokumentation*

Da Unternehmen meist mit ihrer IT-Abteilung oder externen IT-Dienstleistern abgestimmte Vorlagen für Fachkonzepte nutzen, enthält Anhang G.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 Auftretenswahrscheinlichkeit der einzelnen Defekte geschätzt werden. Abbildung 5-11, Abbildung 5-12 und Abbildung 5-13 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.

	Defekt 1	Defekt 2	Defekt 3	Defekt 4	Defekt 5	Defekt 6	Defekt 7	Defekt 8	Defekt 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 5-9: Parameter (Mittelwert, Varianz) normalverteilter Defekthäufigkeiten

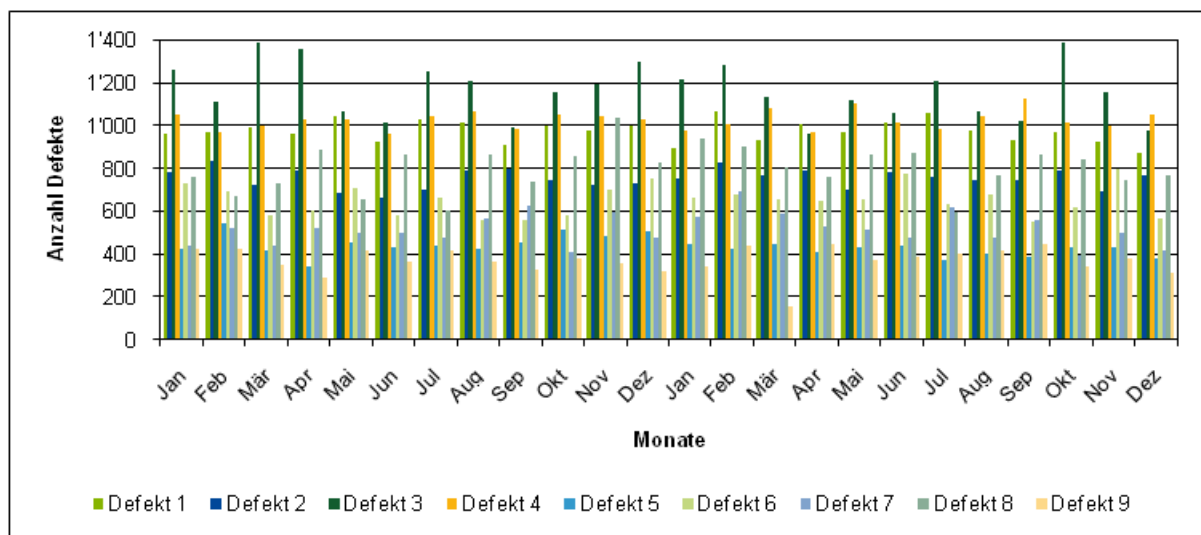


Abbildung 5-11: Simulation von Datendefekten

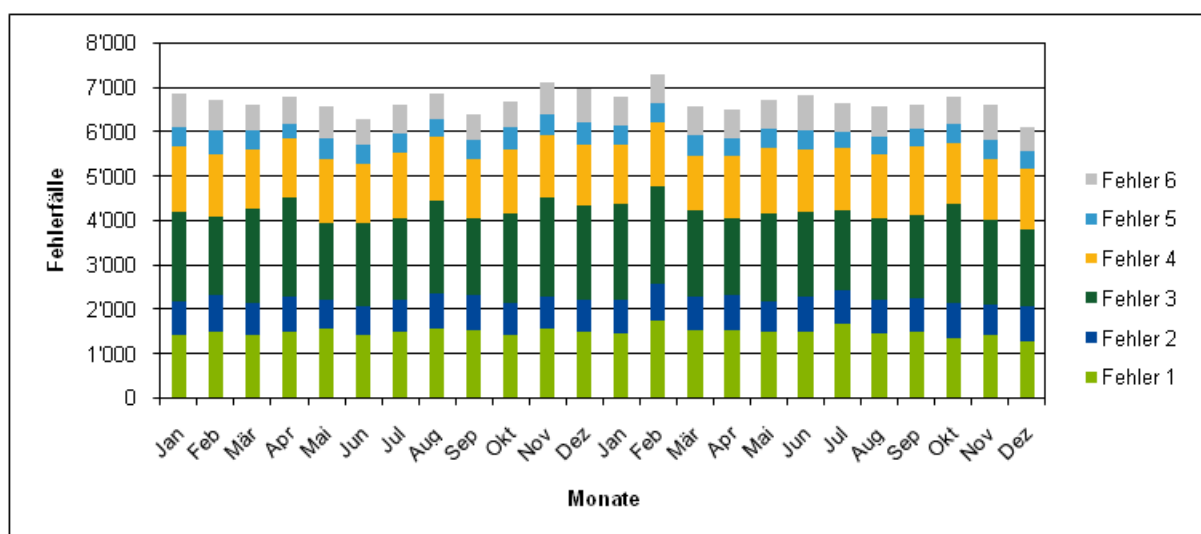


Abbildung 5-12: Durch Datendefekte verursachte Geschäftsprobleme

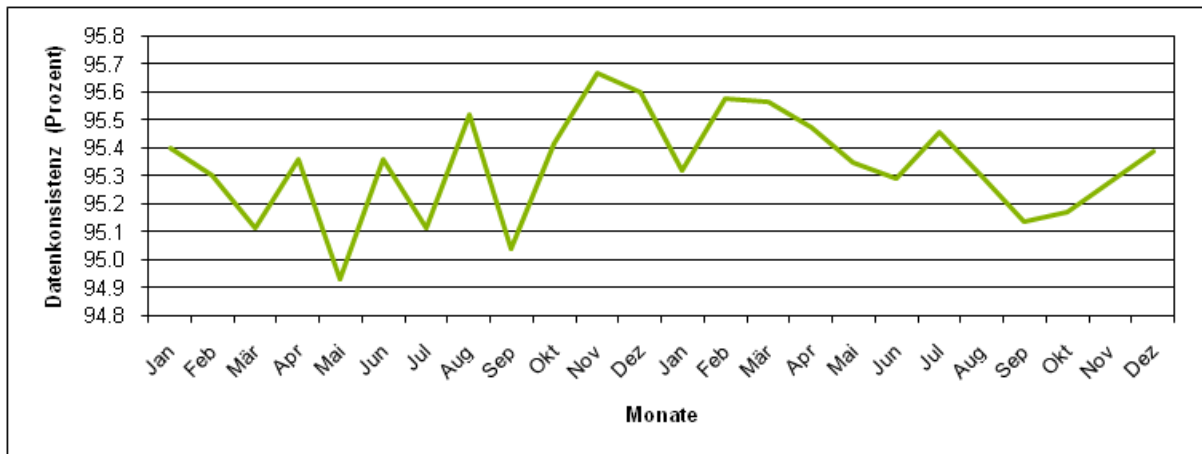


Abbildung 5-13: Überwachung von Datendefekten mit einer DQ-Kennzahl

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 H zeigt die Teilnehmer (s. Tabelle H-1) und die Leitfragen (s. Tabelle H-2) 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 6-1). Die Ausrichtung von DQ-Kennzahlen an strategischen Unternehmenszielen wurde ebenfalls als wichtiges Gestaltungsziel bestätigt.

Relevanz-Perspektive (relevance perspective)
• 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 6-1: 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 6-2). 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)
• 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 6-2: 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 6-3). 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 6-3: 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 6-4). 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 6-4: Bewertungen der Fachexperten für die Konstruktions-Perspektive

Die Fachexperten haben die erkenntnistheoretische Perspektive (engl. epistemological perspective) der Methode (vgl. Anhang H, Tabelle H-2) 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 exemplarischer Kausalketten (vgl. Anhang D.1) und Validierungsregeln (vgl. Anhang F.6) identifiziert.

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, vgl. Abschnitt 1) 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.

Literatur

[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

[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

[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

[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

[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

[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

[Bund 2009]

Bund, Leistungs- und Finanzierungsvereinbarung, Bundesrepublik Deutschland, 2009

[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

[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

[DAMA 2009]

DAMA, The DAMA Guide to the Data Management Body of Knowledge, Technics Publications, Bradley Beach, New Jersey, 2009

[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

[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

[English 1999]

English, L. P., Improving Data Warehouse and Business Information Quality, Wiley, New York et al., 1999

[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

[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

[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 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

[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

[Gutzwiller 1994]

Gutzwiller, T. A., Das CC RIM-Referenzmodell für den Entwurf von betrieblichen, transaktionsorientierten Informationssystemen, Physica, Heidelberg, Deutschland, 1994

[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

[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

[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

[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

[ISO/IEC 2007]

ISO/IEC, ISO/IEC 15939. Systems and software engineering – Measurement process, ISO/IEC 15939:2007(E), International Organization for Standardization, 2007

[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

[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

- [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
- [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
- [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
- [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
- [Mertens 2007]
Mertens, P., Integrierte Informationsverarbeitung 1: Operative Systeme in der Industrie, 16. Aufl., Gabler, 2007
- [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
- [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
- [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
- [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
- [Österle 1995]
Österle, H., Business Engineering: Prozeß- und Systementwicklung, Band 1: Entwurfstechniken, 2. Aufl., Springer, Berlin, Deutschland, 1995
- [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 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
- [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
- [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., Kopcsó, 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

[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

[Redman 1996]

Redman, T. C., *Data Quality for the Information Age*, Artech House, Boston, Massachusetts et al., 1996

[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

[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

[Scheer 1998]

Scheer, A.-W., *ARIS – Vom Geschäftsprozess zum Anwendungssystem*, 3. Aufl., Springer, Berlin et al., 1998

[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

[Schütte/Rotthowe 1998]

Schütte, R., Rotthowe, T., The Guidelines of Modeling – An Approach to Enhance the Quality in Information Models, in: Ling, T. W., Lee, S. L. (Hrsg.), Proceedings of the 17th International Conference on Conceptual Modeling, Singapur, 1998, S. 240-254

[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

[Simon 1996]

Simon, H. A., The sciences of the artificial, 3. Aufl., MIT Press, Cambridge, Massachusetts et al., 1996

[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

[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

[Strong et al. 1997]

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

[vom Brocke 2003]

vom Brocke, J., Referenzmodellierung. Gestaltung und Verteilung von Konstruktionsprozessen, Logos Verlag, Berlin, Deutschland, 2003

[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/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

[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. 2009]

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

[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

Anhang A Entitätstypen des Metamodells der Methode

Tabelle A-1 beschreibt die Entitätstypen des Metamodells der Methode (vgl. Abschnitt 5.1) und ihre Beziehungen zu Entitätstypen des Metamodells des BE-Methodenkerns (vgl. Abschnitt 2.1). Das Metamodell des BE-Methodenkerns spezifiziert HÖNING [2009, S. 252-257].

Metaentitätstyp	Beschreibung
Datenobjektattribut (kurz: Attribut)	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]. Beziehungen Aus einem Datenobjektattribut können Datenelemente instanziiert werden. Ein Datenobjektattribut ist Bestandteil eines Datenobjekttyps.
Datenobjekt	Ein Datenobjekt ist eine Menge mehrerer Datenelementen, die zusammen die Ausprägung eines entsprechenden Datenobjekttyps bilden [Schmidt 2010, S. 222]. 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.
Datenobjekttyp	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. 223]. 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.
Datenqualität	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). Beziehungen Datenqualität ist ein Bestandteil der Prozessleistung und wird von Unterstützungsprozessen (Datenmanagement, Datenqualitätsmanagement) erzeugt.
DQ-Dimension	DQ-Dimensionen repräsentieren bestimmte Aspekte (z. B. Aktualität, Glaubwürdigkeit, Vollständigkeit) des Phänomens <i>Datenqualität</i> [Wang/Strong 1996, S. 6]. Beziehungen Eine DQ-Dimension ist ein spezieller Erfolgsfaktor.

Metaentitätstyp	Beschreibung
	DQ-Dimensionen konzeptualisieren das Phänomen <i>Datenqualität</i> . DQ-Dimensionen können hierarchisch strukturiert sein.
DQ-Kennzahl	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]. Beziehungen DQ-Kennzahlen operationalisieren DQ-Dimensionen. DQ-Kennzahlen können hierarchisch strukturiert sein.
DQM-Massnahme	Eine präventive DQM-Massnahme (z. B. Spezifikation und Nutzung von DQ-Kennzahlen) reduziert das Risiko von Datendefekten (Reduktion der Auftretenswahrscheinlichkeit und/oder der Wirkung eines Datendefekts). Eine reaktive DQM-Massnahme (z. B. Entfernen von Dubletten) behebt Datendefekte nach deren Auftreten [Otto et al. 2010]. Beziehungen DQM-Massnahmen können hierarchisch strukturiert sein.
Geschäftsobjektattribut	Ein Geschäftsobjektattribut (z. B. Alter, Gewicht) kennzeichnet gemeinsame Eigenschaften von Geschäftsobjekttypen. Geschäftsobjektattribute werden durch fachliche Attributtypen und Multiplizitäten charakterisiert [Schmidt 2010, S. 223]. Beziehungen Ein Geschäftsobjektattribut ist Bestandteil eines Geschäftsobjekttyps.
Geschäftsobjekttyp	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]. 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.
Metadaten	Metadaten sind Datenelemente, die die Bedeutung sowie fachliche und technische Eigenschaften anderer Daten beschreiben [Schmidt 2010, S. 224]. Technische Metadaten (z. B. Datentypen, Feldlängen, 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 Unternehmens. Beziehungen Metadaten sind spezielle Datenelemente. Metadaten beschreiben Datenobjekttypen. Metadaten beschreiben Geschäftsobjekttypen.
Stammdatenklasse	Eine Stammdatenklasse gruppiert Stammdatenobjekte des gleichen Typs (z. B.

Metaentitätstyp	Beschreibung
	Kundendaten, Produktdaten). Beziehungen Eine Stammdatenklasse ist ein spezieller Datenobjekttyp. Stammdatenklassen können hierarchisch strukturiert sein.
Stammdatenobjekt	Ein Stammdatenobjekt ist ein spezielles Datenobjekt, dessen Datenelemente sich im Vergleich zu anderen Datenobjekten (z. B. Bewegungsdatenobjekten) seltener ändern [Schmidt 2010, S. 225]. Beziehungen Ein Stammdatenobjekt ist ein spezielles Datenobjekt.

Tabelle A-1: Definition der Metaentitätstypen der Methode

Anhang B Datenqualitätsdimensionen

B.1 Dimensionen nach Larry P. English

ENGLISH [1999, S. 87-118] unterscheidet die in Tabelle B-1 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.
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

Kategorie / Dimension	Beschreibung
	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 B-1: DQ-Dimensionen nach ENGLISH [1999]

B.2 Dimensionen nach David Loshin

Tabelle B-2 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 Bezeichnungen 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 eine Datenspezifikation an neue (nicht berücksichtigte, vgl. <i>Comprehensiveness</i>) 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 Attri-

Kategorie / Dimension	Beschreibung
	bute eines Datenobjektyps).
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 <i>NULL</i> -Werten, (z. B. ob bestimmte Einträge wie 13.12.2099 als <i>NULL</i> -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 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.

Kategorie / Dimension	Beschreibung
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 <i>NULL</i> -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. <i>Privacy, Security</i>).
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).
Unit of Costs	Bestimmt die Kosten für Datenverwaltung und -pflege (z. B. Bereitstellung von Daten, Verbesserung von Datenqualität).

Tabelle B-2: DQ-Dimensionen nach LOSHIN [2001]

B.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 B-3 erläuterten Dimension.

Dimension	Erläuterung
Value Representation Consistency	Bewertet, ob Datenelemente mit gleicher Bedeutung die gleiche Form haben (z. B. <i>St. Gallen</i> und <i>Saint Gallen</i> oder <i>USA</i> und <i>United States of America</i>).

Dimension	Erläuterung
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 <i>rot / blau / gelb</i> 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 B-3: Differenzierung der DQ-Dimension Richtigkeit nach OLSON [2003]

B.4 Dimensionen nach Thomas C. Redman

REDMAN [1996, S. 245-266] differenziert die in Tabelle B-4 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, 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

Kategorie / Dimension	Erläuterung
	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 <i>NULL</i> -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 B-4: DQ-Dimensionen nach REDMAN [1996]

B.5 Dimensionen nach Wang/Strong

WANG/STRONG [1996] haben die 15 in Tabelle B-5 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	
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 feh-

Kategorie / Dimension	Beschreibung
	len) 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 B-5: DQ-Dimensionen nach WANG/STRONG [1996]

Anhang C Kostenarten defekter Daten

C.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 C-1 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.
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. <i>Data cleansing and correction costs</i>).
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

Kostenart	Beschreibung
	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 C-1: Kostenarten defekter Daten nach ENGLISH [1999]

C.2 Kostenarten nach David Loshin

LOSHIN [2001, S. 87-93] beschreibt verschiedene Kostenarten defekter Daten, (s. Tabelle C-2), 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 werden 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

Kategorie / Auswirkung	Beschreibung
	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 C-2: Kostenarten defekter Daten nach LOSHIN [2001]

C.3 Kostenarten nach Jack E. Olson

OLSON [2003, S. 106-107] differenziert die drei in Tabelle C-3 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 C-3: Kostenarten defekter Daten nach OLSON [2003]

Anhang D Identifikation von Kausalketten

D.1 Exemplarische Kausalketten

Tabelle D-1 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	
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.

Kausalkette	Beschreibung
Energieversorgung / Kundendaten	
Kontext	Strom AG, vgl. Abschnitt 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) <i>Bankeinzug</i> 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 5.2. Strategische Ziele des Unternehmens sind die Einhaltung gesetzlicher Vorgaben, 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 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 Stücklistenposi-

Kausalkette	Beschreibung
	tionen 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 5.2. Ein strategisches Unternehmensziel ist die Gewinnung von Neukunden. Über eine Verfügbarkeitsprüfung können potentielle Neukunden prüfen, ob bestimmte Produkte an einer bestimmten Adresse verfügbar sind.
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. <i>Start of Production</i> , 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.
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 D-1: Exemplarische Kausalketten

D.2 Leitfragen zur Identifikation kritischer Datendefekte

Die Leitfragen aus Tabelle D-2 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 sind nicht ausreichend, um zwei Datenobjekte zweifelsfrei voneinander zu unterscheiden (Dubletten). Verschiedene Systeme liefern zu einem bestimmten Sachverhalt verschiedene Datenelemente (Inkonsistenz).	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 ausfüllen können (z. B. aus Zeitmangel oder Unwissenheit)? Passiert es, dass Daten mehrfach vorkommen, obwohl sie nur einmal im System sein dürften? (z. B. verschiedene Adressen eines Kunden)? Gibt es Dateninkonsistenzen zwischen verschiedenen Systemen?
Appropriate Amount of Data	Es werden so viele Datenelemente angeboten, dass ein Mitarbeiter die relevanten Datenelemente nur schwer finden kann. Es ist notwendig, Datenelemente über verschiedene Wege (z. B. Systeme, Bildschirmmasken, Dokumente) zu beziehen, um einen einzelnen Arbeitsschritt zu absolvieren. Es können nicht alle nötigen Datenelemente erfasst werden. Daher werden z. B. Freitextfelder zweck-	Gibt es Eingabeformulare oder Berichte (Reports), die so umfangreiche Daten enthalten, dass Sie Schwierigkeiten haben, relevante Daten zu finden? Gibt es Arbeitsschritte zu denen Sie Daten aus mehreren Anwendungen einholen oder verknüpfen müssen? Fehlen Ihnen in bestimmten Arbeitsschritten Eingabefelder, um für die Bearbeitung notwendige Daten einzutragen? Wenn ja: wo genau bringen Sie Ihre Daten unter? Gibt es Eingabeformulare, auf denen Sie viele Informationen eintragen müssen, die für den Geschäftsfall nicht relevant erscheinen? Wie gehen

DQ-Dimension	Möglicher Datendefekt	Leitfrage
	entfremdet.	Sie damit um?
Believability	Da in vielen Fällen Datenelemente falsch oder unzureichend sind, wird von vornherein schlechte Datenqualität unterstellt. Daher werden Datenelemente erfragt, die eigentlich vorliegen.	Gehen Sie von der Richtigkeit verfügbarer Daten aus, oder nehmen Sie von vornherein Fehler an? Werden bestimmte Daten mit anderen Daten noch einmal validiert (z. B. Rückfrage gegen Kunden oder Recherche in alternativen Datenbehältern)? Welche Daten erscheinen Ihnen unglaubwürdig? Wie erkennen Sie dies?
Completeness	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. Es fehlen Datenelemente.	Fehlen in bestimmten Bildschirmmasken oder Formularen Datenelemente, die eigentlich vorhanden sein sollten? Gibt es Datenelemente, die unverständlich oder zu stark verkürzt dargestellt sind?
Concise Representation	Die Darstellung von Datenelementen ist für die Bearbeitung einer bestimmten Aufgabe unvorteilhaft und könnte präziser sein.	Gibt es Fälle, in denen Ihnen Datenelemente angeboten werden, für deren Verständnis sie erst „lange nachdenken“ oder zusätzliche Informationsquellen verwenden müssen?
Ease of Understanding	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.	Gibt es Daten die Sie verstehen, deren Bedeutung bzw. Auswirkung Ihnen aber nicht klar ist?
Interpretability	Datenelemente sind aufgrund einer nicht bekannten Sprache oder einer nicht korrekt eingesetzten Codierung nicht verständlich.	Gibt es Daten, die in einer für Sie fremden Sprache angezeigt werden? Gibt es Daten, in denen Sonderzeichen (z.B. „%\$?#“) 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?
Representation-al 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	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)?

DQ-Dimension	Möglicher Datendefekt	Leitfrage
	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.	
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 D-2: Leitfragen zur Identifikation kritischer Datendefekte

Anhang E Anforderungen an DQ-Kennzahlen

E.1 Anforderungen der DAMA

Die DAMA [2009, S. 299] beschreibt die in Tabelle E-1 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-1: Anforderungen an DQ-Kennzahlen der DAMA [2009]

E.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-2 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, Daten-

Anforderung	Beschreibung
	objekte) 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.
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-2: Anforderungen an DQ-Kennzahlen nach KAISER ET AL. [2007]

E.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-3 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 Definiti-

Anforderung	Beschreibung
	on, 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.
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-3: Anforderungen an DQ-Kennzahlen

Anhang F Messverfahren

F.1 Bewertung von Datenqualität nach Huang et al.

HUANG ET AL. [1999, S. 87f] stellen einen Fragebogen mit 69 Fragen (s. Tabelle F-1) 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.
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.

Bewertung subjektiver DQ-Dimensionen
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.
This information is difficult to aggregate.
The amount of information is not sufficient for our needs.
This information is incorrect.
This information does not add value to our work.
This information was objectively collected.
It is difficult to interpret the coded information.
The meaning of this information is difficult to understand.
This information is not sufficiently current for our work. I C
This information is easily interpretable.
The amount of information is neither too much nor too little.
This information is accurate.
Access to this information is sufficiently restricted.
This information is presented consistently.
This information has a reputation for quality.
This information is easy to comprehend.
This information is based on facts.
This information is sufficiently complete for our needs.
This information is trustworthy.
This information is relevant to our work.
Using this information increases the value of our work.
This information is presented in a compact form.

Bewertung subjektiver DQ-Dimensionen	Bewertung subjektiver DQ-Dimensionen
This information is appropriate for our work.	formation.
The meaning of this information is easy to understand.	This information is represented in a consistent format.
This information is credible.	This information is easily obtainable.
This information covers the needs of our tasks.	This information comes from good sources.
Representation of this information is compact and concise.	This information is quickly accessible when needed.
This information adds value to our tasks.	This information has sufficient breadth and depth for tasks.
The measurement units for this information are clear.	This information presents an impartial view.
This information is objective.	This information is applicable to our work.
Information can only be accessed by people who should see it.	This information is sufficiently up to date for our work.
This information is sufficiently timely.	This information is reliable.
This information is easy to combine with other in-	

Tabelle F-1: Fragen zur Bewertung von Datenqualität nach HUANG ET AL. [1999]

F.2 Bewertung von Datenqualität nach Lee et al.

LEE ET AL. [2002, S. 143fr] haben 65 Aussagen (s. Tabelle F-2) 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. This information is of doubtful credibility. This information is trustworthy. This information is credible.
Completeness	This information includes all necessary values. This information is incomplete. This information is complete. This information is sufficiently complete for our needs. This information covers the needs of our tasks. This information has sufficient breadth and depth for our task.
Concise Representation	This information is formatted compactly. This information is presented concisely.

Dimension	Zu bewertende Aussagen
	This information is presented in a compact form. The representation of this information is compact and concise.
Consistent Representation	This information is consistently presented in the same format. This information is not presented consistently. This information is presented consistently. This information is represented in a consistent format.
Free of Error	This information is correct. This information is incorrect. This information is accurate. This information is reliable.
Interpretability	It is easy to interpret what this information means. This information is difficult to interpret. It is difficult to interpret the coded information. This information is easily interpretable. The measurement units for this information are clear.
Objectivity	This information was objectively collected. This information is based on facts. This information is objective. This information presents an impartial view.
Relevancy	This information is useful to our work. This information is relevant to our work. This information is appropriate for our work. This information is applicable to our work.
Reputation	This information has a poor reputation for quality. This information has a good reputation. This information has a reputation for quality. This information comes from good sources.
Security	This information is protected against unauthorized access. This information is not protected with adequate security. Access to this information is sufficiently restricted. This information can only be accessed by people who should see it.
Timeliness	This information is sufficiently current for our work. This information is not sufficiently timely. This information is not sufficiently current for our work. This information is sufficiently timely. This information is sufficiently up-to-date for our work.
Understandability	This information is easy to understand. The meaning of this information is difficult to understand. This information is easy to comprehend. The meaning of this information is easy to understand.
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 F-2: Aussagen zur Bewertung von Datenqualität nach LEE ET AL. [2002]

F.3 Bewertung von Datenqualität nach Price et al.

PRICE ET AL. [2008, S. 73f] identifizieren 34 Aussagen (s. Tabelle F-3) 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.	Information layout can easily be modified as 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 some 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.

Tabelle F-3: Aussagen zur Bewertung von Datenqualität nach PRICE ET AL. [2008]

F.4 Validierungsregeln nach David Loshin

Für DQ-Messungen beschreibt LOSHIN [2001, S. 171-197] die neun in Tabelle F-4 erläuterten Kategorien für Validierungsregeln (org. data quality rules).

Kategorie	Beschreibung
Null Value Rules	Prüfen, ob nur Attribute, die <i>NULL</i> -Werte enthalten dürfen, auch <i>NULL</i> -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. <i>fuzzy</i>) Datenspezifikationen oder die Nutzung vorgesehener Navigationspfade auf Websites.

Tabelle F-4: Validierungsregeln nach LOSHIN [2001]

F.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 F-5 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.
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 <i>NULL</i> , <i>BLANK</i> , <i>?</i> , <i>*</i> belegt) ist, d. h. nur angegeben wurde, um z. B. in einem Formular eine

Verfahren / Kategorie	Beschreibung
	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).
Lookup	Prüft Datenelemente gegen andere Datenelemente, die die gleichen Daten enthalten sollten (z. B. Vergleich konkreter Werte, Korrelationsanalysen).
Value Rule Analysis	

Verfahren / Kategorie	Beschreibung
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 <i>blank</i> oder <i>NULL</i>) 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 F-5: Validierungsregeln nach OLSON [2003]

F.6 Exemplarische Validierungsregeln aus partizipativen Fallstudien

Tabelle F-6 zeigt einen Auszug der Validierungsregeln, die durch Anwendung der Technik Spezifikation von Validierungsregeln (vgl. Abschnitt 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 5.2.2 und 5.2.5) sowie eine diese ausgewählten Datenobjekte iterierende Schleife (vgl. Abbildung 5-8) 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> CurrentTime() gibt die aktuelle Zeit zurück. FirstDemand(d) gibt den Zeitpunkt zurück, zu dem das Datenelement <i>d</i> zum ersten Mal gebraucht wird (z. B. Produktionsbeginn).	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> LastChange(a,e) gibt den Zeitpunkt der letzten Belegung des Attributs <i>a</i> mit dem Datenelement <i>e</i> zurück.	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> NumberOfChanges(a) gibt die Anzahl der seit der ersten Eingabe eines Datenelements vorgenommenen Änderungen des Datenelements des Attributs a zurück.	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.
<pre>if NumberOfChanges(d.BoM.Status) > 1 : return 0 return 1</pre>	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.
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> GS1(d1,d2) gibt TRUE zurück, wenn die Gewichte von zwei Datenobjekten d1 und d2 nicht mehr als 20% abweichen, sonst FALSE.	Prüft, ob zwei Datenobjekte mit gleicher GTIN die von GS1 geforderten Eigenschaften <i>keine Gewichtsabweichung von mehr als 20% bei gleicher GTIN</i> haben.
<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>	Prüft, ob zwei Datenobjekte mit gleicher GTIN die von GS1 geforderten Eigenschaften <i>gleiche Stückzahl auf allen logistischen Ebenen bei gleicher GTIN</i> haben (SH: Shrink, PU, Packungseinheit (engl. packaging unit), PA, Palette).
<pre>for each data object d2 if EditDist(d,d2) > threshold : return 0 return 1</pre> EditDist(s1,s2) gibt die minimale Editierdistanz von zwei Zeichenketten s1 und s2 zurück.	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.
Konsistenz	
<pre>if d.GTIN != ArtworkGTIN(d) : return 0 return 1</pre> ArtworkGTIN(d) gibt die in ein Produkt-Artwork eingefügte GTIN zurück (ggf. durch Anwendung eines Texterkennungsalgorithmus).	Prüft die Konsistenz der auf einem Produkt-Artwork abgebildeten GTIN und der systemseitig gespeicherten GTIN.

DQ-Dimension / Regel	Beschreibung
<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>	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.
<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>	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.
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>	Prüft die Vollständigkeit logistischer Daten für ein Produkt.
<pre> if d.FirstName1 == NULL : return 0 if d.Surname1 == NULL : return 0 if d.Salutation == 'Mr and Mr' OR 'Mr and Ms' OR 'Ms and Ms' if d.FirstName2 == NULL : return 0 if d.Surname2 == NULL : return 0 end if return 1 </pre>	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.

Tabelle F-6: Exemplarische Validierungsregeln

Anhang G Dokumentationsvorlagen

G.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 G-1: Dokumentationsvorlage Auftraggeber-Interview

Methode zur Spezifikation geschäftsorientierte Datenqualitätskennzahlen

Aktivität I.1: Betrachtungsbereich definieren Dokumentationsvorlage *Betrachtungsbereiche*

Rollen

Konzern-Datensteward	
	Name, Kontaktdaten
Prozesseigentümer	
	Name, Kontaktdaten

Geschäftsprozess oder -aktivität

Bezeichner	
	Name, ID
Beschreibung	
	Erzeugte Prozessleistung (z. B. Produkte, Service), strategische Bedeutung für das Unternehmen

Prozesskennzahl

Bezeichner	
	Name, ID
Verantwortlichkeit	
	Name, Kontaktdaten
Beschreibung	
	Geschäftsbezug, Skalenniveau, Masseinheit, Messfrequenz, Zielwert

Geschäftsproblem

Bezeichner	
	Name, ID
Beschreibung	
	Häufigkeit des Problems, verursachte Kosten oder andere Effekte, beeinträchtigte Unternehmensziele
Datendefekte als mögliche Ursache	
	Datendefekte ,die als Ursache des Geschäftsproblems in Frage kommen

Copyright © CC CDQ2 (<http://cdq.iwi.unisg.ch>), St. Gallen 2010

Universität St. Gallen, Institut für Wirtschaftsinformatik
Lehrstuhl Prof. Dr. Hubert Osterle

Abbildung G-2: Dokumentationsvorlage für Betrachtungsbereich

G.2 Vorlagen für Aktivität I.2

Methode zur Spezifikation geschäftsorientierte Datenqualitätskennzahlen	
Aktivität I.2: Daten und IT-Systeme identifizieren	Dokumentationsvorlage <i>Datenbehälter und -einschränkung</i>
Rollen	
Konzern-Datensteward	Name, Kontaktdaten
Prozesseigentümer	Name, Kontaktdaten
Dateneinschränkung	
Bezeichner	Name, ID
Geschäftsproblem	Bezeichner, s. Dokumentationsvorlage <i>Betrachtungsbereich</i>
Datenobjektyp	Name, z. B. Kundendaten, Produktdaten
Datenbehälter	Bezeichner
Verantwortlichkeit	Name, Kontaktdaten
Einschränkungen	• • • • • • • • • •
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 G-3: Dokumentationsvorlage für Datennutzung und -einschränkung

Methode zur Spezifikation geschäftorientierte Datenqualitätskennzahlen																	
Aktivität I.2: Daten und IT-Systeme identifizieren	Dokumentationsvorlage <i>Leitfragen</i>																
Rollen <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20%; padding: 5px;">Konzern-Datensteward</td> <td style="padding: 5px;"></td> </tr> <tr> <td style="padding: 5px;">Prozesseigentümer</td> <td style="padding: 5px;"></td> </tr> </table>		Konzern-Datensteward		Prozesseigentümer													
Konzern-Datensteward																	
Prozesseigentümer																	
Leitfragen <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20%; padding: 5px;">DQ-Dimension</td> <td style="padding: 5px;"></td> </tr> <tr> <td style="padding: 5px;">Leitfragen</td> <td style="padding: 5px;"> Aktualität, Änderungshäufigkeit, Genauigkeit, Konsistenz, Objektivität, Rechtzeitigkeit, Vollständigkeit, ... • • </td> </tr> <tr> <td style="padding: 5px;">DQ-Dimension</td> <td style="padding: 5px;"></td> </tr> <tr> <td style="padding: 5px;">Leitfragen</td> <td style="padding: 5px;"> Aktualität, Änderungshäufigkeit, Genauigkeit, Konsistenz, Objektivität, Rechtzeitigkeit, Vollständigkeit, ... • • </td> </tr> <tr> <td style="padding: 5px;">DQ-Dimension</td> <td style="padding: 5px;"></td> </tr> <tr> <td style="padding: 5px;">Leitfragen</td> <td style="padding: 5px;"> Aktualität, Änderungshäufigkeit, Genauigkeit, Konsistenz, Objektivität, Rechtzeitigkeit, Vollständigkeit, ... • • </td> </tr> <tr> <td style="padding: 5px;">DQ-Dimension</td> <td style="padding: 5px;"></td> </tr> <tr> <td style="padding: 5px;">Leitfragen</td> <td style="padding: 5px;"> Aktualität, Änderungshäufigkeit, Genauigkeit, Konsistenz, Objektivität, Rechtzeitigkeit, Vollständigkeit, ... • • </td> </tr> </table>		DQ-Dimension		Leitfragen	Aktualität, Änderungshäufigkeit, Genauigkeit, Konsistenz, Objektivität, Rechtzeitigkeit, Vollständigkeit, ... • •	DQ-Dimension		Leitfragen	Aktualität, Änderungshäufigkeit, Genauigkeit, Konsistenz, Objektivität, Rechtzeitigkeit, Vollständigkeit, ... • •	DQ-Dimension		Leitfragen	Aktualität, Änderungshäufigkeit, Genauigkeit, Konsistenz, Objektivität, Rechtzeitigkeit, Vollständigkeit, ... • •	DQ-Dimension		Leitfragen	Aktualität, Änderungshäufigkeit, Genauigkeit, Konsistenz, Objektivität, Rechtzeitigkeit, Vollständigkeit, ... • •
DQ-Dimension																	
Leitfragen	Aktualität, Änderungshäufigkeit, Genauigkeit, Konsistenz, Objektivität, Rechtzeitigkeit, Vollständigkeit, ... • •																
DQ-Dimension																	
Leitfragen	Aktualität, Änderungshäufigkeit, Genauigkeit, Konsistenz, Objektivität, Rechtzeitigkeit, Vollständigkeit, ... • •																
DQ-Dimension																	
Leitfragen	Aktualität, Änderungshäufigkeit, Genauigkeit, Konsistenz, Objektivität, Rechtzeitigkeit, Vollständigkeit, ... • •																
DQ-Dimension																	
Leitfragen	Aktualität, Änderungshäufigkeit, Genauigkeit, Konsistenz, Objektivität, Rechtzeitigkeit, Vollständigkeit, ... • •																
Copyright © CC CDQ2 (http://cdq.iwi.unisg.ch), St. Gallen 2010 Universität St. Gallen, Institut für Wirtschaftsinformatik Lehrstuhl Prof. Dr. Hubert Osterle																	

Abbildung G-4: Dokumentationsvorlage für Leitfragen zu DQ-Dimensionen

G.3 Vorlagen für Aktivität I.3

Methode zur Spezifikation geschäftsorientierte Datenqualitätskennzahlen	
Aktivität I.3: Kausalketten identifizieren	Dokumentationsvorlage <i>Kausalketten</i>
Rollen	
Konzern-Datensteward	Name, Kontaktdaten
Fachliche(r) Datensteward(s)	Name, Kontaktdaten
Datennutzer	Name, Kontaktdaten
Technische(r) Datensteward(s)	Name, Kontaktdaten
Prozesseigentümer	Name, Kontaktdaten
Geschäftsproblem	
Bezeichner	Name, ID
Prozessaktivität	Bezeichner
Beeinträchtigung	Welche Unternehmensziele werden wie beeinträchtigt? Häufigkeit, Auswirkung (z. B. Mehraufwand, fehlerhafte Produkte)
Datendefekt	
Bezeichner	Name, ID
Datenobjekte	Bezeichner (s. Dateneinschränkung, Aktivität I.2)
DQ-Dimension	s. Dokumentationsvorlage <i>Leitfragen</i>
Verursachung	Wie verursacht der Datendefekt das Geschäftsproblem? Immer? Nur unter bestimmten Umständen?
Copyright © CC CDQ2 (http://cdq.iwi.unisg.ch), St. Gallen 2010 Universität St. Gallen, Institut für Wirtschaftsinformatik Lehrstuhl Prof. Dr. Hubert Osterle	

Abbildung G-5: Dokumentationsvorlage für Kausalkette

G.4 Vorlagen für Aktivität II.1

Methode zur Spezifikation geschäftsorientierte Datenqualitätskennzahlen		
Aktivität II.1: Anforderungen definieren und priorisieren		Dokumentationsvorlage <i>Anforderungen</i>
Rollen		
Konzern-Datensteward		
	Name, Kontaktdaten	
Fachliche(r) Datensteward(s)		
	Name, Kontaktdaten	
Auftraggeber		
	Name, Kontaktdaten	
Technische(r) Datensteward(s)		
	Name, Kontaktdaten	
Anforderungen		
Bezeichner		Priorität
	Name, ID	
Beschreibung		
Bezeichner		Priorität
	Name, ID	
Beschreibung		
Bezeichner		Priorität
	Name, ID	
Beschreibung		
Bezeichner		Priorität
	Name, ID	
Beschreibung		
Copyright © CC CDQ2 (http://cdq.iwi.unisg.ch), St. Gallen 2010 Universität St. Gallen, Institut für Wirtschaftsinformatik Lehrstuhl Prof. Dr. Hubert Osterle		

Abbildung G-6: Dokumentationsvorlage für Anforderungen

G.5 Vorlagen für Aktivität II.2

Methode zur Spezifikation geschäftsorientierte Datenqualitätskennzahlen																	
Aktivität II.2: Datenqualitätskennzahlen spezifizieren	Dokumentationsvorlage <i>Validierungsregeln</i>																
Rollen <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20%; padding: 5px;">Konzern-Datensteward</td> <td style="padding: 5px;"></td> </tr> <tr> <td style="padding: 5px;">Name, Kontaktdaten</td> <td style="padding: 5px;"></td> </tr> <tr> <td style="padding: 5px;">Fachliche(r) Datensteward(s)</td> <td style="padding: 5px;"></td> </tr> <tr> <td style="padding: 5px;">Name, Kontaktdaten</td> <td style="padding: 5px;"></td> </tr> <tr> <td style="padding: 5px;">Technische(r) Datensteward(s)</td> <td style="padding: 5px;"></td> </tr> <tr> <td style="padding: 5px;">Name, Kontaktdaten</td> <td style="padding: 5px;"></td> </tr> </table>		Konzern-Datensteward		Name, Kontaktdaten		Fachliche(r) Datensteward(s)		Name, Kontaktdaten		Technische(r) Datensteward(s)		Name, Kontaktdaten					
Konzern-Datensteward																	
Name, Kontaktdaten																	
Fachliche(r) Datensteward(s)																	
Name, Kontaktdaten																	
Technische(r) Datensteward(s)																	
Name, Kontaktdaten																	
Validierungsregel <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20%; padding: 5px;">Bezeichner</td> <td style="padding: 5px;"></td> </tr> <tr> <td style="padding: 5px;">Name, ID</td> <td style="padding: 5px;"></td> </tr> <tr> <td style="padding: 5px;">Geschäftsproblem</td> <td style="padding: 5px;"></td> </tr> <tr> <td style="padding: 5px;">Bezeichner, s. Dokumentationsvorlage <i>Kausalketten</i></td> <td style="padding: 5px;"></td> </tr> <tr> <td style="padding: 5px;">Datendefekt</td> <td style="padding: 5px;"></td> </tr> <tr> <td style="padding: 5px;">Bezeichner, s. Dokumentationsvorlage <i>Kausalketten</i></td> <td style="padding: 5px;"></td> </tr> <tr> <td style="padding: 5px;">Attribute</td> <td style="padding: 5px;"> </td> </tr> <tr> <td style="padding: 5px;">mögliche Prüfungen</td> <td style="padding: 5px;"> </td> </tr> </table> Welche Eigenschaften der Attribute (z. B. bestimmte Datenelemente, Formate, Vergleiche von Datenelementen) sind ein Defekt?		Bezeichner		Name, ID		Geschäftsproblem		Bezeichner, s. Dokumentationsvorlage <i>Kausalketten</i>		Datendefekt		Bezeichner, s. Dokumentationsvorlage <i>Kausalketten</i>		Attribute		mögliche Prüfungen	
Bezeichner																	
Name, ID																	
Geschäftsproblem																	
Bezeichner, s. Dokumentationsvorlage <i>Kausalketten</i>																	
Datendefekt																	
Bezeichner, s. Dokumentationsvorlage <i>Kausalketten</i>																	
Attribute																	
mögliche Prüfungen																	
Copyright © CC CDQ2 (http://cdq.iwi.unisg.ch), St. Gallen 2010 Universität St. Gallen, Institut für Wirtschaftsinformatik Lehrstuhl Prof. Dr. Hubert Osterle																	

Abbildung G-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	• • •
	Bezeichner, s. Dokumentationsvorlage <i>Kausalketten</i>
Messpunkt	
	z. B. Datenbehälter, s. Dokumentationsvorlage Datenbehälter und -einschränkung
Berechnungs- vorschrift	
	Abbildung der Ergebnisse einer Messung mit Validierungsregeln auf einer Skala, Zuweisung einer Masseinheit
Validierungsregeln	• • • • •
	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 G-8: Dokumentationsvorlage für DQ-Kennzahlen

G.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		
Name, Kontaktdaten		
Fachliche(r) Datensteward(s)		
Name, Kontaktdaten		
Technische(r) Datensteward(s)		
Name, Kontaktdaten		
Anforderungsprüfung		
Bezeichner		Erfüllt? <input style="width: 10%;" type="checkbox"/>
s. Dokumentationsvorlage <i>Anforderungen</i>		
Begründung		
Begründen, falls Anforderung nicht erfüllt, die Spezifikation der DQ-Kennzahlen aber trotzdem in Ordnung ist.		
Bezeichner		Erfüllt? <input style="width: 10%;" type="checkbox"/>
s. Dokumentationsvorlage <i>Anforderungen</i>		
Begründung		
Begründen, falls Anforderung nicht erfüllt, die Spezifikation der DQ-Kennzahlen aber trotzdem in Ordnung ist.		
Bezeichner		Erfüllt? <input style="width: 10%;" type="checkbox"/>
s. Dokumentationsvorlage <i>Anforderungen</i>		
Begründung		
Begründen, falls Anforderung nicht erfüllt, die Spezifikation der DQ-Kennzahlen aber trotzdem in Ordnung ist.		
Bezeichner		Erfüllt? <input style="width: 10%;" type="checkbox"/>
s. Dokumentationsvorlage <i>Anforderungen</i>		
Begründung		
Begründen, falls Anforderung nicht erfüllt, die Spezifikation der DQ-Kennzahlen aber trotzdem in Ordnung ist.		
Copyright © CC CDQ2 (http://cdq.iwi.unisg.ch), St. Gallen 2010 Universität St. Gallen, Institut für Wirtschaftsinformatik Lehrstuhl Prof. Dr. Hubert Osterle		

Abbildung G-9: Dokumentationsvorlage für die Prüfung von Anforderungen

G.7 Vorlagen für Aktivität III.2

Methode zur Spezifikation geschäftsorientierte Datenqualitätskennzahlen	
Aktivität III.2: Spezifikation dokumentieren	Mögliche 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 <i>Kausalkette</i> und <i>geschäftsorientierte DQ-Kennzahl</i>.	
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 Osterle	

Abbildung G-10: Gliederungsvorschlag für Dokumentation

Anhang H 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 H-1 zeigt die Teilnehmer, Tabelle H-2 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 H-1: Teilnehmer des Fokusgruppeninterviews am 9. Februar 2010

Relevance perspective
- 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
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. - 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
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
• Does the method fulfill the requirements to be taken into account? • 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
• 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 H-2: Leitfragen des Fokusgruppeninterviews am 9. Februar 2010