

Zusammenfassung

Kapitel 1 führt in die Rolle der Daten in der Digitalisierung von Wirtschaft und Gesellschaft ein und beschreibt die wichtigsten Geschäftstreiber für Datenqualität. Daten stellen für Unternehmen heutzutage eine strategische Ressource dar, die bewirtschaftet werden muss – nach Zeit-, Kosten- und eben Qualitätsgesichtspunkten. Datenqualitätsmanagement ist die Unternehmensfunktion zur Verbesserung und dauerhaften Sicherung der Datenqualität im Unternehmen. Das Kapitel stellt ein Referenzmodell für das Stammdatenqualitätsmanagement vor und führt die wesentlichen Begriffe und Konzepte ein. Ein Abschnitt zur Konsortialforschung gibt eine Übersicht über die forschungsmethodische Grundlage des Kompetenzzentrums Corporate Data Quality (CC CDQ), das den projektorganisatorischen Rahmen der Inhalte dieses Buchs bildet.

Daten sind das Fundament der digitalisierten Wirtschaft. Die Durchdringung aller Lebens- und Wirtschaftsbereiche mit „digitalen Services“ liefert Daten als Treibstoff für neue Dienstleistungen, neue Kundenzugänge, neue Preismodelle, neue Ökosysteme, also letztlich für einen großen Teil der wettbewerbsentscheidenden Innovationen. Alle Anwendungen der Informationstechnik erzeugen elektronische Daten, sodass eine noch nie dagewesene Datenflut entsteht, die es zu verstehen und zu nutzen gilt.

Ericsson beispielsweise ist ein führender Anbieter von Telekommunikationsprodukten und -dienstleistungen. Das Unternehmen mit Hauptsitz in Stockholm in Schweden bietet u. a. Lösungen für das breitbandige mobile Internet an. Einerseits entstehen also Daten bei der Nutzung von Ericsson-Lösungen. Andererseits wandelt sich das Leistungsangebot von Ericsson selbst immer mehr von der Netzwerktechnologie hin zu digitalen Services. Gemeinsam mit der Container-Reederei Maersk sorgt Ericsson für Informationstransparenz über globale Lieferketten (Ericsson 2012). So kann zum Beispiel der Reifegrad von Bananen auf dem Überseetransport von Südamerika nach Europa permanent überwacht

werden und Transportgeschwindigkeiten sowie die Ladungslöschung im Zielhafen bei Bedarf angepasst werden. Das führt zu verbesserten Güterströmen am Hafen, der Optimierung des Treibstoffverbrauchs von Schiffen – und schließlich zu Kundenzufriedenheit am Obstregal im Supermarkt.

Die unternehmerischen Innovationen ebenso wie die „klassischen“ Treiber der Datenqualität, beispielsweise die Harmonisierung der Geschäftsprozesse, verlangen nach zunehmend hoher Datenqualität. Durch die digitale Vernetzung wirken sich Datenfehler und Datenmissbrauch viel gravierender aus als im Zeitalter der isolierten IT-Anwendungen. So klinken sich organisierte „Hackerbanden“ (Dahlkamp und Schmitt 2014) in den E-Mail-Verkehr zwischen Unternehmen ein, geben sich als Kreditor aus und leiten Zahlungen für Lieferungen und Leistungen auf falsche Konten um. Das fällt häufig so lange nicht auf, bis der richtige Kreditor die Zahlung anmahnt. Dann ist eine Rückabwicklung der Überweisung jedoch meist nicht mehr möglich.

Datenqualität ist kein „Hygienefaktor“, sondern braucht Management. In der digitalisierten Wirtschaft müssen Unternehmen Daten bewirtschaften wie jedes andere Wirtschaftsgut auch, nämlich nach Kosten, Zeit – und eben Qualität. Das erste Kapitel nennt aktuelle Treiber für das Datenqualitätsmanagement und stellt das Framework für Stammdatenqualitätsmanagement vor. Es fasst zudem den Stand der Wissenschaft und Praxis zum Datenqualitätsmanagement zusammen und führt in die Kernkonzepte ein.

Aufbau des Buches

Die Fallstudien in Kap. 2 zeigen, wie bedeutende Unternehmen die Datenqualität zu einer Aufgabe aller Managementebenen machen. Die Qualität der Stammdaten¹ kann nicht in einer zentralen IT-Abteilung gewährleistet, sondern muss am Ort der Datenentstehung und -verwendung, also in den Geschäftsbereichen, sichergestellt werden. Die Fallstudien dokumentieren, wie zehn Unternehmen unterschiedlicher Branchen Datenqualitätsmanagement im Unternehmensalltag verankert haben.

Kapitel 3 stellt Methoden und Werkzeuge vor, die Unternehmen beim Aufbau eines erfolgreichen Stammdatenqualitätsmanagements unterstützen. Alle Methoden wurden mehrfach in der Praxis erprobt.

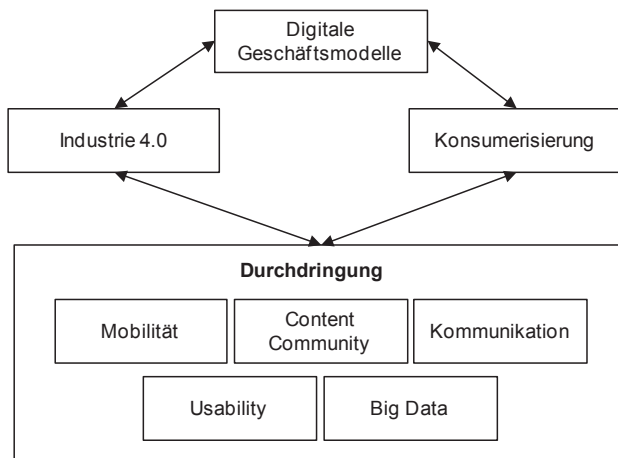
Kapitel 4 fasst die Hauptideen der beschriebenen Lösungsansätze zusammen und präsentiert eine Liste mit Sofortmaßnahmen für besseres Datenqualitätsmanagement.

1.1 Trends der Digitalisierung

Neue Formen der Informationstechnik verändern alle Bereiche von Wirtschaft und Gesellschaft, wie dies z. B. Kagermann (2014) aus der Sicht der Bundesrepublik Deutschland analysiert. Wir fassen die Entwicklung zu vier Trends zusammen (Abb. 1.1).

¹ Dieses Buch verwendet wegen des verbreiteteren Sprachgebrauchs durchgängig den Begriff „Stammdaten“. Gemeint sind damit die Konzernstammdaten, d. h. jene Untergruppe sämtlicher Stammdaten im Unternehmen, die im Rahmen eines unternehmensweiten qualitätsorientierten Datenmanagements bewirtschaftet werden sollten.

Abb. 1.1 Megatrends der Digitalisierung. (eigene Darstellung)



1.1.1 Durchdringung aller Lebens- und Wirtschaftsbereiche

Laut International Telecommunication Union nutzten im Jahre 2013 2,7 Mrd. Menschen das Internet, also knapp 40% der Weltbevölkerung (ITU 2013). Die technologischen Innovationen der letzten 15 Jahre sind für die Durchdringung des Privat- und des Geschäftsbereichs verantwortlich.

- **Mobilität:** Drahtlose Netzwerke und die Miniaturisierung von Computern und anderen Komponenten wie Sensoren und Kameras bringen die digitalen Services an den Ort der Benutzung, sei es im Privatbereich, z. B. als Aufzeichnung einer Wanderroute, oder sei es im Unternehmen, z. B. in der Ferndiagnose einer Maschine.
- **Usability:** Touch Screens und viele Detailverbesserungen wie z. B. die Anmeldung bei digitalen Services über ein Facebook-Konto oder die Sprach-Ein- und Ausgabe haben die Schwelle für die Nutzung drastisch gesenkt. Weitere Erleichterungen wie die Datenbrille (z. B. Google Glass), Gestensteuerung bis hin zur Erkennung von Augenbewegungen zeichnen sich ab.
- **Content und Community:** Unzählige Menschen produzieren einzeln (z. B. in Blogs, Tweets) oder in Gemeinschaften (z. B. Facebook) eine nur noch maschinell „überschaubare“ Menge von Inhalten in Form von Texten, Bildern, Audio und Video. YouTube zählt über eine Milliarde Videoabrufe pro Tag im Juni 2014², Facebook knapp 1,3 Mrd. aktive Benutzer im März 2014³.
- **Kommunikation:** Diese Inhalte werden synchron und asynchron, privat und geschäftlich ausgetauscht. In der Schweiz nutzen z. B. bereits 81% der Bevölkerung täglich oder mehrmals pro Woche das Internet, bei den unter 30-Jährigen sind es sogar 95%.

² Quelle: <https://www.youtube.com/yt/press/statistics.html>.

³ Quelle: <http://newsroom.fb.com/company-info/>.

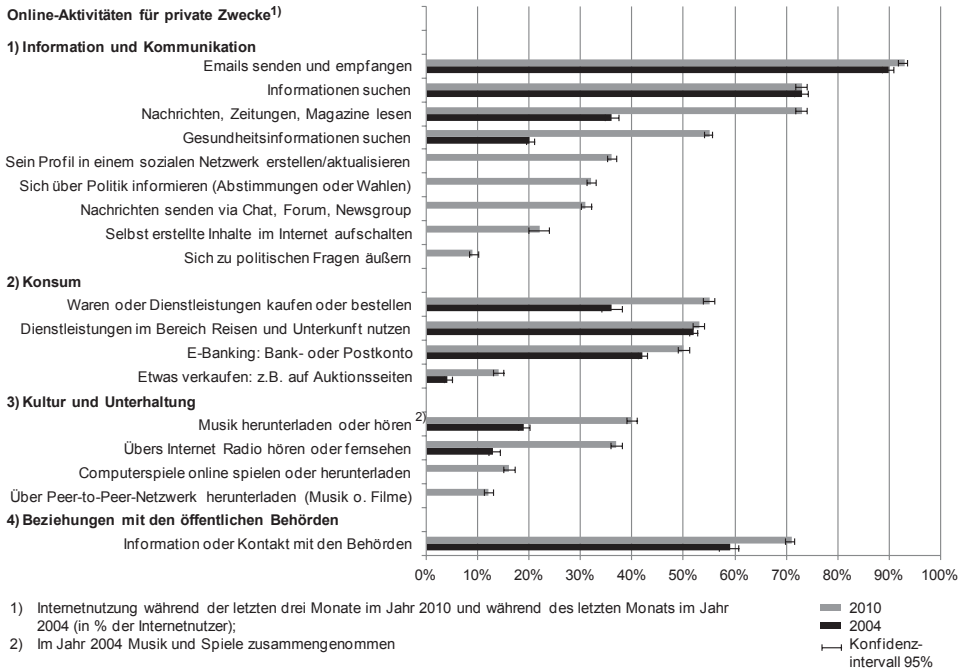


Abb. 1.2 Online-Aktivitäten für private Zwecke in den letzten drei Monaten. (Froidevaux 2012, S. 25)

Kommunikation, z. B. über E-Mail, ist dabei die häufigste Aktivität (BFS 2014). Die Videokommunikation ergänzt immer mehr die herkömmliche Sprachtelefonie und Instant Messaging – Dienste (WhatsApp) werden neben E-Mails zunehmend genutzt.

- Big Data: Bislang unbekannte Datenmengen sind das Ergebnis der Durchdringung von Wirtschaft und Gesellschaft mit digitalen Services und gleichzeitig die Grundlage für die Individualisierung von Services, insbesondere auf Basis von Lokationsinformationen (Abb. 1.2).

In Deutschland nutzte Ende 2013 fast die Hälfte der Bevölkerung (37 Mio. Menschen) ein Smartphone⁴ und ein Fünftel bis ein Viertel der deutschen Bevölkerung nutzt Social Networks über ein Smartphone. Das digitale Networking hat einen enormen Einfluss auf die Meinungsbildung der Menschen in politischen, wirtschaftlichen und privaten Angelegenheiten. Aus Sicht des Datenmanagements sind u. a. folgende Aspekte zu beachten:

- Datensicherheit: Bisher galt das Intranet im Unternehmen als Perimeter, d. h. die Linie, bis zu welcher der Schutz der Daten gesichert wurde. Diese Linie löst sich auf und Unternehmen müssen dazu übergehen, nicht Netze und Anwendungssysteme zu

⁴ Quelle: <http://de.statista.com/statistik/daten/studie/198959/umfrage/anzahl-der-smartphonennutzer-in-deutschland-seit-2010/>.

schützen, sondern die Datenobjekte ertüchtigen, selbst zu wissen, von wem sie gelesen werden dürfen und von wem nicht (O'Brien 2014).

- Datenproduktion: Klassischerweise erfassen Unternehmen Daten zentral (z. B. Kundendaten durch einen zentralen Vertriebsinnendienst). Durch die Verbreitung von Social Media und Social Networks werden jedoch Datennutzer auch immer mehr zu „Datenproduzenten“ (Strong et al. 1997). Kundendaten können durch den Kunden selbst oder von Außendienstmitarbeitern per Smartphone oder Tablet vor Ort erfasst werden. Die Mitarbeiter erwarten, dass die Daten überall verfügbar sind.
- „Streams“ statt „Records“: In Social Networks und durch Social Media erzeugen Millionen von Nutzern Datenströme. Das stellt Unternehmen vor neue Herausforderungen, weil die traditionelle Datenverarbeitung transaktionsorientiert ist, d. h. einzelne Datensätze persistent in Datenbanken geschrieben werden. Die Verarbeitung von Datenströmen aus Social Networks – wie auch aus cyberphysischen Systemen bei Industrie 4.0 – kann aber nicht mehr inkrementell sein, sondern muss kontinuierlich erfolgen (BITKOM 2014).

1.1.2 Industrie 4.0

Der Begriff „Industrie 4.0“ steht für die vierte industrielle Revolution, also die Verschmelzung der physischen mit der virtuellen Welt durch sogenannte „cyber-physische Systeme“ (Bauernhansl et al. 2014). Die Daten werden ohne Zeitverzug, ohne menschliches Zutun und viel detaillierter und exakter als zuvor erfasst. Maschinen werden internetfähig, übernehmen selbständig Aufgaben der Produktion und Datenverarbeitung, und die Daten, die bislang nur in der Fabrik verfügbar waren, sind dem gesamten Unternehmen und seinen Geschäftspartnern zugänglich (Abb. 1.3).

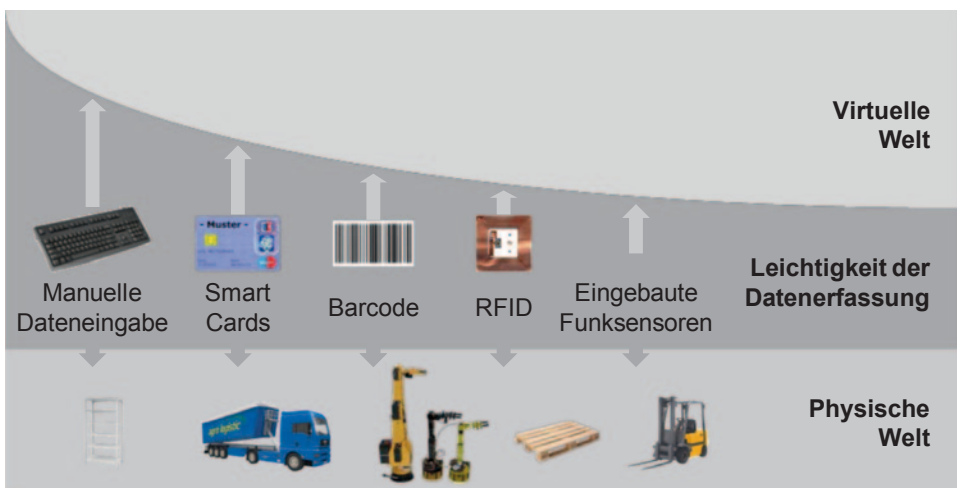


Abb. 1.3 Datenerfassung an der Schnittstelle zwischen virtueller und physischer Welt. (Fleisch 2010; Wahlster 2011, S. 5)

Abb. 1.4 Intelligenter Behälter „InBin“. (Fraunhofer IML 2015)



Industrie 4.0-Szenarien verändern den grundsätzlichen Umgang mit Daten in und zwischen Unternehmen. Das wird an drei Aspekten deutlich:

- Dezentralisierung des Datenmanagements: Die Dinge selbst werden „smart“, d. h. sie produzieren, nutzen und besitzen mehr und mehr Daten und sind zunehmend weniger auf zentrale Steuerungen angewiesen. Infolgedessen übernehmen die Dinge auch verstärkt Aufgaben der Datenverarbeitung, ohne dass es eines zentralen Rechners bedarf.
- Von der „Klasse zur Instanz“: Im Fokus der elektronischen Datenverarbeitung in der Industrie stehen traditionell „Klassen von Dingen“, also Artikel mit einer bestimmten GTIN, Produkte mit einer bestimmten Materialnummer. Industrie 4.0 bedeutet nun, dass auch jede Instanz (jedes Exemplar) einer Klasse von Produkten identifiziert werden kann, also der einzelne Hydraulikzylinder, die einzelne Flasche Hydraulikflüssigkeit (Österle und Otto 2014).
- Kontinuierliche Kopplung von Informations- und Güterfluss: Traditionell zielt die industrielle Datenverarbeitung darauf ab, Informations- und Güterfluss an bestimmten Kontrollpunkten, sogenannten i-Punkten zusammenzuführen. Ein Beispiel ist die Wareneingangsbuchung im Zentrallager bei Anlieferung von Waren. Industrie 4.0-Szenarien nutzen z. B. RFID-Technologie und ermöglichen zu jeder Zeit den Abruf von Status- und Lokationsinformationen einzelner Produkte (Österle und Otto 2014).

Ein Beispiel für eine Industrie-4.0-Anwendung ist der intelligente Behälter inBin, der von der Firma SICK⁵ gemeinsam mit dem Fraunhofer Institut für Materialfluss und Logistik (Fraunhofer IML) entwickelt wurde. Der inBin kennt seine Lokation, erfasst die Temperatur seiner Umgebung und veranlasst selbständig seine Kommissionierung (Abb. 1.4).

⁵ Für einen besseren Lesefluss verzichten wir auf die Nennung der Rechtsformen der erwähnten Unternehmen.

Voraussetzung für den Erfolg von Industrie 4.0 in einzelnen Unternehmen sowie über Supply Chains hinweg ist ein leistungsfähiges Datenmanagement, das folgende Anforderungen erfüllt:

- Beherrschung der Datenvolumina: Das Datenmanagement im Unternehmen muss in der Lage sein, die Massen an Daten zu verarbeiten und sinnvoll auszuwerten (Wrobel et al. 2014).
- Dezentrale Datenverarbeitung: Wenn Maschinen, Behälter, Frachtstücke usw. „intelligent“ werden, bedeutet dies, dass sie Datenverarbeitungsaufgaben selbständig übernehmen. Funktionen der Datenanalyse, der Datenaggregation und Datenbereitstellung finden also nicht mehr zentral in Enterprise Resource Planning (ERP)-Systemen und Data-Warehousing-Systemen statt, sondern lokal vor Ort. Ein Netzwerk von dezentralen intelligenten Geräten ergänzt die zentrale Datenverarbeitung der Unternehmen (Aggarwal et al. 2013).
- Festlegung von Datenstandards: Zeit-, Kosten- und Qualitätsvorteile durch den Einsatz cyber-physischer Systeme und der automatische Datenaustausch lassen sich nur dann realisieren, wenn sich für die Datenbeschreibungen und den Datenaustausch Standards etablieren. Diese Standards müssen mindestens innerbetrieblich, besser jedoch über ganze Supply Chains hinweg gelten (Otto et al. 2014). So entwickelt die MobiVoc-Initiative beispielsweise ein Datenvokabular für neue Mobilitätslösungen⁶.

1.1.3 Konsumerisierung

Jeder Einzelne von uns nutzt heute eine Vielzahl unterschiedlicher Konsumentenservices, die uns in verschiedenen Lebenslagen unterstützen (Österle 2014). Abbildung 1.5 zeigt zehn Lebensbereiche, in denen Menschen digitale Services nutzen, von der Navigationsunterstützung bis zum Hören von Musik, vom Preisvergleich bis zur Fernsteuerung der Beleuchtung im eigenen Haus. Der Bereich Kommunikation ist beispielhaft um zwei weitere Ebenen erweitert, um einen Eindruck von der Vielfalt der Services zu vermitteln. Eine ausführlichere, aber nie vollständige MindMap der digitalen Konsumentenservices findet man auf il.iwi.unisg.ch/appmap (Amiona 2014).

Dabei steigen die Erwartungen des Konsumenten, dass digitale Services vermehrt individuell auf ihn zugeschnitten sind. Unternehmen reagieren auf diese Konsumerisierung der Informationstechnik, indem sie ihre Geschäftsprozesse an den Bedürfnissen des Konsumenten, also dem Konsumentenprozess, ausrichten. Dieser besteht aus sämtlichen Aktivitäten, die der Einzelne zur Erfüllung verschiedenster Bedürfnisse (z. B. Einkaufen, Sport treiben, Reisen) in einer Lebenssituation zu verrichten hat.

Die Konsumerisierung führt zu einer neuen Rolle des Konsumenten im Wirtschaftsleben (Konsumentenzentrierung). Er ist nicht mehr Endpunkt bzw. Senke unidirektionaler

⁶ Siehe hierzu auch <http://www.mobivoc.org>.

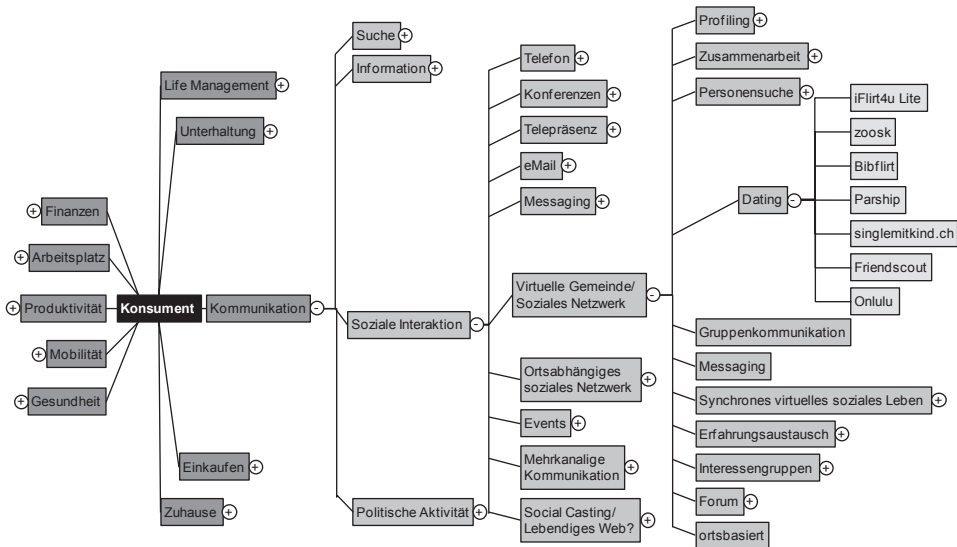


Abb. 1.5 Zehn Lebensbereiche und Beispiele für ihre digitalen Services. (Amiona 2014)

Waren- und Informationsflüsse, sondern beeinflusst über Plattformen wie Foodwatch.org die öffentliche Meinung von Produkten und Unternehmen und agiert sowohl als Verbraucher als auch als Produzent von Waren und Dienstleistungen. Beispiele sind die Stürme der Entrüstung, die über die Firma Nestlé wegen der Nutzung von Palmöl in KitKat-Schokoladenriegeln hereinbrach, und das Crowdsourcing von Programmierleistungen.

Abbildung 1.6 zeigt exemplarisch, wie sich der Fluss von Produktinformation beim Konsumgüterhersteller Beiersdorf innerhalb eines Zeitraums von fünf Jahren gewandelt hat. Von 2007 auf 2012 ist einerseits die Zahl an Akteuren im Unternehmensnetzwerk gestiegen, weil Unternehmen wie Apple und Google sowie Online-Händler wie Zalando Produktinformationen von z. B. Nivea nutzen und verteilen. Dieses erweiterte Unternehmensnetzwerk wird in Anlehnung an die Ökologie auch als „Ökosystem“ bezeichnet. Andererseits ist der Konsument hinsichtlich der „Macht über die Daten“ im Netzwerk von der Peripherie ins Zentrum gerückt, da nahezu alle Unternehmen des Netzwerks mit dem Konsumenten interagieren (Schierning 2012).

Nestlé pflegt nicht nur klassische Unternehmensdaten, sondern auch Konsumentendaten. Nestlé hat 94 Mio. Fans auf Facebook und 16 Mio. Views seines Contrex-Videos auf YouTube. Dazu kommen Daten von Onlineshops, auf denen Nespresso z. B. mehr als 50% der Kaffee-Kapseln verkauft.

Konsumentenzentrierung bedeutet für Unternehmen eine Abkehr von der traditionellen unternehmenszentrierten Sicht auf den Endkunden. Nicht mehr der Entwurf und die Verbesserung der Interaktion mit dem Konsumenten aus Sicht des Unternehmens steht im Vordergrund des Handelns („Inside-out-Ansatz“), sondern der integrale Konsumentenprozess über die Grenzen einzelner Unternehmen hinweg („Outside-in-Ansatz“).

Die Konsumerisierung führt zu neuen Anforderungen an das Datenmanagement:

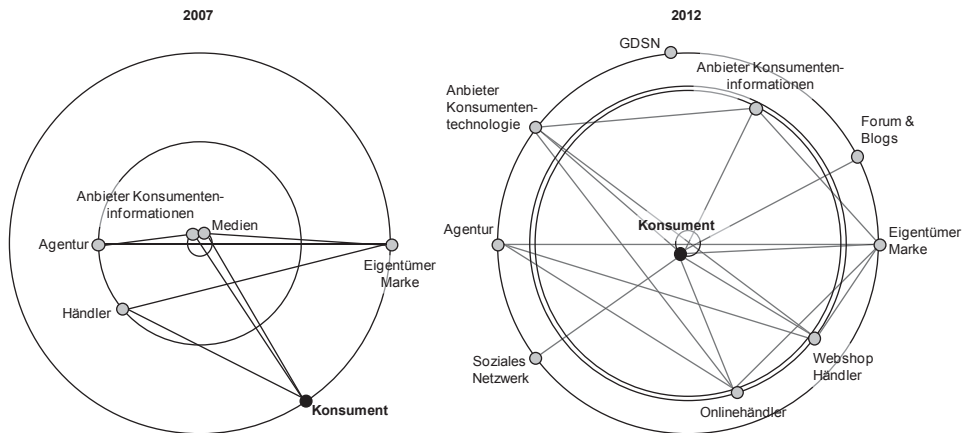


Abb. 1.6 Netzwerkanalyse der Produktinformationsflüsse bei Beiersdorf. (Schierning 2012, S. 9)

- **Daten-Ownership:** Wem gehören die Daten? Die facettenreiche Diskussion um den Datenschutz und Aussagen wie von Mark Zuckerberg von facebook, dass Datenschutz keine „soziale Norm“ mehr sei (Johnson 2010), zeigen, dass der Trend der Konsumerisierung das traditionelle Verständnis zum Eigentum und Besitz immaterieller Güter überholt hat. Sogenannte Daten-Broker sammeln persönliche Internetdaten in legalen Grauzonen (Anthes 2015). Für Unternehmen bedeutet dies, sich auf eine heterogene Rechtslage zum Datenschutz einzustellen. Gesetzgeber sind gefragt, einheitliche Rahmenbedingungen zu schaffen.
- **Datenintegration:** Die Menschen nutzen nicht mehr allein einen Kommunikationskanal, um mit einem Unternehmen in Verbindung zu treten, sondern viele verschiedene. Das Schweizer Einzelhandelsunternehmen Migros identifiziert neun verschiedene Kanäle (offline und online), über die es mit dem Konsumenten kommuniziert. Die Vielfalt reicht von Briefpost, Online Shops und E-Mail bis zu SMS. Weil der Konsument erwartet, über alle Kanäle eindeutig identifiziert zu werden und gleiche Preise und Rabatte zu den Migros-Produkten angeboten zu bekommen, muss das Unternehmen konsistente, aktuelle und vollständige Daten zu den Kunden sowie den Produkten über alle Kanäle hinweg verfügbar haben (Schemm 2012).
- **Kombination von „strukturierten“ und „unstrukturierten“ Daten:** Infolge der Konsumerisierung stellen Unternehmen nicht allein traditionelle alphanumerische Datenformate wie Beschreibungstexte, Gewichts- und Preisangaben zu Produkten bereit, sondern vermehrt Produktvideos, Marketing-Texte, Inhaltsstoffe usw. Die Unterscheidung zwischen Produktdaten, die meist in zentralen ERP- oder Product Lifecycle Management (PLM)-Systemen gespeichert sind, und multimedialen Produktinformationen, die häufig über eine Vielzahl interner Anwendungssysteme sowie externe Dienstleister (z. B. Werbeagenturen) verteilt sind, kann dann nicht mehr aufrecht erhalten werden (Österle und Otto 2014).

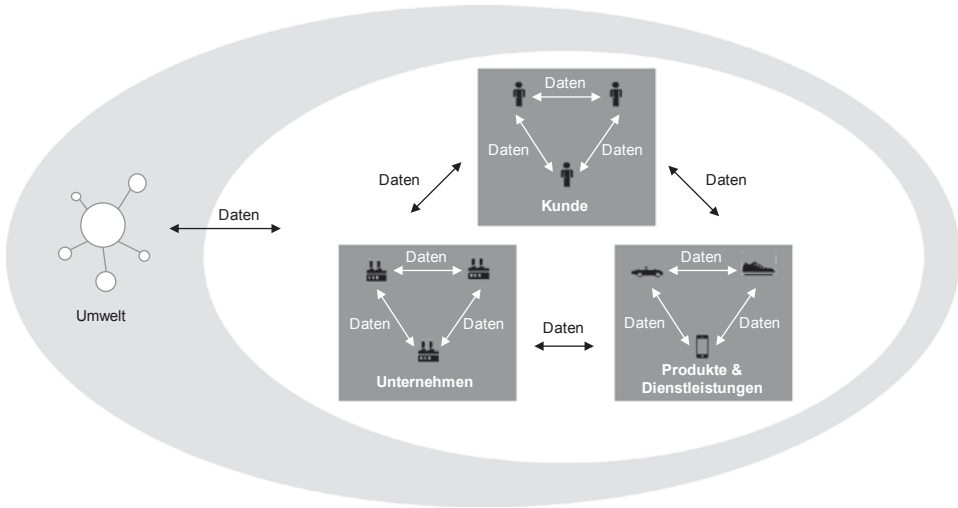


Abb. 1.7 Digitale Geschäftsmodelle. (Brenner und Herrmann 2012, S. 20)

1.1.4 Digitale Geschäftsmodelle

Die Durchdringung von Wirtschaft und Gesellschaft und damit von Industrie und Konsumenten mit digitalen Services führt zu neuartigen Geschäftsmodellen abseits klassischer Unternehmen⁷. Beispiele aus dem Konsumentenbereich sind Google, aber auch Airbnb, idealo und viele weitere Unternehmen, die eine große Zahl von Konsumenten und Geschäftskunden mit einer großen Zahl von Anbietern zusammen bringen. Diese Unternehmen nehmen eine Vermittlerrolle zwischen Leistungserstellung und -bezug verschiedener Akteure ein. Aus einer eher technischen Sicht wird vielfach auch vom „Internet der Dienste“ gesprochen. Vier Entwicklungen prägen diese Geschäftsmodelle:

- **Datenzentrierung:** Neue Geschäftsmodelle der internetbasierten Servicewirtschaft nutzen Daten als strategische Ressource (siehe Abb. 1.7). Die Deutsche Post bietet z. B. über den Dienst GEOVISTA hochauflösende Geoinformationen für den Einzelhandel, die Versicherungswirtschaft, die Immobilienwirtschaft sowie die öffentliche Verwaltung und andere Kunden an („Daten als Produkt“)⁸.
- **Industriekonvergenz:** Traditionelle Branchengrenzen verlieren an Bedeutung. Innovationstreiber beim autonomen Fahren ist Google; klassische Autobauer sind potenzielle Lizenznehmer für die Technologie. Amazon hat sich von einem Buchhändler zu einem Fulfillment-Experten gewandelt, der seine besonderen Fähigkeiten wie die skalierbare

⁷ Die deutsche Smart-Service-Welt-Initiative untersucht Prinzipien solcher Geschäftsmodelle und leitet Handlungsempfehlungen ab (Smart Service Welt Working Group 2014).

⁸ Quelle: <https://www.deutschepost.de/de/g/geovista.html>.

IT-Infrastruktur oder Logistikdienstleistungen Unternehmen aus vielfältigen Branchen und sogar Konsumenten anbietet.

- **Hybride Services:** Vielfach verbinden digitale Geschäftsmodelle digitale Dienstleistungen mit klassischen „Offline-Services“. Ein Beispiel sind Carsharing-Modelle, die das digitale Mieten und Finden von Autos inklusive Bezahlung (meist unterstützt durch SmartPhone-Apps) mit der klassischen Dienstleistung Mobilität kombinieren.
- **Konsumentenprozess:** Das Internet der Dienste richtet sich an das Individuum, also den einzelnen Konsumenten, den Patienten, den Servicetechniker oder den Shopper. Das Ziel ist die „Ende-zu-Ende“-Unterstützung in Lebenssituationen wie Einkaufen, Arbeiten, Mobilität, Therapie oder Vorsorge (Österle und Senger 2011).

1.2 Treiber der Datenqualität

Digitale Geschäftsmodelle und das Internet der Dienste basieren auf der Ressource Daten. Datenqualität ist damit für Unternehmen kein „Hygienefaktor“ mehr oder gar Selbstzweck von Stabsabteilungen, sondern ist kritisch für die Operational Excellence. Datenqualität ist definiert als ein Maß für die Eignung der Daten für bestimmte Anforderungen in Geschäftsprozessen, in denen sie verwendet werden (Otto et al. 2011). Im Folgenden wird „Datenmanagement“ stets unter besonderer Berücksichtigung des Datenqualitätsmanagements behandelt.

Zu den wichtigsten Treibern für das qualitätsorientierte Datenmanagement gehören:

- 360-Grad-Blick auf den Kunden
- Unternehmenszukäufe und -zusammenschlüsse
- Compliance
- Berichtswesen
- Operational Excellence

1.2.1 360-Grad-Blick auf den Kunden

Das Wissen über den Kunden ist der Ausgangspunkt für Marketing und Verkauf, aber auch für die Produkt- und Dienstleistungsentwicklung. Deshalb müssen Unternehmen in der Lage sein, sämtliche Informationen zu den Bedürfnissen des Kunden verfügbar zu haben. Bei Konsumenten sind das z. B. das Internet-Surf-Verhalten, die Einkäufe und die Bezugsgruppen in sozialen Netzen, bei Geschäftskunden seine Adressen, Tochterunternehmen, Kontaktdaten und Namen von Ansprechpartnern, sowie Daten zu gekauften Produkten und bestehenden Verträgen.

Das Unternehmen Bühler, ein global tätiger Hersteller von Produktionsanlagen mit Spezialisierung auf die Nahrungsmittelindustrie, stellt beispielsweise seinen Mitarbeitern

im Kundendienst und im Vertrieb einen digitalen Kundensteckbrief zur Verfügung. Dieser beantwortet Fragen wie:

- Wie hoch ist der Umsatz mit dem Kunden (und allen seinen Tochterunternehmen) im aktuellen Geschäftsjahr?
- Welche unserer Anlagen und Dienstleistungen nutzt der Kunden an welchem Standort?
- Wann laufen Wartungsverträge aus?
- Welcher Mitarbeiter hatte in den letzten drei Monaten Kontakt zu welchen Kundenmitarbeitern? Was waren Ergebnisse dieser Kontakte?
- Wie profitabel ist die Kundenbeziehung?

Der 360-Grad-Blick auf den Kunden stellt zahlreiche Anforderungen an das qualitätsorientierte Datenmanagement:

- Datenqualität: Kundendaten müssen konsistent, aktuell und vollständig über alle Funktionsbereiche (Vertrieb, Service etc.) verfügbar sein.
- Datenlebenszyklus: Es muss klar definiert sein, wie Kundendaten ins Unternehmen gelangen, wo sie erfasst und gespeichert werden, wer sie anreichert und ändert und in welche Geschäftsprozesse und Systeme sie einfließen.
- Datenschutz: Vor allem bei Konsumentendaten muss sichergestellt sein, dass die Datenschutzbestimmungen eingehalten werden, also u. a. Kundendaten gelöscht werden, wenn dies gewünscht ist.
- Data Governance: Unternehmen müssen klar festlegen, wer im Unternehmen für welche Kundendaten verantwortlich ist. Ist der Außendienstmitarbeiter für die Kundenadresse verantwortlich oder der Vertriebsinnendienst? Darf der Servicemitarbeiter den Kundenstatus in „aktiv“ ändern? Wer sammelt die E-Mails mit diesem Kunden oder dessen Facebook-Fotos?

1.2.2 Unternehmenszukäufe und -zusammenschlüsse

Unternehmenszukäufe und –zusammenschlüsse sind ein wichtiges Instrument von Unternehmensstrategien. In der chemischen Industrie hat z. B. die BASF seit 2005 u. a. die Elektronikchemikaliensparte von Merck, die Feinchemiefirma Orgamol, den Katalysatorhersteller Engelhard, die Bauchemikaliensparte von Degussa sowie den Spezialchemiekonzern Ciba übernommen. Die Zukäufe wurden in einheitliche Applikationssysteme und Geschäftsprozesse integriert.

Ein weiteres Beispiel für Unternehmensintegrationen liefert Nestlé. Das Unternehmen führt über 2000 unterschiedliche Marken, die in mehr als 440 Fabriken in fast 90 Ländern der Erde produziert und in über 190 Ländern verkauft werden⁹. Von dem Gesamtumsatz

⁹ Quelle: <http://www.nestle.com/media/facts-figures>.

Implementierungsstatus GLOBE-Projekt		
GLOBE unterstützt > 90 Nestlé Einheiten / Märkte / Geschäftsbereiche, die gemanagt werden müssen		
93% des Gesamtumsatzes	300.000 Benutzer	800 Fabriken (> 800 Fabriken (~ 380 Nestlé, ~ 430 Lohnpacker)
1.900 physische Distributionszentren	590 Vertriebsbüros	6.700 SAP Betriebscodes in Betrieb
510 Vertriebsorganisationen in Betrieb	Rechnungsgenerierung im Wert von 40 Mio. CHF / Stunde	350 Datenstandards geschaffen und andauernd in Betrieb

Abb. 1.8 Eckdaten zum zentralen System GLOBE bei Nestlé. (nach Muthreich 2013, S. 18)

in Höhe von mehr als 92 Mrd. Schweizer Franken im Jahre 2013 laufen 93% auf dem zentralen Enterprise Resource Planning (ERP)-System „GLOBE“. Abbildung 1.8 zeigt einige Eckdaten zu GLOBE.

Das GLOBE-Programm verfolgt seit seinem Start 2001 drei Ziele, nämlich die unternehmensweite Nutzung von „Best Practices“ auf Basis gemeinsamer Geschäftsprozesse, die Einführung eines standardisierten Anwendungssystems sowie die Nutzung von Daten als „Asset“. Voraussetzung dafür ist ein leistungsfähiges Datenmanagement, das insbesondere viele Unternehmenszukäufe der letzten Jahre integriert.

- Datenstandards: Für die Erfassung, Pflege und Verwendung der Stammdaten wie Kunden, Lieferanten und Materialien und Produkte müssen verbindliche Vorgaben gelten.
- Datenerfassung an der Quelle: Aufgrund der Größe und Komplexität des Unternehmens können Daten nicht zentral erfasst werden, sondern so nah wie möglich an der Datenquelle.
- Datenqualität: Die Größe des GLOBE-Systems lässt es nicht zu, Daten verunreinigt ins System zu bringen und nachträglich zu reinigen. Die Daten müssen stattdessen bei erstmaligem Erfassen richtig sein („first time right“-Prinzip).
- Datenintegration: Ein integriertes System wie GLOBE lässt keine „Datensilos“ zu, sondern alle Geschäftsbereiche, Funktionen und Märkte arbeiten mit einer integrierten Datenbasis. Die Datenintegration kann ihr Potenzial jedoch nur entfalten, wenn im Unternehmen ein Umdenken einsetzt: Weg von „My Data“ und hin zu „Our Data“.

1.2.3 Compliance

Die zunehmende Regulierungsdichte zwingt die Unternehmen, eine große und weiter steigende Zahl gesetzlicher und behördlicher Vorgaben und Vorschriften zu erfüllen. Zwei prominente Beispiele dazu sind:

- Die EU-Verordnung REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) regelt die Registrierungspflicht von Chemikalien, welche in der EU in Verkehr gebracht werden, und legt die Datenanforderungen für die Registrierung fest („no data, no market“) (European Commission 2006). Zu den Datenanforderungen gehören u. a. Angaben zur Herstellung und sicheren Verwendung von Chemikalien. Unternehmen müssen diese Anforderungen im Stammdatensatz für Materialien mitführen und für Berichtszwecke aktuell, konsistent, vollständig und in der richtigen Form vorhalten.
- Unter dem Schlagwort Solvency II vereinheitlicht die Europäische Kommission das Versicherungsaufsichtsrecht, insbesondere im Hinblick auf das sogenannte Solvenzkapital. Bestandteile der Richtlinie sind Vorgaben für das Risikomanagement sowie Berichterstattungspflichten für Versicherungsunternehmen. Diese Vorgaben resultieren in Forderungen nach einem unternehmensweit einheitlichen Management von Markt-, Kerngeschäfts- und Finanzdaten (Salchegger und Dewor 2008).

Das Pharmaunternehmen Novartis muss z. B. aufgrund behördlicher und gesetzlicher Auflagen Daten zu klinischen Studien und zu Wirkstoffen in Produkten vollständig, aktuell und korrekt bereitstellen können. Als Voraussetzung dafür schafft das Unternehmen einen durchgängigen, unternehmensweiten „Regulatory Submission“-Prozess. Das Datenmanagement spielt dabei eine besondere Rolle:

- Datenkonsistenz: Nicht nur die Daten selbst, sondern auch die Metadaten (Definitionen, Wertelisten usw.) müssen über Systeme, Geschäftsprozesse und Funktionen hinweg konsistent sein.
- Datenlebenszyklus: Der gesamte Lebenszyklus der Daten von ihrer Entstehung bis zur Archivierung und zum Löschen muss definiert sein.
- Data Governance: Es muss definiert sein, wer im Unternehmen für welche Daten welche Rechte zur Definition und zur Nutzung hat.

1.2.4 Berichtswesen

Unternehmen geben zwischen 1 und 5% ihres Umsatzes für die Anschaffung und den Betrieb leistungsfähiger Unternehmenssoftware (z. B. SAP Business Suite) aus (Reynolds 2010; Equey et al. 2008), können aber oftmals grundlegende Fragen nicht beantworten. Beispiele für diese Fragen sind:

- Aus wie vielen Produkten besteht unser Sortiment?
- Wie hoch ist das Beschaffungsvolumen mit den größten zehn Lieferanten?
- Welchen Umsatz haben wir im vergangenen Geschäftsjahr mit unserem größten Kunden gemacht?

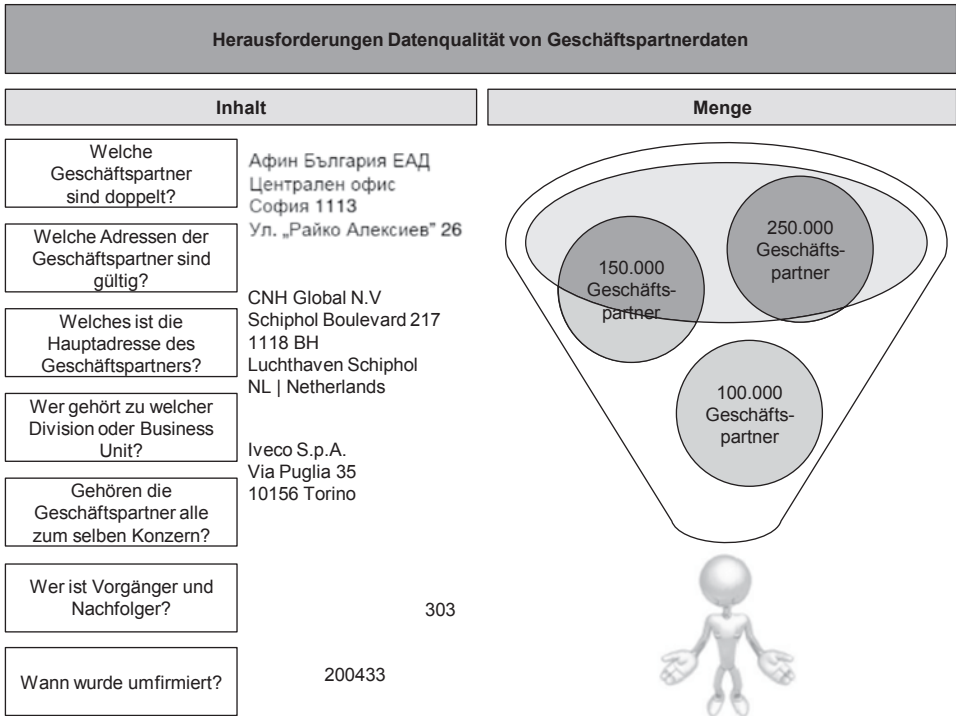


Abb. 1.9 Datenqualitätsherausforderungen beim Kundenumsatz-Reporting bei ZF Friedrichshafen. (Möller 2012, S. 24)

Der Grund dafür ist nicht etwa das Unvermögen, die Systeme zu nutzen, oder ein niedriger Reifegrad des IT-Betriebs, sondern das Fehlen einer sogenannten „Single Source of the Truth“. Große Unternehmen bestehen aus einer Vielzahl von Sparten, Standorten und Geschäftsprozessen, in denen sich über den Lauf der Zeit jeweils ein eigenes Bild der Realität (Kunden, Materialien, Lieferanten usw.) entwickelt hat. Wenn dann z. B. im Rahmen der Lieferantenentwicklung die Beschaffungsvolumina aller Standorte, Sparten usw. bei einem Lieferanten und all seinen Tochterunternehmen ermittelt werden sollen, passen diese unterschiedlichen Abbildungen der Realität nicht zusammen.

Exemplarisch zeigt Abb. 1.9 einige Datenqualitätsherausforderungen am Beispiel der ZF Friedrichshafen AG, die für ein aussagekräftiges Geschäftspartner-Reporting bewältigt werden müssen.

Bestandteile des Datenmanagements für ein vertrauenswürdigen Berichtswesen sind:

- Datenmodell: Voraussetzung für die Single Source of Truth ist ein straffes Management der Kunden-, Produkt- und Lieferantendaten, sodass alle Objekte eindeutig identifizierbar sind, die unternehmensweit wichtig sind.
- Datenqualität: Die Datennutzung im Reporting gibt die Anforderungen an die Datenqualität vor, also welches Maß an Aktualität, Vollständigkeit und Konsistenz bestimmte Attribute der Kunden-, Produkt- und Lieferantendaten erfüllen müssen.

- Datenarchitektur: Die Datenarchitektur definiert einerseits das Konzerndatenmodell, legt aber andererseits auch fest, welche Systeme Single Source of the Truth für welche Datenobjekte bzw. -attribute sind und in welche anderen Systeme die Daten von dort verteilt werden.

1.2.5 Operational Excellence

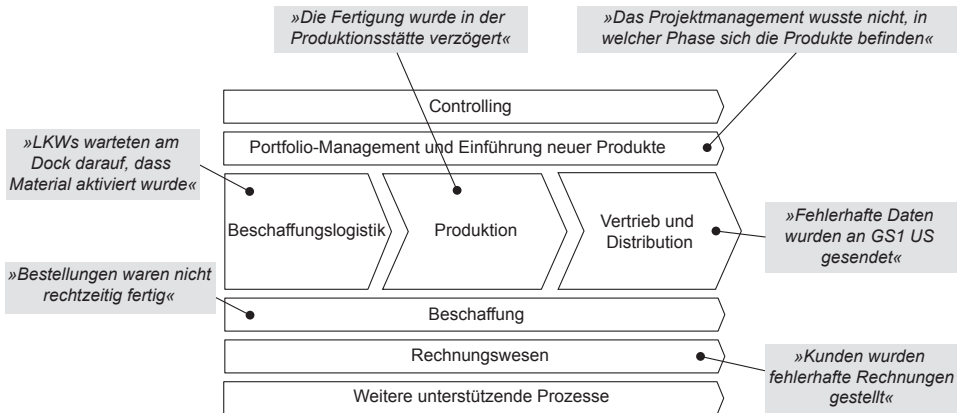


Abb. 1.10 Geschäftsprozessprobleme durch schlechte Datenqualität bei Johnson & Johnson. (Otto 2014, S. 20)

Mit der Standardisierung und Automatisierung von Geschäftsprozessen nutzen Unternehmen Skaleneffekte und verringern gleichzeitig ihre Komplexität. Voraussetzung dafür ist ein einheitliches Verständnis über die Daten im Unternehmen, welche in allen Geschäftsbereichen genutzt werden. Denn die Standardisierung der Geschäftsprozesse ist nicht möglich, wenn z. B. Materialstammdaten in Teilprozessen oder Regionen unterschiedlich definiert sind und unterschiedlich erzeugt und verwendet werden.

2007 litt die Konsumgütersparte von Johnson & Johnson in den USA unter vielen Problemen mit der Datenqualität in Geschäftsprozessen (siehe Abb. 1.10 und Kap. 2.7).

Weniger als 30% der logistischen Daten zu Artikeln, d. h. Angaben zu den Abmessungen und zum Gewicht der Artikel, befanden sich innerhalb der erlaubten Fehlertoleranz von 5%. Anders ausgedrückt: Mehr als 70% der logistischen Daten waren falsch. Johnson & Johnson erneuerte sein Stammdatenmanagement und erreichte 2013 einen Six-Sigma-Level hinsichtlich seiner Datenqualität¹⁰. Voraussetzungen dafür waren:

¹⁰ Six Sigma ist ein Qualitätsmanagementansatz, der als Leistungsziel nur 3,4 Fehler pro eine Million Instanzen vorsieht (Shah et al. 2008). Nach Wang et al. (1998) können Qualitätsmanagementansätze für physische Güter auch auf immaterielle Güter wie Daten übertragen werden.

- Data Governance: Eine zentrale Stelle im Unternehmen legt fest, wie die Daten definiert sind, wie sie angelegt, verwendet und gelöscht werden und welche Qualität sie haben müssen, damit die Geschäftsprozesse reibungslos funktionieren.
- Datenqualitätsmessung: Monatlich wird die Datenqualität der wichtigsten Attribute gemessen. Ein Datensatz, der nicht sämtliche der mehreren hundert Geschäftsregeln erfüllt, gilt als defekt.
- Workflow-gestützte Anlage: Die Erfassung und Änderung der Daten ist klar geregelt und durchgängig durch ein Workflow-Managementsystem gestützt.

1.2.6 Datensicherheit und Privatheit

Privatpersonen geben im Internet vermehrt freiwillig private Daten preis, um Angebote wie soziale Netzwerke zu nutzen. Während in den USA der Datenschutz kaum gesetzlich geregelt ist, wird der Schutz personenbezogener Daten in der Europäischen Union als Grundrecht definiert (EU 2010). Daten dürfen nur mit Einwilligung der betroffenen Personen oder auf Basis einer gesetzlichen Grundlage verarbeitet werden. Für Unternehmen haben diese Vorgaben Auswirkungen auf Werbemaßnahmen oder Analysen des Kundenverhaltens. Allerdings ist die Rechtslage nicht immer eindeutig angesichts der Vernetzung durch das Internet sowie infolge der Auslagerung von IT-Aktivitäten in Länder, in denen andere Datenschutzbestimmungen gelten.

Unternehmen haben im Sinne der Informationssicherheit auch die Aufgabe, die von ihnen verwalteten personenbezogenen Daten gegen den unberechtigten Zugriff Dritter zu schützen. Dabei schaden Datenlecks Internet-Unternehmen besonders intensiv. Deshalb versuchte im Dezember 2013 ein Telekommunikationsunternehmen gerichtlich die Publikation eines Artikels zu verhindern, in dem öffentlich wurde, dass Angaben zu Bankkonten von 7500 Kunden sowie zu 5,6 Mio. E-Mail-Abonnenten entwendet worden sind. Durch ein Gerichtsurteil ist die superprovisorische Verfügung inzwischen aufgehoben und die Veröffentlichung hat stattgefunden (Schmid 2014).

1.3 Herausforderungen und Anforderungen des Datenqualitätsmanagements

Unternehmen stehen vor der Aufgabe, dass sie einerseits die gesellschaftlichen Trends der Digitalisierung (Kap. 1.1) nutzbar machen und gleichzeitig Antworten auf die großen Treiber für Datenqualität (Kap. 1.2) finden müssen. Daraus ergibt sich eine Anzahl von konkreten Herausforderungen und Leistungsanforderungen, die Unternehmen beim qualitätsorientierten Management von Stammdaten (kurz Datenqualitätsmanagement DQM) berücksichtigen müssen.

Tab. 1.1 Die zehn großen Datenmanagement-Herausforderungen^a

Rang	Herausforderung	Punktwert
1	Datenqualität	52
2	Transparenz über Datennutzung	36
3	Redundante Datenpflege	33
4	Manuelle Datenpflege	31
5	Limitationen zentraler Datenarchitekturen (Inflexibilität, Bürokratie etc.)	25
6	Semantische Integration	18
	Trennung zwischen „strukturierten“ und „unstrukturierten“ Daten	18
8	Datenschutz	14
9	Trennung zwischen OLAP (Online Analytical Processing) und OLTP (Online Transactional Processing)	5
10	Management von „Klassen“ von Dingen, keine Instanzen	1

^a Die Teilnehmer der Befragung waren aufgefordert, aus einer Liste von zehn Einträgen die fünf größten Herausforderungen im Datenmanagement zu nennen, wobei der Wert „1“ die größte und „5“ die fünftgrößte Herausforderung darstellte. Bei 17 Teilnehmern ergeben sich ein Maximalwert von 85 und ein Minimalwert von 0 Punkten pro Eintrag.

1.3.1 Herausforderungen im Umgang mit Daten

Das Kompetenzzentrum Corporate Data Quality (CC CDQ) an der Universität St. Gallen und dem Fraunhofer Institut für Materialfluss und Logistik in Dortmund greift die größten Herausforderungen im Datenmanagement auf und entwickelt dafür neue Lösungsansätze. Im Rahmen des CC CDQ wurden im April 2013 Datenmanager aus unterschiedlichen Industrien zu den größten Herausforderungen für das Datenmanagement befragt (Österle und Otto 2014). Tabelle 1.1 zeigt als Ergebnis dieser Fokusgruppe die Rangfolge der Herausforderungen.

Die Verbesserung und die Sicherung der **Datenqualität** gelten den Teilnehmern der Fokusgruppe mit Abstand als die größte Herausforderung. Datenqualität ist ein Maß dafür, in welchem Umfang die Daten geeignet sind, die Anforderungen der Geschäftsprozesse zu erfüllen, in denen sie verwendet werden (Otto et al. 2011). Datenqualität lässt sich in verschiedenen Datenqualitätsdimensionen messen, wie z. B. Konsistenz, Aktualität und Vollständigkeit.

Als zweitgrößte Herausforderung sehen die Teilnehmer der Fokusgruppe die **Transparenz über die Datennutzung**. Insbesondere in großen Unternehmen mit komplexen Anwendungssystemlandschaften ist oftmals unklar, wo und wie Daten ins Unternehmen gelangen, in welchem System sie federführend gespeichert sind und was nach ihrer Verteilung in lokale Anwendungssysteme mit ihnen geschieht. White und Radcliffe (2010) verwenden in diesem Zusammenhang den Begriff der mangelnden „Downstream Visibility“ von Daten.

Redundante Datenpflege gilt als drittgrößte Herausforderung. Ein Beispiel ist die Erfassung und Pflege von Lieferantenstammdaten in unterschiedlichen Geschäftsbereichen

desselben Unternehmens oder bei verschiedenen Unternehmen mit überlappender Lieferantenbasis. Typischerweise erfasst und pflegt jedes Unternehmen die Daten separat, obwohl alle die gleichen Daten benötigen. An wie vielen Orten werden z. B. die Adressdaten von IBM gepflegt? Wie oft muss ein Konsument seine Adresse und andere Informationen eingeben? Wäre es nicht von Vorteil, wenn Portale wie Facebook versuchten, dem Konsumenten redundante Identifikationen abzunehmen und seine Identität allen Interessierten und Berechtigten verfügbar zu machen?

Eine vierte Herausforderung ist die **manuelle Datenpflege**, die bei Medienbrüchen in der Datenverarbeitung auftritt (Fleisch und Österle 2004). Ein Beispiel für einen Medienbruch ist das Abtippen oder Einscannen von Kundenstammdaten von einer Visitenkarte in ein Customer Relationship Management (CRM)-System. Manuelle Datenpflege ist anfälliger für Fehler und gefährdet damit die Datenqualität.

Als fünfte große Herausforderung gelten die **Limitationen zentraler Datenarchitekturen**. Denn zukünftig werden immer mehr Daten von externen Quellen beschafft und zur Zeit des Bedarfs im Geschäftsprozess zur Verfügung gestellt. Ein Beispiel sind Angaben zum CO₂-Ausstoß bei der Produktion und Distribution von Konsumgütern. Konsumgüterhersteller, die zu derartigen Angaben z. B. in Frankreich verpflichtet sind (AFNOR 2009), werden diese Daten nicht in zentralen (ERP)-Systemen führend verwalten, sondern auf autorisierte Datenbanken von Drittanbietern zugreifen.

Die **semantische Integration** von Daten ist die sechswichtigste Herausforderung im Datenmanagement. In der Datenintegration ist Semantik definiert als die Interpretation von Daten in einem bestimmten Anwendungsfall (Ziegler und Dittrich 2007). Ein Beispiel ist der Begriff „Kunde“, der in der Buchhaltung eines Unternehmens als „aktiver Kunde“ verstanden wird und im Vertrieb als „potentieller Kunde“.

Ebenfalls auf dem sechsten Rang ist die **Trennung zwischen „strukturierten“ und „unstrukturierten“ Daten** genannt. Als strukturierte Daten werden alphanumerische Daten bezeichnet, die oftmals gemäß einem relationalen Datenbankschema organisiert sind. Als unstrukturiert gelten Texte, Audios, Videos, Bilder, Tweets und Zeichnungen. Die Trennung zwischen diesen beiden Datenarten stellt Unternehmen vor Probleme, wenn z. B. im Berichtswesen neben Umsätzen auch Daten aus Social-Networking-Plattformen oder Verbraucherportalen analysiert werden sollen (Baars und Kemper 2008).

Platz 8 der größten Herausforderungen im Datenmanagement nimmt der **Datenschutz** ein. Sony wurde z. B. 2011 Opfer eines Hacker-Angriffs auf sein PlayStation-Netzwerk, bei dem auch Daten von Nutzern gestohlen wurden. Wurde das Unternehmen zu Beginn des Jahres 2013 in Großbritannien zu Strafzahlungen in Höhe von 250.000 GBP verurteilt, weil der Vorfall nach Ansicht des Information Commissioner's Office (ICO) hätte „verhindert werden können“ (BBC 2013), so wog doch der Reputationsverlust viel schwerer.

Die neuntgrößte Herausforderung im Datenmanagement ist die **Trennung zwischen „Online Analytical Processing“ (OLAP) und „Online Transactional Processing (OLTP)“**. Häufig werden Daten in OLTP-Systemen wie ERP-Systemen erfasst, gepflegt und anschließend extrahiert, um dann nach Transformations- und Bereinigungsverfahren in OLAP-Systeme wie Data Warehouses und Business Intelligence-Anwendungen importiert

zu werden. Die Herausforderung besteht darin, dass die Daten somit redundant gehalten werden, was zusätzliche Aufwände hervorruft und die Datenqualität gefährdet.

Auf Platz 10 rangiert die Herausforderung, in Zukunft **nicht** allein **Klassen von Entitäten** zu bewirtschaften, sondern **Instanzen**. Durch RFID-Technologien kann ein Spezialmaschinenbauer jedes einzelne Ersatzteil identifizieren. Dieser Ansatz unterscheidet sich von der klassischen Datenverarbeitung, bei der ein Stammdatensatz inkl. einer Teilenummer die Teileklasse beschreibt und Bestandsdaten den jeweiligen Lagerbestand. Das Datenmanagement muss mit dieser Zunahme der Datensätze umgehen können.

1.3.2 Anforderungen an das Datenqualitätsmanagement

Die Beispiele in den Kap. 1.1 und 1.2 haben gezeigt, dass sich diese Anforderungen durch die Digitalisierung noch verschärfen. Tabelle 1.2 fasst die wichtigsten Anforderungen an ein erfolgreiches Datenqualitätsmanagement zusammen, die von den „Top 8“ der Herausforderungen abgeleitet werden können.

Tab. 1.2 Anforderungen an das Datenqualitätsmanagement

Rang	Herausforderung	Anforderung
1	Datenqualität	Definition und Messbarkeit: Das moderne Datenqualitätsmanagement muss festlegen, welche Datenqualität die Geschäftsprozesse für einen reibungslosen Ablauf benötigen. Dabei gilt: Nur was gemessen wird, lässt sich managen. Die Datenqualität muss deshalb kontinuierlich gemessen werden und bei Abweichung vom Sollwert muss das Datenqualitätsmanagement Maßnahmen zur Erhöhung der Datenqualität einleiten.
2	Transparenz über Datennutzung	Transparenz und Verantwortlichkeit: Der Lebenszyklus der Daten, beginnend mit ihrer Entstehung im Unternehmen und erstmaligen Erfassung in einem Informationssystem über ihre Nutzung bis zur Archivierung und Löschung, muss bekannt und gemäß den Anforderungen der Geschäftsprozesse definiert sein. Das Datenqualitätsmanagement muss diesen Lebenszyklus steuern und überwachen. Unternehmen müssen die Definition von Daten sowie ihre Nutzung klar regeln. Dafür müssen Verantwortlichkeiten im Unternehmen geschaffen und zugeordnet sein. Beispielsweise definiert beim unternehmensweiten Umsatzreporting eine zentrale Stelle in der Finanzbuchhaltung oder im Vertrieb die Kundenstammdaten, damit in allen Geschäftsbereichen derselbe Kunde auch eindeutig identifiziert und gleich verwendet wird.
3	Redundante Datenpflege	Prävention: Datenqualitätsmanagement darf nicht erst dann beginnen, wenn die Daten bereits defekt sind, sondern muss vorbeugend wirken – wie bei anderen „Assets“ im Unternehmen auch (z. B. vorbeugende Wartung von Produktionsanlagen, Maßnahmen zur Gesundheitsprävention bei Mitarbeitern).

Tab. 1.2 (Fortsetzung)

Rang	Herausforderung	Anforderung
4	Manuelle Datenpflege	Automatisierung: Die Datenvolumina, ihre Vielfalt und ihre Änderungsrate steigen. Um der daraus resultierenden Komplexität Herr zu werden, müssen Unternehmen Datenverarbeitungsaufgaben (z. B. die Anlage von Daten, ihre Qualitätsmessung, ihre Änderung und Bereitstellung) so weit wie möglich automatisieren, z. B. über Workflows oder Geschäftsregeln.
5	Limitationen zentraler Datenarchitekturen	Flexibilisierung und Verteilung: Datenarchitekturen definieren ein unternehmensweit einheitliches Modell der Konzerndaten (das Konzerndatenmodell) und bestimmen außerdem die Datenverteilungs- und Datenhaltungsarchitektur. Sie haben traditionell den Nachteil, dass sie nur mit hohem bürokratischem Aufwand erstellt werden können und selten aktuell gehalten werden. Moderne Datenarchitekturen müssen hinreichend flexibel an neue Anforderungen angepasst werden können und sowohl klassische unternehmensinterne als auch externe Datenobjekte enthalten. Die Herausforderung besteht darin, diese Flexibilität zu ermöglichen, aber gleichzeitig für die Kerngeschäftsobjekte weiterhin unternehmensweit maßgebend zu sein.
6	Semantische Integration	Einheitlichkeit: Konzerndaten als unternehmensweite Stammdaten müssen eindeutig identifiziert und einheitlich verwendet werden. Das Datenqualitätsmanagement muss dafür die Konzerndaten integrieren. Grundsätzlich stehen zwei Architekturvarianten dafür zur Verfügung: Entweder werden die Daten in einem System zusammengeführt oder die Daten verbleiben in verschiedenen Systemen und die Systeme werden über Schnittstellen und Datenaustausch miteinander verbunden.
	Trennung „strukturierte“ und „unstrukturierte“ Daten	Heterogenität der Datentypen: Im Kontext von „Big Data“ wird häufig von „strukturierten“ und „unstrukturierten Daten“ gesprochen, um die wachsende Heterogenität der vorkommenden Datenarten zu beschreiben, mit denen Unternehmen umgehen müssen. Damit ist gemeint, dass neben Daten, die in ERP-Systemen in relationalen Datenbanken abgelegt sind, zunehmend auch Daten wie Videos und Bilder sowie unkonventionelle externe Daten, z. B. über Internetnutzung oder Social Media-Kanäle, verwendet werden. Diese können wertvolle Erkenntnisse über den Markt und Konsumentenvorlieben liefern. Solche „unstrukturierten“ Daten brauchen aber neue Werkzeuge zur Datenanalyse und werden üblicherweise nicht in relationalen Datenbanken gespeichert.
	Datenschutz	Gerade multinationale Unternehmen sind mit unterschiedlichen Datenschutzvorgaben konfrontiert. Das Datenqualitätsmanagement muss sicherstellen, dass diese Regeln eingehalten werden. Problemquellen hierfür sind, dass Richtlinien oft nicht transparent sind, sie sich häufig ändern (was ebenfalls unbekannt ist) und niemand genau weiß, inwieweit sie in den Systemen umgesetzt sind. Daher wird Datenschutz von vielen Unternehmen eher als Behinderung angesehen denn als Opportunität.

1.4 Framework für Stammdatenqualitätsmanagement

Die oben genannten Anforderungen müssen in der Praxis an den speziellen Bedürfnissen und Möglichkeiten jedes Unternehmens orientiert werden, damit ein erfolgreiches unternehmensweites Datenqualitätsmanagement entstehen kann. Denn Datenqualität heißt nicht Qualität um jeden Preis, sondern Qualität gemäß den Anforderungen der Unternehmensstrategie, der Geschäftsprozesse und der Strukturorganisation sowie des Informationssystems.

1.4.1 Framework-Überblick

Das Framework für (Konzern-)Stammdatenqualität bietet eine Lösung für diese Gestaltungsaufgabe, indem es den Ansatz des Business Engineering auf das unternehmensweite Datenqualitätsmanagement überträgt (siehe Abb. 1.11). Generell ist Business Engineering die methodenorientierte und modellbasierte Konstruktionslehre für Unternehmen des Informationszeitalters (Österle und Winter 2003). Gestaltet werden Artefakte auf den drei Ebenen „Strategie“, „Organisation“ und „Systeme“ in sechs Gestaltungsbereichen (Otto 2011b; Otto et al. 2011). Jeder Gestaltungsbereich hat eigene Ergebnistypen (Dokumente).

1.4.2 Strategieebene

Die „Datenqualitätsmanagementstrategie“ richtet das Datenqualitätsmanagement an den Unternehmenszielen aus (siehe Tab. 1.3).

Ein Beispiel für den Zusammenhang zwischen Datenqualitätsmanagement und den Zielen des Unternehmens findet sich bei der DB Netz AG, die für die Eisenbahninfrastruktur in Deutschland zuständig ist. Zur Eisenbahninfrastruktur gehören das Gleisnetz, Tunnels, Brücken, Bahnhöfe etc. Eine Leistungs- und Finanzierungsvereinbarung regelt die Mittelzuwendung des Bundes an die DB Netz AG im Sinne einer Bezuschussung für Instandhaltungsarbeiten an der Eisenbahninfrastruktur. Die Höhe des jährlichen Zuschusses hängt – in gewissen Grenzen – direkt von der Qualität des Infrastrukturkatasters ab, in welchem u. a. Anzahl, Wartungszustand und gewisse Leistungsparameter (zum Beispiel zulässige Geschwindigkeiten) sämtlicher Infrastrukturanlagen erfasst werden. Eine hohe Konsistenz, Aktualität, Vollständigkeit und Verfügbarkeit der Stammdaten zu Infrastrukturanlagen beeinflusst also positiv die Finanzausstattung des gesamten Unternehmens.

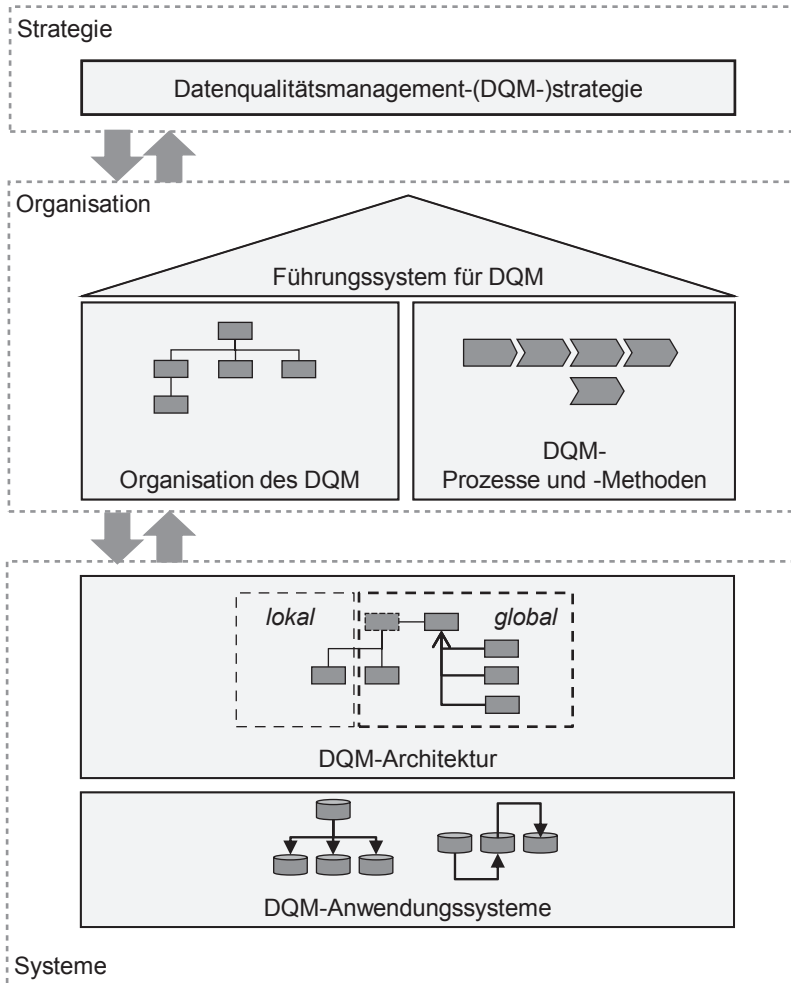


Abb. 1.11 Framework für unternehmensweites Datenqualitätsmanagement. (nach Otto et al. 2011, S. 10)

1.4.3 Organisatorische Ebene

Die organisatorische Ebene umfasst drei Gestaltungsbereiche, nämlich das Führungssystem für Datenqualitätsmanagement (auch: „Datenqualitäts-Controlling“ oder „Qualitätssicherung“), die DQM-Organisation sowie Prozesse und Methoden für DQM.

Datenqualitätsmanagement kann nur dann zielgerichtet betrieben werden, wenn quantifiziert wird, was „gute“ (Stamm-)daten sind. Dazu muss die Qualität der Daten gemessen werden. Datenqualitätskennzahlen sind ein quantitatives Maß für Datenqualität

Tab. 1.3 Ergebnisse der Datenqualitätsstrategie

Ergebnistyp	Prüffragen zum Gestaltungsfortschritt
Handlungsauftrag	Ist der Handlungsauftrag organisatorisch zugeordnet? Weiß die betreffende Stelle um die Aufgaben und Verantwortungen?
Zieldefinition	Sind die Ziele des Datenqualitätsmanagements, also z. B. die kritischen Daten, aus den Unternehmenszielen abgeleitet?
Leitlinien	Sind die Leitlinien des Datenqualitätsmanagements entworfen und kommuniziert?

Tab. 1.4 Ergebnisse des Führungssystems

Ergebnistyp	Prüffragen zum Gestaltungsfortschritt
Kennzahlensystem für Stammdatenqualität	Gibt es Kennzahlen für Datenqualität? Haben die Kennzahlen Geschäftsbezug? Ist das Messverfahren definiert? Gibt es Zielwerte für Datenqualität?
Zielsystemintegration	Sind die Ziele für Datenqualität in das Zielsystem des Unternehmens integriert (z. B. in die Jahreszielvereinbarungen von Mitarbeitern)?

(Hüner 2011)¹¹. Entscheidend beim Aufbau eines Kennzahlensystems für Datenqualität ist herauszufinden, was gemessen werden soll und was gemessen werden kann. Kennzahlensysteme für Datenqualität müssen sich an den fachlichen Erfordernissen orientieren und sind – soweit möglich – mit den Kennzahlen für Geschäftsprozesse zu koppeln. Tabelle 1.4 stellt die Ergebnisse des Gestaltungsbereichs „Führungssystem“ dar.

Weil das Management von Stammdaten ein Querschnittsthema ist, müssen die Aufgaben des Datenmanagements über die einzelnen Divisionen und Geschäftsbereiche des Unternehmens hinweg koordiniert werden. Diesem Zweck dient die Organisation des DQM. Sie ist in vielen Unternehmen eine virtuelle Organisation, in welcher die Mitarbeiter disziplinarisch in ihren ursprünglichen Berichtslinien verbleiben und zusätzlich in einer neuen fachlichen Berichtslinie eingebunden sind (Tab. 1.5).

Die Organisation des Datenqualitätsmanagements manifestiert sich in den Rollen des DQM sowie der Zuordnung von Verantwortlichkeiten zu diesen Rollen. In der Praxis haben sich verschiedene Rollen herausgebildet, um die Aufgaben eines unternehmensweiten Datenqualitätsmanagements wahrzunehmen. Neben der Identifikation und Beschreibung der Rollen im Datenqualitätsmanagement müssen die Verantwortlichkeiten definiert sein. Verantwortlichkeiten geben an, welche Aufgabenbereiche und Rechte (z. B. Anweisungs-, Planungs-, Entscheidungs-, Mitspracherechte) einer Rolle im Stammdatenmanagement zugeordnet sind. Ein Aufgabenbereich ist z. B. die Entwicklung eines einheitlichen Datenmodells für die übergreifend verwendeten Geschäftsobjekte. Hauptverantwortlich

¹¹ Die Nomenklatur und Systematik von Datenqualitäts-Messsystemen ist angelehnt an die Vorgaben zu Messsystemen in der Softwareentwicklung der IEEE Software Society (vgl. IEEE 1998). Ein Kennzahlensystem ist eine spezielle Art eines Messsystems. Siehe Kap. 3.2.2 für Details zu einem Datenqualitäts-Kennzahlensystem.

Tab. 1.5 Ergebnisse der Organisation

Ergebnistyp	Prüffragen zum Gestaltungsfortschritt
Rollen	Sind alle Rollen im Datenmanagement identifiziert, beschrieben und zugeordnet? Nehmen Rolleninhaber die Rolle wahr?
Verantwortlichkeiten	Sind Rollen Entscheidungsbereichen zugeordnet? Werden Entscheidungen gemäß der Zuordnung getroffen? Sind die Verantwortlichkeiten im Unternehmen kommuniziert? Ist bei der Zuordnung das Kongruenzprinzip gewahrt (d. h. Umfang der Aufgabe muss zu Kompetenz und Pflichten passen)?

Tab. 1.6 Ergebnisse der Prozesse und Methoden

Ergebnistyp	Prüffragen zum Gestaltungsfortschritt
Lebenszyklus-Management für Stammdaten	Ist für jede Stammdatenklasse klar definiert, in welchen Aktivitäten der Geschäftsprozesse die Daten erzeugt, verändert, erweitert, gelesen, gelöscht und archiviert werden? Ist der Datenpflegeprozess für diesen Lebenszyklus entworfen, modelliert und implementiert?
Funktionsbeschreibungen	Gibt es klare Funktionsbeschreibungen für die Aktivitäten des Datenqualitätsmanagements im Unternehmen? Sind standardisierte Verfahren definiert und kommuniziert? Sind die Aktivitäten in die Geschäftsprozessarchitektur des Unternehmens eingebettet?

(im Sinne des englischen „*responsible*“) dafür ist in vielen Fällen der Konzern-Datensteward, der auch für den Aufbau des Stammdatenmanagements zuständig ist. Neben dem Konzern-Datensteward werden oft weitere Rollen von Datenverantwortlichen definiert, die das notwendige fachliche und technische Wissen haben, um z. B. das Datenmodell freizugeben oder Nachbesserungen zu fordern. Der Data Owner (auch „Dateneigner“) ist im Sinne des englischen „*accountable*“ für bestimmte Datenobjekte verantwortlich und meist ein Vertreter des Managements (z. B. Leiter Zentraleinkauf, Leiter Supply Chain Management) eines Fachbereichs.

Der vierte Gestaltungsbereich „DQM-Prozesse und -Methoden“ bezieht sich auf das Lebenszyklusmanagement für Stammdaten sowie diejenigen Prozesse, nach denen die Mitarbeiter des Datenqualitätsmanagements arbeiten (Tab. 1.6).

Eine der wichtigsten Ursachen für schlechte Datenqualität ist das Fehlen einer gesamthaften Bewirtschaftung einzelner Stammdatenklassen. Unternehmen sind nach Funktionen (z. B. Einkauf, Vertrieb), Ländern bzw. Märkten und Geschäftsprozessen (z. B. „Order-to-cash“, „Make-to-Stock“) organisiert. Deshalb gibt es in nur wenigen Unternehmen eine Stelle, welche den Gesamtüberblick darüber hat, wo ein Stammdatenum erfasst, geändert, verwendet und zum Löschen markiert wird.

Die Aufgabe, Ursachen und Auswirkungen niedriger Stammdatenqualität zu analysieren, ist deshalb sehr komplex. Ursachen sind zumeist Aktionen, die innerhalb von Anwendungssystemen mit den Daten ausgeführt werden (z. B. anlegen, ändern, ergänzen,

löschen). Diese Aktionen wiederum haben Auswirkungen auf Geschäftsprozesse, deren Qualität sich durch Kennzahlen quantifizieren lässt.

1.4.4 Informationssystemebene

Tab. 1.7 Ergebnisse der Unternehmensdatenarchitektur

Ergebnistyp	Prüffragen zum Gestaltungsfortschritt
Kerngeschäftsjobjektmodell	Sind die Kerngeschäftsjobjekte bekannt? Sind sie eindeutig definiert und beschrieben? Sind Abhängigkeiten untereinander bekannt? Sind unternehmensweite Merkmale bestimmt und definiert?
Stammdatendatenmodell	Gibt es ein Stammdatenmodell, welches aus dem Kerngeschäftsjobjektmodell abgeleitet ist?
Datenhaltungs- und Datenverteilungsjarchitektur	Ist die Datenhaltungsarchitektur (führendes System, Zentralsystem etc.) für jede Stammdatenklasse definiert? Sind die Datenflüsse zwischen den Systemen analysiert?

Die Systemebene umfasst zwei Gestaltungsbereiche, nämlich die DQM-Architektur und die Anwendungssysteme für das Datenqualitätsmanagement.

Die Ergebnisse des Gestaltungsbereichs „DQM-Architektur“ fasst Tab. 1.7 zusammen.

Das Kerngeschäftsjobjektmodell ist ein zentrales Ergebnis des Datenqualitätsmanagements, weil es die Voraussetzung für ein einheitliches Verständnis der Daten und damit auch für die intendierte Nutzung der Daten ist. Seine Entwicklung und sein Fortschritt muss durch Einbeziehung der Fachbereiche erfolgen, weil nur dort das Wissen um die Bedeutung der Stammdaten in den Geschäftsprozessen verfügbar ist. Für die informationstechnische Umsetzung wird das Kerngeschäftsjobjektmodell in ein Konzerndatenmodell überführt.

Die Datenverteilungs- und Datenhaltungsarchitektur beschreibt, welche Daten in welchen Systemen gespeichert werden und zeigt die Datenflüsse zwischen den Systemen. Schließlich bildet die Anwendungssystemlandschaft für das Datenqualitätsmanagement den sechsten Gestaltungsbereich. Die Ergebnisse dieses Gestaltungsbereichs sind in Tab. 1.8 dargestellt.

Dieser Gestaltungsbereich bezieht sich auf die Analyse, den Entwurf, die Implementierung und Verbesserung derjenigen Anwendungssysteme, welche zur Unterstützung des Datenqualitätsmanagements benötigt werden. Dazu gehören zum einen spezielle Stammdatenmanagementsysteme wie SAP Netweaver MDM und zum anderen Softwarewerkzeuge zur Verwaltung des Kerngeschäftsjobjektmodells. In der Auswahl dieser Anwendungssysteme für das Stammdatenmanagement müssen Aspekte der Datenmodellierung, des Datenqualitätsmanagements, der Sicherheit, der Benutzungsschnittstellen, der Datenverteilungsarchitekturen und insbesondere die Art der Integration, sowohl bezogen auf Systeme als auch auf die zu integrierenden Informationen, betrachtet werden. Das Fraun-

Tab. 1.8 Ergebnisse der Anwendungssysteme

Ergebnistyp	Prüffragen zum Gestaltungsfortschritt
Auslegeordnung	Welche Softwarefunktionalität wird für das Stammdatenmanagement heute und in Zukunft benötigt?
Bebauungsplan	Welche Anwendungssysteme stellen welche Funktionalität bereit? Welche Kriterien werden bei der Auswahl und Bewertung der Anwendungssysteme herangezogen?

hofer-Institut für Arbeitswirtschaft und Organisation liefert einen ausführlichen Vergleich etablierter Systeme (Kokemüller 2009).

1.5 Begriffsdefinitionen und Grundlagen

Das Management von Stammdaten und ihrer Qualität ist kein grundsätzlich neues Thema, die Durchdringung aller Wirtschaftsbereiche mit digitalen Services hat nur seinen Stellenwert drastisch gesteigert. Es wird in Forschung und Praxis diskutiert, seit Unternehmen Informationssysteme zur Unterstützung von Geschäftsprozessen einsetzen. Es ist deshalb sowohl für die Forschung als auch die Praxis von besonderer Bedeutung zu wissen, welche Lösungen bereits existieren, welche sich behauptet haben und welche nicht. Dafür gilt es zunächst die zentralen Konzepte und Begriffe des Datenqualitätsmanagements zu klären. Abbildung 1.12 stellt die wichtigsten Begriffe und ihre Beziehung zueinander dar.

1.5.1 Daten und Information

Daten beschreiben die Eigenschaften von Geschäftsobjekten, also materielle und immaterielle Objekte der Realwelt (Boisot und Canals 2004). Zwar gibt es viele Arbeiten zur Unterscheidung von Daten und Information, aber ein eindeutiges und akzeptiertes Verständnis dazu hat sich bisher nicht durchgesetzt (Boisot und Canals 2004; Badenoch et al. 1994). Eine Lehrmeinung fasst Informationen als Wissen auf, das während der menschlichen Kommunikation ausgetauscht wird, während eine zweite eine Informationsverarbeitungsperspektive einnimmt, in der Daten die Bausteine von Information sind (Oppenheim et al. 2003). Demnach werden Daten zu Informationen „verarbeitet“ (Van den Hoven 1999; Holtham 1995; Wang 1998). Nach ISO/IEC 2382-1 sind **Daten** die formalisierte, d. h. für die weiterführende Verarbeitung, Interpretation und Kommunikation geeignete Repräsentation der Eigenschaften von Geschäftsobjekten (ISO/IEC 1993).

Die logische Datenorganisation unterscheidet verschiedene Aggregationsebenen (Chen 1976; Levitin und Redman 1998; Yoon et al. 2000). Datenelemente bilden die unterste Aggregationsebene. Datenelemente sind die Instanzierungen der Attribute von Datenobjekten (z. B. Nachname eines Kunden). Die zweite Aggregationsebene bilden Datensätze.

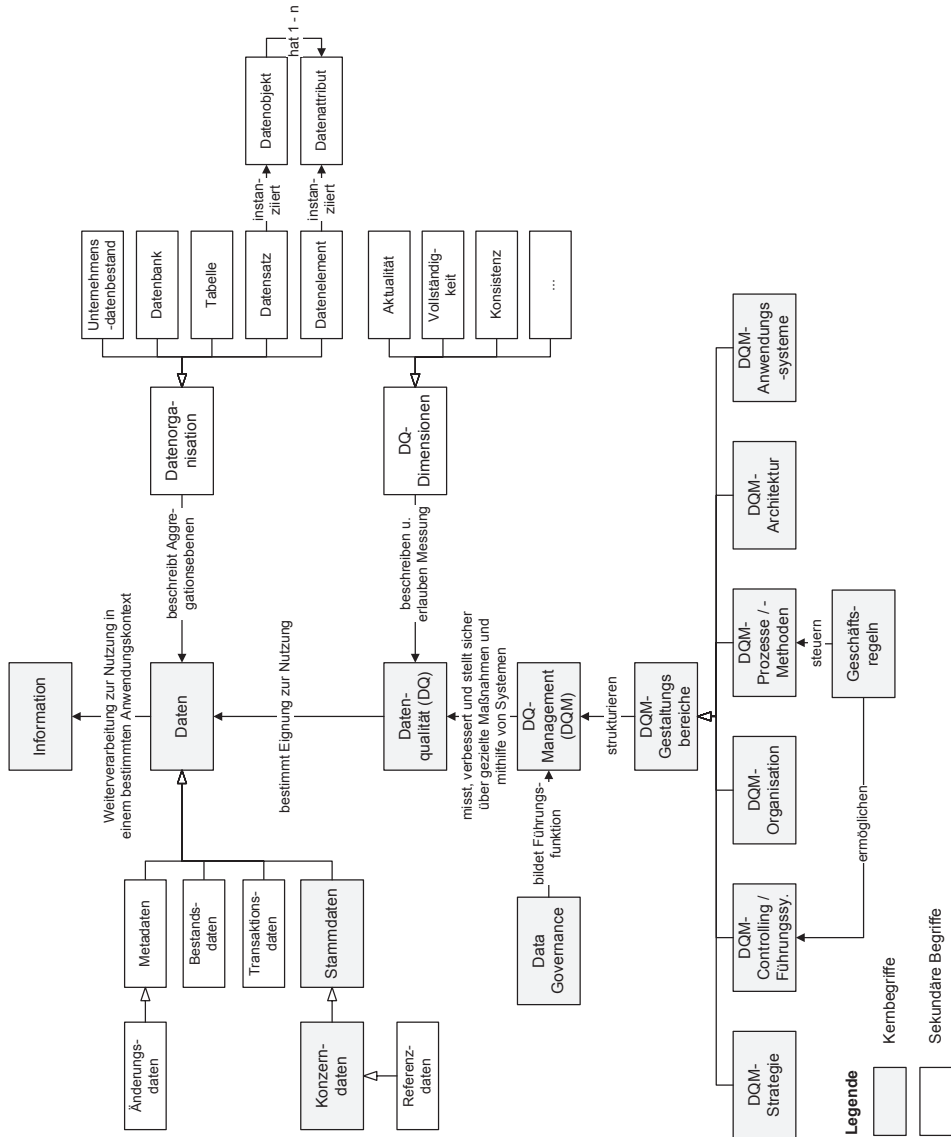


Abb. 1.12 Begriffslkarte für das Datenqualitätsmanagement. (eigene Darstellung)

Ein Datensatz ist die Instanziierung eines Datenobjekts. Zum Beispiel bildet ein Kundenstammdatensatz mit allen Merkmalen das Geschäftsobjekt „Kunde“ so ab, dass alle Geschäftsprozesse rund um den Kunden (Vertrieb, Service, Debitorenbuchhaltung) reibungslos ablaufen. Auf der dritten Aggregationsebene fassen Tabellen mehrere Datensätze zusammen, z. B. in einer Kundenstammdatentabelle. Datenbanken wiederum aggregieren mehrere Tabellen. Eine Kundenmanagement-Datenbank könnte alle Kundenstammdaten sowie die zugehörigen Vertriebsdaten enthalten. Die Gesamtheit aller Datenbanken im Unternehmen bildet schließlich den Unternehmensdatenbestand (engl. Data Resource).

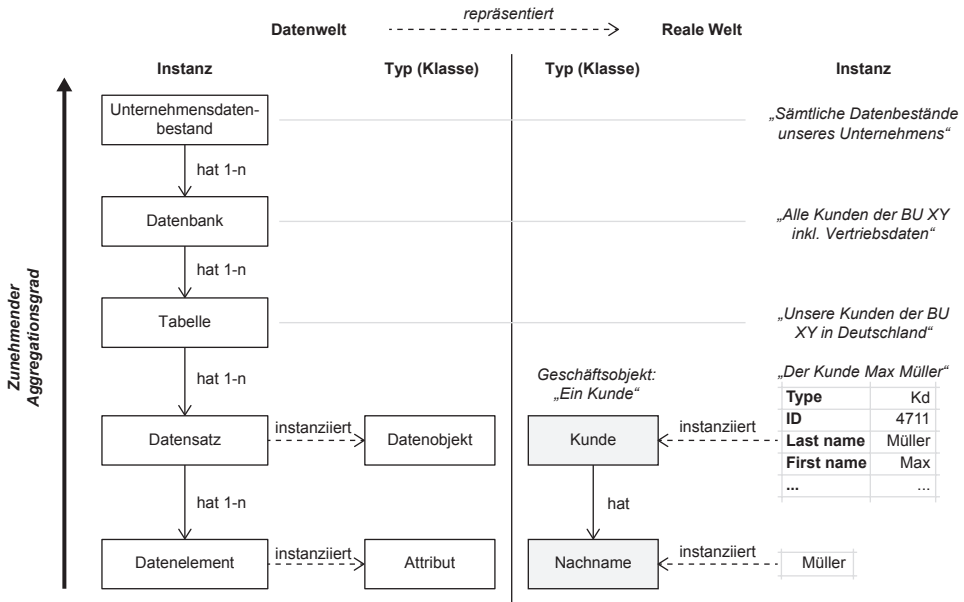


Abb. 1.13 Übersicht der logischen Datenorganisation. (eigene Darstellung)

Abbildung 1.13 visualisiert diese Beziehungen. Diese Darstellung der Datenorganisation richtet sich nach dem relationalen Datenmodell. Im idealen, semantisch eindeutigen Fall gibt es eine 1:1-Beziehung zwischen der Datenwelt und der realen Welt, d. h. ein Datenobjekt bildet genau ein Geschäftsobjekt ab. In der Realität gibt es jedoch oft mehrere Datenobjekte parallel, die dasselbe Geschäftsobjekt repräsentieren. In diesem Fall ist das Datenqualitätsmanagement gefordert, Richtlinien, Prozesse und Systeme zu etablieren, die erlauben, das für einen bestimmten Kontext „richtige“ Datenobjekt zu identifizieren und für die Nutzung in Geschäftsprozessen bereitzustellen.

1.5.2 Stammdaten

Der Standard ISO 8000 der International Organization for Standardization (ISO) (ISO 2009) definiert **Stammdaten** als Informationsobjekte, „welche unabhängig und fundamental für eine Organisation sind. [Stammdaten] müssen referenziert werden, um Transaktionen durchführen zu können.“ Diese Daten müssen innerhalb eines Unternehmens über mehrere Organisationseinheiten hinweg eindeutig identifiziert und einheitlich interpretiert werden. Dabei sind **Konzerndaten** sogenannte „globale“ Stammdaten, die für das gesamte Unternehmen gelten. Im Gegensatz dazu gelten „lokale“ Stammdaten nur z. B. für einen Geschäftsbereich, einen Standort oder eine Unternehmensfunktion. Im Fokus der hier vorgestellten Methoden und Werkzeuge stehen die Konzerndaten. Da sich im allgemeinen Sprachgebrauch sowie in kommerziellen Softwarelösungen jedoch der Begriff

„Stammdaten“ durchgesetzt hat, verwendet das Buch überwiegend diese gängigere Bezeichnung.

Stammdaten verändern sich im Gegensatz zu Bewegungs- oder Bestandsdaten vergleichsweise selten. Diese Daten werden darum manchmal auch als „static data“ (statische Daten) bezeichnet. In der Praxis ist es nicht möglich, eine allgemeingültige Liste aller Stammdatenklassen zu definieren. So gelten z. B. Daten zu Verträgen in der Energie- und Versicherungswirtschaft als Stammdaten, in der Telekommunikationsbranche hingegen aufgrund der im Vergleich kurzen Laufzeiten und häufigen Änderungen faktisch als Bewegungsdaten.

Die Firma Bosch z. B. klassifiziert folgende Daten als Stammdaten (Hatz 2008):

- Kunden
- Kundenhierarchien
- Materialien
- Lieferanten
- Mitarbeiter
- Kontenpläne
- Organisationseinheiten

In der Praxis ist weniger relevant, wie Stammdaten von Bewegungs- und Bestandsdaten abzugrenzen sind, sondern vielmehr, welche einzelnen **Attribute** einer Stammdatenklasse von einem zentralen Stammdatenmanagement verwaltet werden müssen. Denn eine zentrale Organisationseinheit kann diese Aufgabe aufgrund der Komplexität einzelner Stammdatenklassen nicht in vollem Umfang unternehmensweit übernehmen. Grundsätzlich beantwortet sich die Frage, welche Attribute einer Stammdatenklasse zum Umfang des zentralen Stammdatenmanagements gehören, aus der Analyse der strategischen Anforderungen jedes einzelnen Unternehmens.

Dabei können nach (White 2010; White und Radcliffe 2007) folgende Unterscheidungsmerkmale helfen:

- Organisatorische Reichweite: Unterscheidung zwischen „globalen“, also unternehmensweit genutzten Konzerndaten und lokalen Daten
- Datentyp: Unterscheidung zwischen „strukturierten“ Daten, welche typischerweise in relationalen Datenbanken verwaltet werden, und „unstrukturierten“ Daten wie Produktinformationen (z. B. Bilder, Werbetexte, Applikationsvideos)
- Ort der Metadatendefinition: Unterscheidung zwischen interner Definition von Bedeutung, Formaten sowie Standardwerten einerseits und externer Definition andererseits (zum Beispiel bei Länder- und Währungs-codes der ISO sowie Klassifikationsstandards wie eCl@ss und UN/SPSC)

Stammdaten, die extern definiert sind, heißen **Referenzdaten**. Beispiele sind, wie oben erwähnt, Ländercodes und Währungs-codes sowie Geodaten. **Metadaten** (wörtlich „Daten

über Daten“) beschreiben und definieren Eigenschaften von anderen Daten (DAMA 2008, S. 84). Eine Unterart von Metadaten sind z. B. Änderungsdaten. Diese halten fest, *wann*, *wie* und *von wem* ein bestimmtes Datum geändert wurde.

1.5.3 Datenqualität

Datenqualität ist ein mehrdimensionales, kontextabhängiges Konzept (Wang und Strong 1996). Es gibt also nicht ein einziges Merkmal, das Datenqualität vollständig beschreibt. Vielmehr gibt es verschiedene Datenqualitätsdimensionen, die in ihrer Gesamtheit die Qualität von Daten beschreiben. Typische Datenqualitätsdimensionen sind¹²:

- **Korrektheit:** Stimmen die Daten sachlich richtig mit den Eigenschaften des Objekts in der realen Welt überein, das sie repräsentieren sollen?
- **Konsistenz:** Stimmen mehrere Datenversionen desselben realen Objekts, die z. B. in unterschiedlichen Informationssystemen gehalten werden, miteinander überein?
- **Vollständigkeit:** Sind alle Werte/Attribute eines Datensatzes komplett vorhanden?
- **Aktualität:** Stimmen die Daten zu jedem Zeitpunkt mit dem aktuellen Status des realen Objekts überein und werden die Daten angepasst, wenn es sich ändert?
- **Verfügbarkeit:** Sind die Daten für Datennutzer zum gewünschten Zeitpunkt problemlos zugänglich?

Kontextabhängigkeit bedeutet, dass Datenqualität für einen Geschäftsvorfall ausreichend sein kann, für einen anderen hingegen ungenügend. Zum Beispiel ist für die Logistikabteilung eines Automobilzulieferers die korrekte Lieferadresse eines Kunden einschließlich der korrekten Frachtrampe essentiell zur Bedienung eines Auftrags. Für die Vertriebsabteilung desselben Unternehmens sind Korrektheit und Konsistenz der Lieferadresse dagegen unerheblich, da ihr z. B. für eine Auswertung über den mit diesem Kunden im vergangenen Jahr eingekommenen Umsatz allein Unternehmensname und Landeszuordnung ausreicht. Deshalb ist Datenqualität insgesamt definiert als ein Maß für die Eignung der Daten für bestimmte Anforderungen in Geschäftsprozessen, in denen sie verwendet werden (Otto et al. 2011).

Datenqualität ändert sich über die Zeit, weil die Daten lediglich ein Abbild der Wirklichkeit darstellen, sich diese Wirklichkeit aber verändert. Zum Beispiel ziehen Kunden um und Lieferanten ändern ihre Rechtsform. Abbildung 1.14 stellt einen typischen Datenqualitätsverlauf dar, wie er in vielen Unternehmen zu finden ist.

Viele Unternehmen beginnen erst mit der Lösung der Datenqualitätsprobleme, wenn die Datenqualität bereits unter ein Maß gesunken ist, das reibungslose Geschäftsprozesse

¹² Literatur und Praxis definieren eine Vielzahl unterschiedlicher Datenqualitätsdimensionen. Quellen finden sich z. B. bei DAMA (2009) oder bei der International Association for Information and Data Quality <http://iaidq.org/main/glossary.shtml#D>. Hier sind die fünf wichtigsten Dimensionen wiedergegeben.

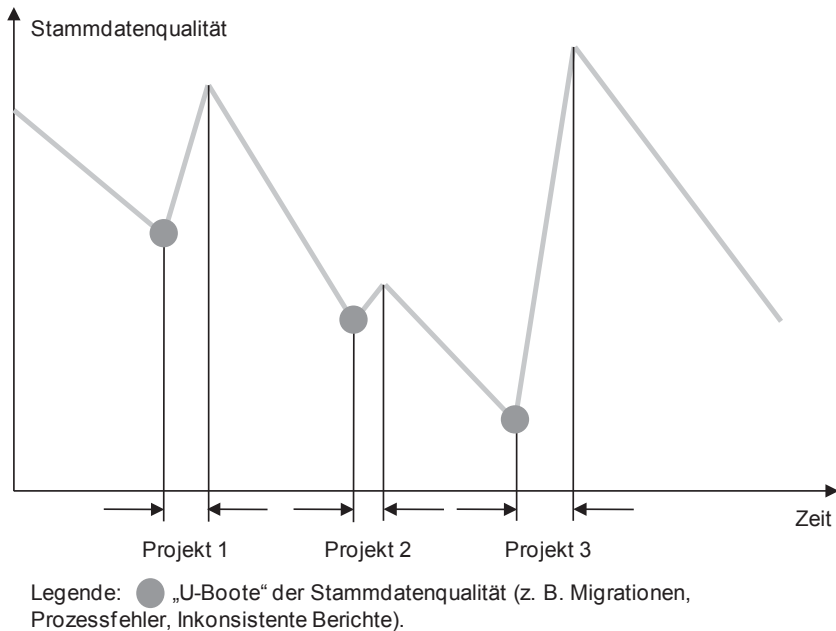


Abb. 1.14 Typischer Datenqualitätsverlauf über die Zeit in Unternehmen. (Otto 2014, S. 21)

ermöglicht. Migrationsprobleme, hoher manueller Aufwand in Geschäftsprozessen, Management Reports mit unterschiedlichen Werten zur gleichen Kennzahl sind Beispiele für die Datenqualitätsprobleme, die plötzlich sichtbar werden, nachdem die Datenqualität zuvor schleichend abgefallen ist.

1.5.4 Datenqualitätsmanagement (DQM)

Die Analyse, Verbesserung und Sicherung der Datenqualität ist Aufgabe des DQM. Wiederum unter Übernahme produktionswirtschaftlicher Ansätze versteht die Data Management Association (DAMA) unter DQM sämtliche Aktivitäten, Verfahren und Systeme, die unter Nutzung von Methoden des Qualitätsmanagements die Eignung der Daten zur Nutzung messen, verbessern und sichern (DAMA 2008).

DQM unterscheidet generell zwischen präventiven und reaktiven Maßnahmen. Präventives DQM zielt darauf ab, Datendefekte mit negativer Auswirkung auf die Datenqualität zu vermeiden. Im Gegensatz dazu zielt das reaktive DQM darauf ab, bestehende Datendefekte zu entdecken und zu beheben.

Der reaktive Ansatz hat mehrere Nachteile:

- Ressourcen für die Verbesserung der Datenqualität (z. B. Software, Expertenwissen, Beratungsleistungen) sind nicht eingeplant und budgetiert und damit im Zweifel nicht verfügbar.
- Rein reaktives Datenqualitätsmanagement geht häufig damit einher, dass Datenqualität nicht gemessen wird. In solchen Fällen fehlen dem Unternehmen häufig Zielwerte, d. h. es ist nicht überprüfbar, ob die (reaktiven) Maßnahmen zur Verbesserung der Datenqualität ausreichen oder gar über das Ziel „hinausschießen“.
- Das Total Quality Management, ein Qualitätsmanagementansatz aus dem Operations Management, zeigt, dass die Summe der Kosten aller reaktiven Maßnahmen im Qualitätsmanagement die Kosten eines präventiven Qualitätsmanagements übersteigt (Reid und Sanders 2005). Das gilt für materielle Güter wie für immaterielle (z. B. Daten).

Ein Beispiel für eine präventive DQM-Maßnahme ist die Nutzung von automatischen Prüfregeln (auch „Geschäftsregeln“, siehe folgender Abschnitt) bei einer manuellen Dateneingabe. Ein wichtiges und von vielen Unternehmen verwendetes Stammdatenum ist die DUNS-Nummer (Data Universal Numbering System), ein Zahlencode zur Identifikation von Unternehmen. Für diesen Fall sind Dienste verfügbar, die in Echtzeit die Gültigkeit der in ein systemgestütztes Formular eingegebenen Nummer überprüfen und so eine fehlerhafte Eingabe verhindern. Ein Beispiel für eine reaktive Maßnahme im selben Fall ist die nachträgliche Datenbereinigung von Duplikaten im gleichen oder in verschiedenen Datenbeständen.

Grundsätzliches Ziel von Unternehmen ist es, Datendefekte möglichst zu erkennen, bevor sie eintreten, um Risiken und Kosten infolge mangelnder Datenqualität zu vermeiden. Jedoch verursachen nicht nur defekte Daten Kosten, sondern auch das DQM. DQM-Kosten fallen sowohl für präventive als auch für reaktive Maßnahmen an. In Analogie zum Qualitätsmanagement generell (Campanella 1999) kann ein überproportionaler Zusammenhang zwischen Datenqualität und DQM-Kosten unterstellt werden. Die Grenzkosten des DQM nehmen also zu. Im Gegensatz dazu nehmen die Folgekosten schlechter Daten ab, je höher die Datenqualität ist. Die gesamtkostenoptimale Datenqualität ergibt sich dann als Minimum der Summenfunktion aus DQM-Kosten und Folgekosten defekter Daten (Eppler und Helfert 2004) (Abb. 1.15). Aufgabe des DQM ist es somit, eine kostenoptimale Kombination aus präventiven und reaktiven Maßnahmen zu finden. In der Praxis treten dabei häufig Schwierigkeiten auf, weil das Rechnungswesen typischerweise viele DQM-Kosten nicht ausweist. Das gilt insbesondere für die Folgekosten, die in verschiedenen Unternehmensfunktionen anfallen und in weiten Teilen kaum zu quantifizieren sind.

1.5.5 Geschäftsregeln (Business Rules)

Die automatische Prüfung von Geschäftsregeln (engl. Business Rules) ist ein wichtiges Mittel sowohl für das proaktive als auch für das reaktive Datenqualitätsmanagement. Geschäftsregeln definieren die Ausführung und die Leistung von Geschäftsprozessen (Ross und Lam 2011). Geschäftsregeln formalisieren damit die Richtlinien (engl. business poli-

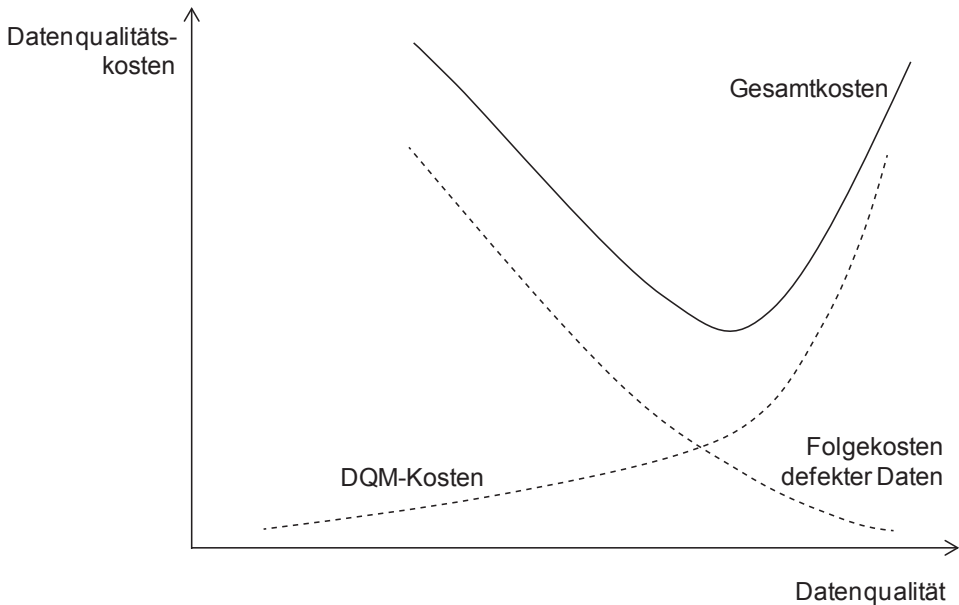


Abb. 1.15 Datenqualitätskosten. (nach Eppler und Helfert 2004, S. 318)

cies) im Unternehmen (OMG 2008). Sie helfen häufig dabei, dass Wissen „in den Köpfen“ der Mitarbeiter zu explizieren, damit es zur Wiederverwendung und Automatisierung in IT-Systemen festgehalten werden kann. Geschäftsregeln sind das Bindeglied zwischen dem Geschäftsprozess- und dem Datenmanagement. Denn neben der Steuerung von Geschäftsprozessen dienen Geschäftsregeln auch dazu, Datenqualität zu messen und zu kontrollieren. In der Praxis werden diese Regeln teilweise auch als „Validierungsregeln“ oder „Prüfregeln“ bezeichnet.

Einfache Geschäftsregeln definieren z. B. Pflichtfelder in Eingabemasken. So könnte eine einfache Regel für die Anlage eines Kundenstamms lauten, dass ein Kundenstammdatensatz nur dann vollständig ist, wenn er eine Rechtsform (GmbH, AG usw.) besitzt. Wird dieses Feld nicht ausgefüllt, kann der Datensatz nicht weiter bearbeitet werden. Im Standard SBVR (Semantics of Business Vocabulary and Business Rules) der Object Management Group werden solche Regeln als **strukturelle Regeln** (engl. structural rules) bezeichnet (OMG 2008). Ein komplexeres Beispiel wäre beispielsweise eine Regel, die steuert, nach welchen Kriterien über die weitere Behandlung einer Beschwerde im Kundenservice entschieden werden sollte. Eine solche Regel wird in SBVR als **operative Regel** (engl. operative rule) bezeichnet. Tabelle 1.9 zeigt ein Beispiel für strukturelle und operative Regeln.

Formale Sprachen wie SBVR oder BPMN (Business Process Model and Notation) dokumentieren Geschäftsregeln so, dass sie sich einfach in Programmcode umsetzen lassen bzw. direkt maschinenlesbar sind. Die Gesamtmenge von Geschäftsregeln eines Unternehmens kann in einer Rule Engine verwaltet werden. Es ist Aufgabe des Datenqualitäts-

Tab. 1.9 Beispiel für eine Geschäftsrichtlinie und Geschäftsregeln. (Ofner 2013, S. 51)

Geschäftsrichtlinie	Neue Lieferanten müssen in Übereinstimmung mit (globalen, national oder regionalen) gesetzlichen Vorgaben und Bestimmungen aufgenommen werden. Der Prozess muss nachvollziehbar sein.
Strukturelle Regel	Es ist notwendig (engl. nach SBVR-Standard: „It is necessary that“), dass jeder Lieferant einen Sicherheitskennzeichen hat.
Operative Regel	Es ist unerlässlich (engl. nach SBVR-Standard: „It is obligatory that“), dass das Sicherheitskennzeichen jedes Lieferanten, der Produkte in die USA liefert, auf „ja“ steht.

managements, die für die Sicherstellung hoher Datenqualität relevanten Geschäftsregeln zu identifizieren, zu dokumentieren, umzusetzen und zu pflegen.

Zusammengefasst bieten Geschäftsregeln folgende Vorteile für das DQM:

- **Automatisierung:** Sie automatisieren Teile von Geschäftsprozessen und können diese damit beschleunigen und manuelle Aufwände reduzieren (z. B. durch automatisches Vor-Ausfüllen IT-gestützter Formulare).
- **Fehlervermeidung:** Geschäftsregeln helfen dabei, Flüchtigkeits- und vorsätzliche Fehler zu vermeiden (Prüfregeln verhindern, dass obligatorische Felder „vergessen“ werden können). Dies ist eine Maßnahme für proaktives Datenmanagement.
- **Messung und Steuerung der Datenqualität:** Datensätze können erst kontinuierlich oder periodisch auf Regelkonformität überprüft werden, nachdem Geschäftsregeln definiert und in IT-Systeme implementiert wurden. Nur so können automatisiert Fehler in Datensätzen entdeckt werden, bevor diese Fehler in Geschäftsprozessen verursachen. Dieses Wissen kann erstens dazu verwendet werden, Maßnahmen zur Verbesserung der Datenqualität zu initiieren (Maßnahme des reaktiven Datenqualitätsmanagements). Zweitens kann das Ergebnis der Datenqualitätsmessung über Geschäftsregeln in einem Datenqualitäts-Indikatorwert (auch „Datenqualitätsindex“ oder „Datenqualitäts-Zielwert“ genannt) ausgedrückt werden. Mit diesem Wert lässt sich die Veränderung der Datenqualität in einem System im Laufe der Zeit nachvollziehen und z. B. auch die Datenqualitätsentwicklung zwischen verschiedenen Geschäftseinheiten vergleichen.
- **Bewusstsein für DQM:** Eine Initiative zur Bestimmung von Geschäftsregeln trägt dazu bei, bei allen beteiligten Mitarbeitern ein Bewusstsein für die Bedeutung von DQM im Allgemeinen zu schaffen. Damit erhöhen sich die Chancen, dass Mitarbeiter auch solche Datenqualitätsprobleme bemerken, die nicht mit Regeln geprüft werden können.

Allerdings ist die Verwendung von Geschäftsregeln auch mit einigen Herausforderungen verbunden:

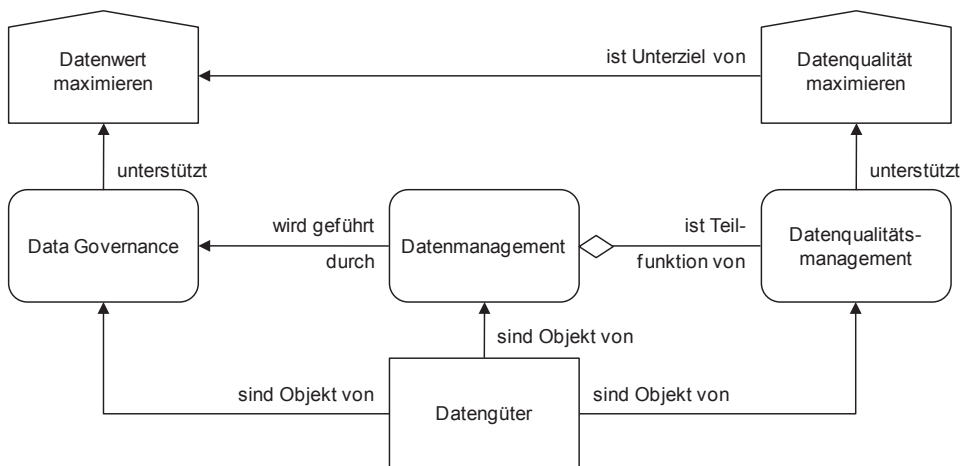
- **Explication:** Da das Wissen über Geschäftsregeln meist nicht in Dokumenten vorliegt (zumindest nicht an einer einzelnen Stelle), sondern „in den Köpfen der Mitarbeiter“ verborgen ist, besteht die größte Herausforderung darin, dieses Wissen zu dokumentieren¹³.
- **Verringerung der Prozessflexibilität:** Sind Geschäftsregeln zu restriktiv definiert, können sie bei seltenen oder unvorhergesehenen Prozessvarianten zu Mehraufwänden oder erst recht zu falschen Eingaben führen, wenn Mitarbeiter Regeln „umgehen“ müssen.
- **Überbewertung der Aussagekraft eines Datenqualitätsindex:** Ein guter Datenqualitätsindex sagt lediglich aus, dass die Daten gemessen an allen bekannten und implementierten Regeln korrekt sind. Falls diese Regeln aber unvollständig oder veraltet sind, wird das Datenqualitätslevel überschätzt. Daraus folgt, dass Geschäftsregeln aktiv verwaltet werden müssen: Wenn sich die Geschäftswirklichkeit ändert, müssen auch die Regeln geändert werden. Darüber hinaus müssen Datenqualitätsmanager akzeptieren, dass niemals sämtliche Datenqualitätsprobleme mit automatischen Prüfregelel erfassbar sind. Deshalb darf die Schulung und fortwährende Motivation der Mitarbeiter, im Alltag mit „gesundem Menschenverstand“ auf Datenqualität zu achten, auch nach der Implementierung von Geschäftsregeln nicht vernachlässigt werden.

1.5.6 Data Governance

Data Governance unterscheidet sich von DQM. Data Governance verfolgt das Ziel, den Wert der Daten im Unternehmen zu maximieren (Otto 2011a). Die wertmäßige Betrachtung von Daten im Sinne eines Anlageguts geht ebenfalls zurück auf die Übertragung von Konzepten zur Bewirtschaftung materieller Güter auf den Umgang mit Daten (Horne 1995). Heute diskutieren Forschung und Praxis, ob der Wert von Daten für Unternehmen auch finanzbuchhalterisch erfasst werden kann und soll (Atkinson und McGaughey 2006). Grundsätzlich besitzen Daten nur dann einen Wert, wenn sie genutzt werden. Ihre Eignung zur Nutzung wiederum ist definiert als Datenqualität (s. o.). Niedrige Datenqualität schmälert den Wert der Datengüter im Unternehmen, weil ihre Nutzbarkeit gering ist (Even und Shankaranarayanan 2007). Unternehmen sind also bestrebt, mit dem DQM eine von der Geschäftsstrategie geforderte Datenqualität zu erreichen.

Die Beziehung zwischen DQM und Data Governance folgt der von der ISO vorgeschlagenen Unterscheidung zwischen Governance und Management (ISO/IEC 2008). In diesem Sinne bildet Data Governance die Führungsfunktion für das DQM. Denn Data

¹³ Der japanische Wissenschaftler Ikujiro Nonaka beschrieb 1991 in seinem Aufsatz „The Knowledge-Creating Company“ diesen Prozess als „Artikulation von implizitem Wissen“ („articulation of tacit knowledge“) (Nonaka 2007). Wir verwenden hier bewusst den Begriff „Explication“, da er dem heutigen Sprachgebrauch und den jüngeren auf Nonaka aufbauenden Quellen entspricht.



Legende: Ziel Funktion Daten

Abb. 1.16 Datenqualitätsmanagement und Data Governance. (Otto 2011a, S. 236)

Governance legt fest, welche Entscheidungen im Umgang mit Daten zu treffen sind und wer sie trifft. Aufgabe des DQM ist es, die datenqualitätsrelevanten Entscheidungen zu fällen und umzusetzen.

Abbildung 1.16 stellt diesen Zusammenhang grafisch dar.

1.6 Kompetenzzentrum Corporate Data Quality

Das Kompetenzzentrum Corporate Data Quality (CC CDQ) ist ein Konsortialprojekt, das Lösungen für die dauerhafte Sicherung von Datenqualität in großen Unternehmen erforscht, entwickelt und überprüft.

Konsortialforschung findet im Verbund von Forschungseinrichtungen und Unternehmen statt, die an einem Thema von gemeinsamem Interesse arbeiten. Konsortialforschung hat mehrere Ziele (Österle und Otto 2010):

- Forscher und Praxispartner definieren gemeinsam die Forschungsziele, bewerten die laufende Arbeit und evaluieren die Projektergebnisse.
- Mehrere Partnerunternehmen bringen ihre Expertise ein und gewähren den Forschern Zugang zu ihrem Wissen.
- Die Forschungsergebnisse sind Artefakte (z. B. Methoden, Modelle oder Prototypen), die zur Lösung praktischer Probleme beitragen.
- Der Gestaltungsprozess ist mehrfach iterativ und umfasst Iterationszyklen über vier Phasen und mehrere Partnerunternehmen.
- Die Partnerunternehmen testen die Artefakte in ihrem betrieblichen Umfeld.
- Die Partnerunternehmen finanzieren das Projekt mindestens in Teilen.

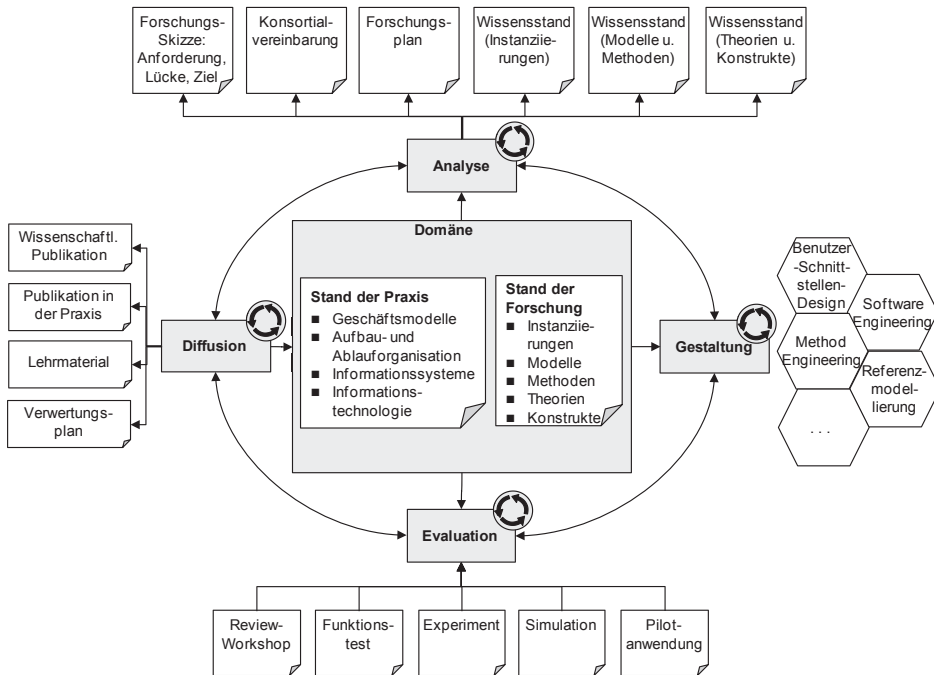


Abb. 1.17 Konsortialforschung. (Österle und Otto 2010, S. 278)

- Forscher und Praktiker nehmen über einen signifikanten Zeitraum hinweg an dem Projekt teil (in der Regel zwei Jahre).
- Die Forschungsergebnisse werden der Öffentlichkeit zugänglich gemacht.

Konsortialforschung läuft, nach ähnlichen Prinzipien wie der Action Design Research (Sein et al. 2011), zyklisch ab, d. h. Ergebnisse werden kontinuierlich an den Anforderungen der Forschung und Praxis gemessen, weiterentwickelt und getestet. Abbildung 1.17 stellt diesen Ablauf dar.

In regelmäßigen Workshops treffen sich die Mitglieder des Konsortiums und tauschen Erfahrungen aus, besprechen offene Punkte und arbeiten gemeinsam an Methoden, Modellen und Lösungen. Sogenannte bilaterale Projekte überprüfen die Ergebnisse anschließend unternehmensindividuell.

Das CC CDQ startete im November 2006 mit dem Ziel, Antworten auf folgende Fragen zum Datenqualitätsmanagement zu finden:

- Welchen Beitrag liefert Stammdatenqualität zu den Unternehmenszielen?
- Wie steht das eigene Unternehmen im Vergleich zu anderen?
- Wie lässt sich die Leistung des Datenqualitätsmanagements messen?
- Was sind Kosten und Nutzen der Datenqualität?
- Wie etablieren Unternehmen Data Governance in der Organisation?

- Was ist das richtige Maß an Standards und Regulierung für die Daten?
- Wie ist ein gemeinsames Verständnis über die Stammdaten im ganzen Unternehmen zu schaffen?
- Was ist die richtige Datenarchitektur?
- Welchen Beitrag leisten innovative Technologien wie das Semantic Web und In-Memory Computing?

Seit 2006 nahmen 30 Unternehmen am CC CDQ teil. Durchschnittlich sind ca. zehn bis vierzehn Unternehmen gleichzeitig aktiv und die Verweildauer im Konsortium reicht von zwei bis über sieben Jahre. Tabelle 1.10 listet sämtliche aktuellen und ehemaligen Unternehmen auf, die seit der Gründung des CC CDQ im Jahr 2006 Partner im Kompetenzzentrum waren.

In den meisten Fällen nehmen Vertreter sowohl aus Fachbereichen (z. B. Supply Chain Management, Finanzen) und Informatikabteilung am CC CDQ teil. Dadurch ist gewährleistet, dass immer eine fachliche und eine informationstechnische Perspektive auf das Thema eingenommen werden.

Die Fallstudien in Kap. 2 und die Werkzeuge und Methoden in Kap. 3 sind Ergebnisse des CC CDQ.

Tab. 1.10 Mitglieder des CC CDQ seit Gründung im Jahr 2006 (in alphabetischer Reihenfolge)

Unternehmen	Dauer der Mitgliedschaft im CC CDQ
ABB Ltd.	2014 – Gegenwart ^a
AO Foundation	2011 – 2012
AstraZeneca PLC	2012 – Gegenwart
Bayer AG	2006 – Gegenwart
Beiersdorf AG	2010 – Gegenwart
Corning Cable Systems GmbH	2012
Daimler AG	2007–2008
DB Netz AG	2008–2009; 2014 – Gegenwart
Drägerwerk AG & Co. KGaA	2013 – Gegenwart
eCl@ss e. V.	2014 – Gegenwart
E.ON SE	2007–2008
Ericsson AB	2014 – Gegenwart
ETA SA	2006–2008
Festo AG & Co. KG	2010– Gegenwart
Hewlett-Packard GmbH	2008–2010
IBM Deutschland GmbH	2007–2011
Kion Information Management Service GmbH	2010–2012
Merck KGaA	2014 – Gegenwart
Migros-Genossenschafts-Bund	2009
Nestlé SA	2008 – Gegenwart

Tab. 1.10 (Fortsetzung)

Unternehmen	Dauer der Mitgliedschaft im CC CDQ
Novartis Pharma AG	2008–2010; 2014
Osram GmbH	2013 – Gegenwart
Robert Bosch GmbH	2007 – Gegenwart
SAP AG	2012–2014
Schweizerische Bundesbahnen SBB	2013 – Gegenwart
Schaeffler AG	2014 – Gegenwart
Siemens Enterprise Communications GmbH & Co. KG	2010–2013
Swisscom (Schweiz) AG	2012 – Gegenwart
Syngenta Crop Protection AG	2009–2013
Telekom Deutschland GmbH	2008–2010
ZF Friedrichshafen AG	2007 – Gegenwart

^a Entspricht dem Zeitpunkt der Fertigstellung des Manuskripts im Februar 2015

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung-Nicht kommerziell 4.0 International Lizenz (<http://creativecommons.org/licenses/by-nc/4.0/deed.de>) veröffentlicht, welche für nicht kommerzielle Zwecke die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Literatur

- AFNOR (Association Française de Normalisation): General principles for an environmental communication on mass market products. La Plaine Saint-Denis, 2009
- Aggarwal, Charu C.; Ashish, Naveen; Sheth, Amit P.: The Internet of Things: A Survey from the Data-Centric Perspective. In: Aggarwal, C. C. (Hrsg.): *Managing and Mining Sensor Data*. New York: Springer, 2013, S. 383–428. – DOI 10.1007/978146146309212
- Amiona: AppMapp. – Internetquelle. Aktualisierungsdatum 2014-02-14. Abgerufen 2014-12-05 von: <http://www.amiona.com/index.php/downloads/viewdownload/3-appmap/2-amiona-appmap>
- Anthes, Gary: Data Brokers Are Watching You. In: *Communications of the ACM* 58 (2015), Nr. 1, S. 28–30. – DOI 10.1145/2686740
- Atkinson, Keith; McGaughey, Ronald: Accounting for data: a shortcoming in accounting for intangible assets. In: *Academy of Accounting and Financial Studies Journal* 10 (2006), Nr. 2, S. 85–95
- Baars, Henning; Kemper, Hans-Georg: Management Support with Structured and Unstructured Data: An Integrated Business Intelligence Framework. In: *Information Systems Management* 25 (2008), Nr. 2, S. 132–148. – DOI 10.1080/10580530801941058
- Badenoch, Douglas; Reid, Christine; Burton, Paul; Gibb, Forbes; Oppenheim, Charles: The value of information. In: Feeney, Mary; Grieves, Maureen (Hrsg.): *The value and impact of information*. London: Bowker-Saur, S. 9–75
- Bauernhansl, Thomas ; ten Hompel, Michael ; Vogel-Heuser, Birgit (Hrsg.): *Industrie 4.0 in Produktion, Automatisierung und Logistik*. Wiesbaden : Springer Fachmedien, Vieweg+Teubner, 2014. – DOI 10.1007/978-3-658-04682-8

- BBC (British Broadcasting Corporation): Sony fined over ‚preventable‘ PlayStation data hack. – Internetquelle. Aktualisierungsdatum 2013-05-10. Abgerufen 2015-02-05 von <http://www.bbc.co.uk/news/technology-21160818>.
- Bitkom: Big-Data-Technologien – Wissen für Entscheider: Leitfaden. Berlin: Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V. 2014
- Boisot, Max; Canals, Agustí: Data, information and knowledge: have we got it right? In: *Journal of Evolutionary Economics* 14 (2004), Nr. 1, S. 43–67. – DOI 10.1007/s0019100301819
- Brenner, Walter; Herrmann, Andreas: Das Modell des Managements in der digitalen vernetzten Welt. In: Stadler, R.; Brenner, W.; Herrmann, A. (Hrsg.): *Erfolg im digitalen Zeitalter: Strategien von 17 Spitzenmanagern*, 1. Auflage. Frankfurt: Frankfurter Allgemeine Buch, 2012. – ISBN 978-3899813005
- Campanella, Jack: *Principles of quality costs : principles, implementation and use*. Milwaukee : ASQ Quality Press, 1999
- Chen, Peter Pin-Shan: The Entity-Relationship Model: Toward a Unified View of Data. In: *ACM Transactions on Database Systems* 1 (1976), Nr. 1, S. 9–36. – DOI 10.1145/320434320440
- Dahlkamp, Jürgen; Schmitt, Jörg: Die Dotcom-Räuber. In: *Der Spiegel* 48 (2014-11-24), S. 68–69. – Abgerufen 2015-01-15 von: <http://www.spiegel.de/spiegel/print/d-130458636.html>
- DAMA International: DAMA-DMBOK Functional Framework. Version 3.02. Lutz: Technics Publications, 2008
- DAMA International: *The DAMA Guide to The Data Management Body of Knowledge (DAMAD-MBOK Guide)*. Bradley Beach: Technics Publications, 2009. – ISBN 978-0-9771400-8-4
- Eppler, Martin J.; Helfert, Markus: A Framework for the Classification of Data Quality Costs and an Analysis of Their Progression. In: Chengalur-Smith, I., Raschid, L., Long, J., Seko, C. (Hrsg.): *Proceedings of the 9th International Conference on Information Quality (ICIQ-04) (2004-11-05)*. Cambridge, 2004, S. 311–325
- Equy, Catherine; Kusters, Rob J.; Varone, Sasha; Montandon, Nicolas: Empirical Study of ERP Systems Implementation Costs in Swiss SME’s. In: Cordeiro, J., Felipe, J (Hrsg.): *Proceedings of the 10th International Conference on Enterprise Information Systems (ICEIS) (2008-06-12/16)*. Barcelona, 2008, S. 143–148
- Ericsson: Maersk Line and Ericsson Bring Mobile Connectivity to the Oceans. Press release 2012-11-01. Stockholm: Ericsson Corporate Public & Media Relations, 2012. – Internetquelle. Abgerufen 2015-01-15 von: <http://hugin.info/1061/R/1576938/491546.pdf>
- European Commission: Regulation No 1907/2006 of the European Parliament and of the Council. In: *Official Journal of the European Union L 396 (2006)*, 49, S. 1–849
- Europäische Union: Charta der Grundrechte der Europäischen Union. In: *Amtsblatt der Europäischen Union C 83/389*, 2010
- Even, Adir; Shankaranarayanan, Ganesan: Utility-driven assessment of data quality. In: *ACM SIGMIS Database* 38 (2007), Nr. 2, S. 75–93. – DOI 10.1145/1240616.1240623
- Fleisch, Elgar: *What is the Internet of Things? An Economic Perspective*. St. Gallen/Zürich, Universität St. Gallen/ETH Zürich, Institut für Technologiemanagement, Auto-ID Labs, 2010. – White Paper. Auto-ID Labs White Paper WP-BIZAPP-053, 2010-01. Abgerufen 2015-02-08 von: <http://www.im.ethz.ch/education/HS10/AUTOIDLABS-WP-BIZAPP-53.pdf>
- Fleisch, Elgar; Österle, Hubert: *Auf dem Weg zum Echtzeit-Unternehmen*. In: Alt, R.; Österle, H. (Hrsg.): *Real-time Business*. Vol 1. Berlin: Springer. S. 3–17. – DOI 10.1007/97836421710861
- Fraunhofer IML (Fraunhofer Gesellschaft für Materialfluss und Logistik): *InBin – Der intelligente Behälter*. – Internetquelle. Abgerufen 2015-01-29 von: http://www.impl.fraunhofer.de/de/themengebiete/automation_eingebettete_systeme/Produkte/IntelligenterBehaelter.html
- Froidevaux, Yves: *Internet in den Schweizer Haushalten : Ergebnisse der Erhebung Omnibus IKT 2010*. Neuchâtel : Office fédéral de la statistique (OFS) (Bundesamt für Statistik (BFS)), 2012

- Hatz, Albert: BOSCH Master data Management. (6. Competence Center Corporate Data Quality Workshop (CC CDQ 1), 2008-01-16). St. Gallen, 2008. – Präsentation. Auf Anfrage von den Autoren verfügbar
- Holtham, Clive: Resolving the imbalance between information and technology. In: Best, D. (Hrsg.): *The Fourth Resource: Information and its Management*. Aldershot: Aslib/Gower, 1995, S. 41–58
- Horne, Nigel W.: Information as an Asset: The Board Agenda. In: *Computer Audit Update* 9 (1995), S. 5–11. – DOI 10.1016/0960259395902465
- IEEE Software Society: IEEE Standard for a Software Quality Metrics Methodology. IEEE Standard 1061TM-1998 (R2004). New York, 1998. – Reaffirmed Standard on 24 June 2004 by the IEEE-SA Standards Board and on 21 January 2005 by American National Standards Institute, ISBN 1-55937-529-9
- ISO/IEC: Information technology: Vocabulary Part 1: Fundamental terms (ISO/IEC 2382-1:1993). Genf: ISO/IEC, 1993
- ISO/IEC: ISO/IEC 38500: Corporate governance of information technology. Geneva: ISO/IEC, 2008
- ISO: ISO 8000-110:2009: Data quality, Part 110, Master data: Exchange of characteristic data: Syntax, semantic encoding, and conformance to data specification. Genf: ISO, 2009
- ITU (International Telecommunication Union): ICT Facts and Figures. The World in 2013. Geneva, February 2014. – Internetquelle. Abgerufen 2014-06-11 von: <http://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx>
- Johnson, Bobbie: Privacy no longer a social norm, says Facebook founder. *The Guardian*. – Internetquelle. Aktualisierungsdatum 2010-01-11. Abgerufen 2014-02-08 von <http://www.theguardian.com/technology/2010/jan/11/facebook-privacy>
- Kagermann, Henning: Industrie 4.0 und Smart Services. In: Brenner, W.; Hess, T. (Hrsg.): *Wirtschaftsinformatik in Wissenschaft und Praxis: Festschrift für Hubert Österle*. Berlin Heidelberg: Springer, 2014, S. 243–248
- Kokemüller, Jochen: *Stammdatenmanagementsysteme 2009: Eine Marktübersicht zu aktuellen Systemen*. Stuttgart: Fraunhofer-Verlag, 2009
- Levitin, Anany V.; Redman, Thomas C.: Data as a Resource: Properties, Implications, and Prescriptions. In: *Sloan Management Review* 40 (1998), Nr. 1, S. 89–101
- Möller, Henning: *Stammdatenqualität als Basis für das globale Kundenumsatz-Reporting der ZF Friedrichshafen AG*. St. Gallen, 2012. – Gastreferat in der Vorlesung „Enterprise Systems“ der Universität St. Gallen, 2012-05-07
- Muthreich, Karsten: *BIG DATA @ Nestlé* (37. Competence Center Corporate Data Quality Workshop (CC CDQ 4), 2013-12-11). Hamburg, 2011. – Präsentation. Auf Anfrage von den Autoren verfügbar
- Nonaka, Ikujiro: The Knowledge-Creating Company. In: *Harvard Business Review* 85 (2007), Nr. 7/8, S. 162–171
- O'Brien, Ralph: Is Traditional Security Dead? In: *ITNOW* 56 (2014), Nr. 1, S. 26–27. – DOI 10.1093/itnow/bwu012
- Ofner, Martin: *Datenqualität aus Prozessperspektive: Methoden und Modelle*. St. Gallen, Universität St. Gallen, Institut für Wirtschaftsinformatik, Dissertation, 2013
- OMG: *Semantics of Business Vocabulary and Business Rules (SBVR): V1.0*. Needham, MA: Object Management Group, 2008
- Oppenheim, Charles; Stenson, Joan; Wilson, Richard M.S.: Studies on Information as an Asset I: Definitions. In: *Journal of Information Science* 29 (2003), Nr. 3, S. 159–166. – DOI 10.1177/01655515030293003
- Österle, Hubert: Business oder Life Engineering? In: *HMD Praxis der Wirtschaftsinformatik* 51 (2014), Nr. 6, S. 744–761. – DOI 10.1365/s407020140097x

- Österle, Hubert; Otto, Boris: Consortium Research. In: *Business & Information Systems Engineering* 2 (2010), Nr. 2, S. 283–293. – DOI 10.1007/s115760100238y
- Österle, Hubert; Otto, Boris: Das datenzentrierte Unternehmen: Eine Business-Engineering-Perspektive. In: Schuh, G.; Stich, V. (Hrsg.): *Enterprise-Integration: Auf dem Weg zum kollaborativen Unternehmen*. Wiesbaden: Springer, S. 91–105 (2014)
- Österle, Hubert; Senger, Enrico: Prozessgestaltung und IT: Von der Unternehmens- zur Konsumentensicht. In: *Controlling & Management* 55 (2011), Nr. 2, S. 80–88. – DOI 10.1365/s1217601203363
- Österle, Hubert; Winter, Robert: Business Engineering. In: Österle, H.; Winter, R. (Hrsg.) *Business Engineering*. Vol 2. Berlin: Springer, 2003, S. 3–18
- Otto, Boris: Data Governance. In: *Business & Information Systems Engineering* 3 (2011a), Nr. 4, S. 241–244. – DOI 10.1007/s1259901101628
- Otto, Boris: Quality Management of Corporate Data Assets. In: Praeg, C.P.; Spath, D. (Hrsg.): *Quality Management for IT Services: Perspectives on Business and Process Performance*. Hershey, PA: IGI Global, 2011b, S. 193–209. – DOI 10.4018/978-1-61692-889-6
- Otto, Boris: Stammdatenqualität: Das Rückgrat moderner logistischer Systeme. In: Fraunhofer Institut für Materialfluss und Logistik IML Dortmund (Hrsg.): *Logistik entdecken* 15 (2014), S. 20–21
- Otto, Boris; Kokemüller, Jochen; Weisbecker, Anette; Gizanis, Dimitros: Stammdatenmanagement: Datenqualität für Geschäftsprozesse. In: *HMD – Praxis der Wirtschaftsinformatik* 48 (2011), Nr. 279, S. 5–16. – DOI 10.1007/BF03340582
- Otto, Boris; Abraham, Rene; Schlosser, Simon: Toward a Taxonomy of the Data Resource in the Networked Industry. In: *Proceedings of the 7th International Scientific Symposium on Logistics (ISSL 2014) (2014-06-4/5)*, Köln, 2014, S. 382–421
- Reid, R. Dan; Sanders, Nada R.: *Operations Management: An Integrated Approach*. Hoboken: John Wiley, 2005
- Reynolds, George W.: *Information Technology for Managers*. Boston, Ma: Course Technology, 2010
- Ross, Ronald G.; Lam, Gladys S.W.: *Building Business Solutions: Business Analysis with Business Rules*. Bissonnet: Business Rule Solutions LLC, 2011
- Salchegger, Markus; Dewor, Eva: Höchste Zeit für ein Solvency II Data Management. In: *Versicherungswirtschaft* 20 (2008), S. 1714–1718
- Schemm, Jan W.: Consumer-Centricity at Migros (Focus Group Workshop „Consumer-Centric Information Management (CCIM)“, 2012-06-13). St. Gallen, 2012. – Präsentation. Auf Anfrage von den Autoren verfügbar
- Schierning, Andreas: Consumer-Centric Information Management: Exploring Product Information (Business Engineering Forum, 2012-09-20). Bregenz, 2012. – Präsentation. Auf Anfrage von den Autoren verfügbar
- Schmid, Andreas: Bankdaten von 7500 Swisscom-Kunden sickern aus Leck. In: *Neue Zürcher Zeitung*, Ausg. 2014-02-15/16, S. 1 und S. 11
- Schweizer Bundesamt für Statistik: Internetnutzung Schweiz. – Internetquelle. Aktualisierungsdatum 2014-03-06. Abgerufen 2014-07-10 von: http://www.bfs.admin.ch/bfs/portal/de/index/themen/16/04/key/approche_globale.indicator.30106.301.html?open=1#1
- Sein, Maung K.; Henfridsson, Ola; Purao, Sandeep; Rossi, Matti; Lindgren, Rikard: Action Design Research. In: *MIS Quarterly* 35 (2011), Nr. 1, S. 37–56
- Shah, Rachna; Chandrasekaran, Aravind; Linderman, Kevin: In pursuit of implementation patterns: the context of Lean and Six Sigma. In: *International Journal of Production Research* 46 (2008), Nr. 23, S. 6679–6699. – DOI 10.1080/00207540802230504
- Smart Service Welt Working Group: *Smart Service Welt: Recommendations for the Strategic Initiative Web-based Services for Businesses*. Berlin: Acatech – National Academy of Science and Engineering, 2014

- Strong, Diane M.; Lee, Yang W.; Wang, Richard Y.: Data Quality in Context. In: Communications of the ACM 40 (1997), Nr. 5, S. 103–110. – DOI 10.1145/253769253804
- Van den Hoven, John: Information resource management: Stewards of data. In: Information Systems Management 16 (1999), Nr. 1, S. 88–91. – DOI 10.1201/1078/43187.16.1.19990101/31167.13
- Wahlster, Wolfgang: Vorarbeiten zu Industrie 4.0: Die Allianzen ADiWa und SemProM (Promotorengruppe KOMMUNIKATION der Forschungsunion (FU), Strategieworkshop „Industrie 4.0“, acatech Hauptstadtbüro, Berlin, 2011-06-27). Berlin, 2011. – Abgerufen 2015-01-29 von: http://www.res-com-project.org/tl_files/music_academy/Dokumente/Vorarbeiten_zur_Industrie_4_0_Die_Allianz_ADiWa_und_SemProM.pdf
- Wang, Richard Y.: A product perspective on total data quality management. In: Communications of the ACM 41 (1998), Nr. 2, S. 58–65. – DOI 10.1145/269012.269022
- Wang, Richard Y.; Lee, Yang W.; Pipino, Leo L.; Strong, Diane M.: Manage Your Information as a Product. In: Sloan Management Review 39 (1998), Nr. 4, S. 95–105
- Wang, Richard Y.; Strong, Diane M.: Beyond Accuracy: What Data Quality Means to Data Consumers. In: Journal of Management Information Systems 12 (1996), Nr. 4, S. 5–33
- White, Andrew: The Five Vectors of Complexity That Define Your MDM Strategy. Stamford, CT: Gartner Inc., 2010
- White, Andrew; Radcliffe, John: Four Dimensions of MDM: Understanding the Complexity. Stamford, CT: Gartner Inc., 2007
- White, Andrew; Radcliffe, John: Hype Cycle for Master Data Management. Stamford, CT: Gartner Inc., 2010
- Wrobel, Stefan; Voss, Hans; Köhler, Joachim; Beyer, Uwe; Auer, Sören: Big Data, Big Opportunities. In: Informatik-Spektrum (2014). – DOI 10.1007/s0028701408064
- Yoon, Vicotria Y.; Aiken, Peter G.; Guimaraes, Tor: Managing Organizational Data Resources: Quality Dimensions. In: Information Resources Management Journal 13 (2000), Nr. 3, S. 5–13. – DOI 10.4018/irmj.2000070101
- Ziegler, Patrick; Dittrich, Klaus R.: Data Integration: Problems, Approaches, and Perspectives. In: Krogstie, J.; Opdal, A.L.; Brinkkemper, S. (Hrsg.): Conceptual Modelling in Information Systems Engineering. Berlin: Springer, 2007, S. 39–58. – DOI 10.1007/978-3-540-72677-7_3