

I.2 Ursachen und Ausprägungen schlechter Datenqualität

Die Ausführungen in diesem Kapitel beziehen sich in erster Linie auf den in Kapitel I.1 eingeführten Begriff der anwenderorientierten Qualitätsvorstellung. Wie in diesem einführenden Kapitel dargelegt, basiert die Qualität von Daten auf den subjektiven Anforderungen der Anwender und hängt dabei im Wesentlichen von der Anforderung der Anwendung ab.

Im Umfeld von Business-Intelligence-Projekten ist diese Sichtweise sicherlich am besten geeignet, da sich die Qualität der Daten unmittelbar auf die in diesen Projekten entwickelten Anwendungen auswirkt. Dieses Kapitel beschreibt die Ursachen und Ausprägungen schlechter Datenqualität basierend auf Erfahrungen aus solchen Business-Intelligence-Projekten.

Business-Intelligence-Projekte und deren Ergebnisse initiieren in vielen Fällen den erstmaligen Kontakt der Anwender mit den Daten; den Anwendern ermöglicht dies oftmals eine komplett neue Sichtweise auf die Daten bzw. eine Verknüpfung verschiedener Bestände. Damit einher geht jedoch die Unsicherheit bezüglich der Datenqualität aufgrund fehlender Vergleichsmöglichkeiten und großer Datenvolumina. Das führt dazu, dass die Akzeptanz von Business-Intelligence-Projekten massiv unter schlechter Datenqualität bzw. unter subjektiv wahrgenommenen Datenqualitätsproblemen leidet.

Business-Intelligence-Anwendungen sind meist das letzte Glied in der Kette der Datenaufbereitung und -bereitstellung. Aus diesem Grund scheitern Business-Intelligence-Projekte oftmals wegen fehlender Akzeptanz. Daher ist es für eine erfolgreiche Umsetzung von Projekten unumgänglich, bei den potenziellen Anwendern ein Verständnis für schlechte Datenqualität aufzubauen. Die Capgemini-Studie „IT Trends 2009“⁴⁰ zeigt, dass das Thema Datenqualität im Umfeld von Business-Intelligence-Projekten immer wichtiger wird. Im Rahmen dieser Studie wurde Datenqualität von den befragten Unternehmen bereits das dritte Jahr in Folge als das mit Abstand wichtigste Thema in Bezug auf Business Intelligence genannt.

Dies beinhaltet eine Sensibilisierung für die Ursachen und Ausprägungen schlechter Datenqualität und den Umgang damit. Diese Problematik muss bereits im Vorfeld von Projekten einbezogen werden und es sind Maßnahmen aufzuzeigen, um negativen Auswirkungen entgegenzusteuern.

Dabei sind die Ursachen schlechter Datenqualität vielfältig, und oftmals ist deren Vermeidung auch unter größtmöglichen Anstrengungen nicht immer durchgängig möglich. Daher ist eine dahingehende Sensibilisierung der Anwender ein wichtiger Baustein für die erfolgreiche Umsetzung von Business-Intelligence-Projekten.

I.2.1 Geschäftstreiber

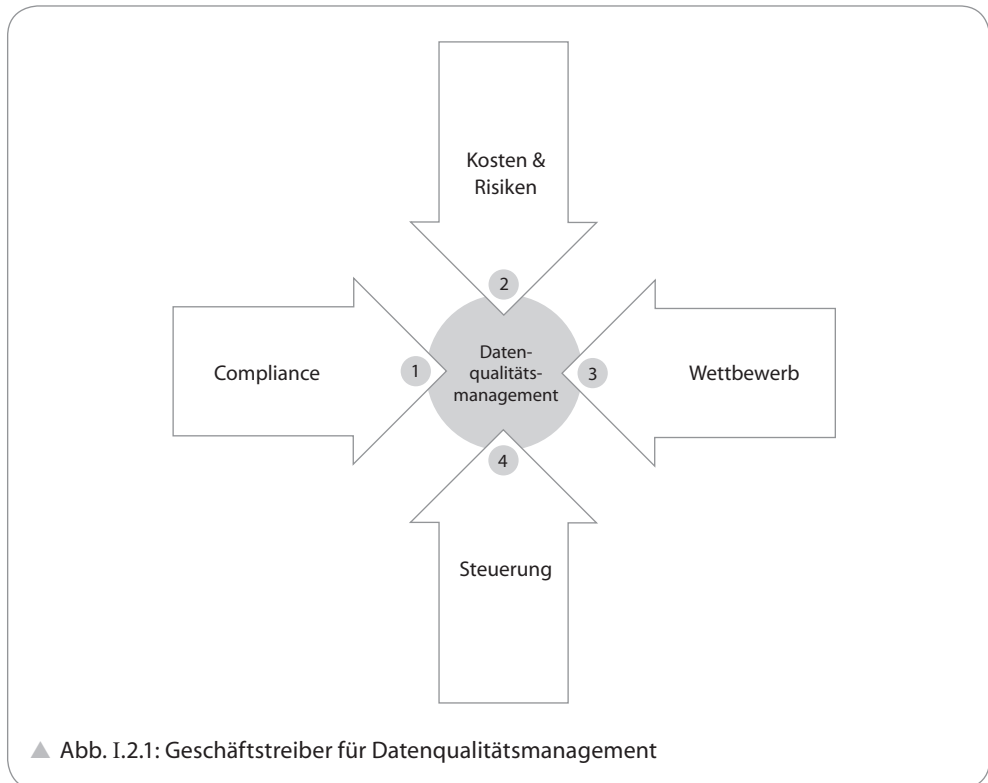
Schlechte Datenqualität hat Auswirkungen auf viele unterschiedliche Bereiche. Eine Verbesserung der Datenqualität ist deshalb generell ein wünschenswertes Unterfangen, das prinzipiell auf wenig Widerstand in den Unternehmen trifft. In der Praxis ist die Sicherung der Datenqualität jedoch eine komplexe Angelegenheit. Budget- und Ressourcenbeschränkungen

⁴⁰ Vgl. Capgemini-Studie „IT Trends 2009“, Rubrik Business Intelligence.

sind in den meisten Unternehmen die Realität. Daher sind in jedem Fall Kosten-Nutzen-Betrachtungen anzustellen, um eine Argumentation hinsichtlich der Mittelbereitstellung aufzubauen.

Diese Betrachtungen können sowohl quantitativer als auch qualitativer Natur sein (siehe Kapitel I.3). In diesem Abschnitt sind im Folgenden einige Geschäftstreiber dargestellt, die ein aktives Management der Datenqualität erfordern. Diese Geschäftstreiber geben oftmals den initialen Anstoß zur Verbesserung mangelhafter Datenqualität. Zudem erfordern die zunehmende Integration von Geschäftsprozessen und Anwendungen über Abteilungsgrenzen hinweg sowie Ansätze wie Serviceorientierte Architekturen (SOA) und Right Time Business Intelligence saubere Daten. Während einer Right-Time-Datenintegration bleibt schlichtweg keine Zeit mehr für die Durchführung umfassender Datenqualitätsmaßnahmen.

Abbildung I.2.1 stellt vier wesentliche Geschäftstreiber dar, die eine der Anwendung⁴¹ entsprechende, akzeptable Datenqualität speziell im Rahmen der Datenbereitstellung durch Business-Intelligence-Anwendungen voraussetzen.



Diese vier Geschäftstreiber sind grundsätzlich unabhängig von bestimmten Branchen und hängen oftmals mit Business-Intelligence-Projekten bzw. mit der Datenbereitstellung und Datenverwendung generell zusammen. Aus den Anforderungen dieser Geschäftstreiber heraus entsteht in weiterer Folge oft die Forderung nach einem nachhaltigen Datenqualitätsmanagement.

⁴¹ Die Begriffe „Anwendung“ und „System“ werden synonym verwendet.

I.2.1.1 Compliance

Externe regulatorische Vorgaben sowie die interne Revision erfordern die Definition und die Umsetzung einheitlicher Standards und Mechanismen für ein aktives Datenqualitätsmanagement. Zielsetzung ist die Etablierung unternehmensweiter Strukturen zur Sicherstellung der Datenqualität bzw. der standardisierte Umgang mit schlechter Datenqualität und auftretenden Datenqualitätsproblemen.

Regulatorische Vorgaben sind zum Beispiel der Sarbanes-Oxley Act, die International Financial Reporting Standards sowie Basel-II-Vorgaben für Finanzdienstleistungen. Es ist zudem aufgrund der jüngsten Turbulenzen auf den Finanzmärkten zu erwarten, dass weitere regulatorische Anforderungen auf die Unternehmen zukommen.

I.2.1.2 Kosten und Risiken

Das Erkennen und die frühzeitige Behebung von Datenqualitätsproblemen sowie ein laufendes aktives Datenqualitätsmanagement reduzieren das tägliche operative Risiko in den Unternehmen.

Das aktive Beheben schlechter Datenqualität mindert generell die Projektrisiken und führt zu geringeren Kosten bzw. zu reduzierten Durchlaufzeiten von Projekten (insbesondere auch von Business-Intelligence-Projekten). Besonders betroffen sind dabei die Entwicklungen von Schnittstellen und Datenmigrationen zwischen Anwendungen. Dies ist insbesondere relevant, da die Forderung nach immer agileren und schnelleren Projekten keine Zeit für projektinterne Datenqualitätsmaßnahmen mehr lässt.

I.2.1.3 Wettbewerb

Im Wettbewerb um den Kunden ist eine dem Verwendungszweck entsprechende Datenqualität Voraussetzung, z. B. für den Erfolg von Kampagnen sowie bei der effizienten Entwicklung neuer Produkte. Für eine Verbesserung der Kundenbeziehung ist eine hohe Qualität der Daten ebenso förderlich (zum Beispiel durch eine korrekte Anrede).

I.2.1.4 Steuerung

Die für die Unternehmenssteuerung verantwortlichen Personen müssen sich auf die Qualität der Daten verlassen können, da diese eine unmittelbare Auswirkung auf die Qualität der Entscheidungen im Unternehmen hat. Grundsätzlich hängt die Akzeptanz durch die Anwender von Business-Intelligence-Anwendungen in großem Ausmaß von der Qualität der bereitgestellten Daten ab.

I.2.1.5 Zusammenfassung

Wie einleitend bereits kurz erwähnt, ist ein wesentlicher Punkt, dass die Wahrnehmung der Qualität der Daten durch den Anwender weitgehend subjektiv ist. Daten können für verschiedene Anwender unterschiedliche Qualität besitzen. Dies kann unter Umständen bedeuten, dass eine für einen Anwender (z. B. aus dem externen Rechnungswesen/Wirtschaftsprüfer) schlechte Datenqualität eine ausreichende bis gute Datenqualität für einen anderen Anwender (zum Beispiel aus dem Vertrieb für Kampagnenmanagement) darstellt, und vice versa.

Diese Unterscheidung ist gerade im Umfeld von Business-Intelligence-Projekten von außerordentlicher Wichtigkeit. Eine 100-prozentige Qualität der Daten wird praktisch nicht erreichbar sein. Der Grad der benötigten Qualität ist daher immer in Abhängigkeit von der beabsichtigten Anwendung zu evaluieren.

I.2.2 Ausprägungen schlechter Datenqualität

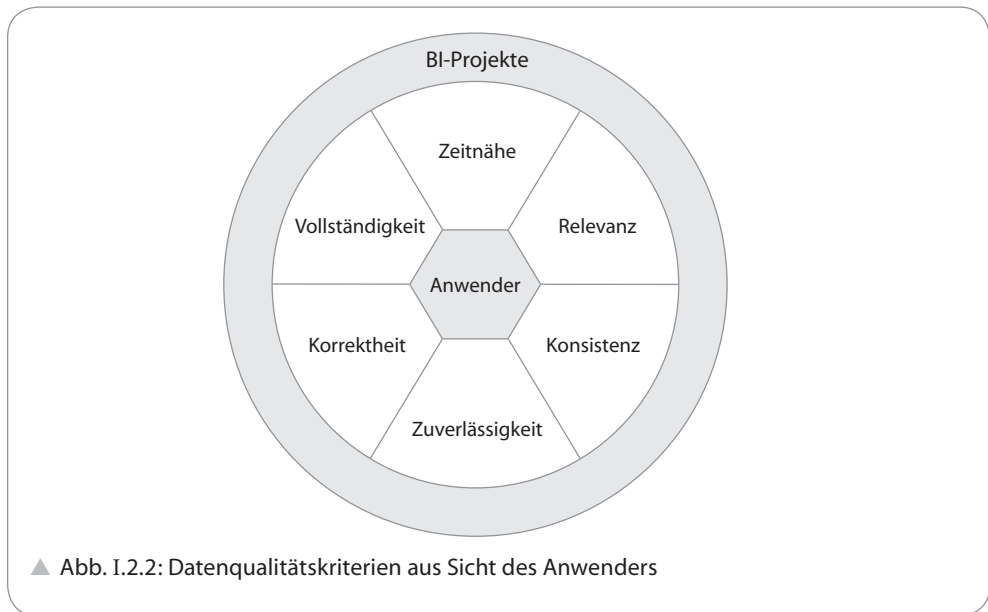
Die oben dargestellten Geschäftstreiber erfordern jeweils eine bestimmte Qualität der Daten. Die Ausprägungen von als schlecht wahrgenommener Datenqualität sind grundsätzlich sehr vielfältig und orientieren sich im Wesentlichen am subjektiven Empfinden der Anwender.

Die Ausprägungen schlechter Datenqualität, wie sie in diesem Abschnitt zur Darstellung gelangen, lassen sich anhand der bereits in Kapitel I.1 definierten Datenqualitätskriterien festmachen.

Die hier ausgewählten Ausprägungen schlechter Datenqualität bzw. deren Auswirkungen in der realen Welt sind insbesondere im Rahmen verschiedener Business-Intelligence-Projekte sehr stark von Anwendern reklamiert worden.

Der Anspruch dieses Abschnitts ist keine generelle Definition und Auflistung aller möglichen Ausprägungen schlechter Datenqualität. Der Fokus liegt auf Business-Intelligence-Projekten sowie der Wahrnehmung schlechter Datenqualität durch die Anwender.

Abbildung I.2.2 beschränkt sich daher auf die wesentlichsten Ausprägungen schlechter Datenqualität, dargestellt mittels sechs ausgewählter Datenqualitätskriterien.



Im Folgenden sollen beispielhaft die Inhalte dieser sechs Datenqualitätskriterien beschrieben werden, mit der Zielsetzung, ein Verständnis für die wesentlichen Ausprägungen schlechter Datenqualität im Umfeld von Business-Intelligence-Projekten zu schaffen.

I.2.2.1 Zeitnähe

Die Zeitnähe hat aus Sicht der Anwender von Business-Intelligence-Anwendungen grundsätzlich zwei Ausprägungen, welche die subjektiven Anforderungen an die Qualität der Daten beeinflussen.

Einerseits ist dies der Anlieferungszeitpunkt der Daten aus den Quellsystemen bzw. der Zeitpunkt der Verfügbarkeit der Daten für den Anwender in den Business-Intelligence-Anwendungen. Daten besitzen aus Sicht der Anwendung dann eine schlechte Qualität, wenn sie nicht zu dem vom Anwender benötigten Zeitpunkt zur Verfügung stehen. Dieses Problem potenziert sich mit der Verknüpfung von Daten verschiedener Bereitstellungszeitpunkte.

Die zweite Ausprägung betrifft die Aktualität der angelieferten Dateninhalte. Darunter fällt zum Beispiel die Aktualität der Pflege von Adressdaten. Ein anderes Beispiel aus dem Einsatzfeld von Banken ist die regelmäßige Neubewertung von Sicherheiten für Kredite.

I.2.2.2 Korrektheit

Das Datenqualitätskriterium Korrektheit beschreibt die inhaltliche und formale Komponente der Daten. Korrekte Daten beinhalten die inhaltlich richtigen Informationen in den vordefinierten Formaten der Attribute. Somit müssen zum Beispiel die Geburtsdaten von Personen nicht nur inhaltlich richtig sein, sondern die Anlieferung hat auch im vordefinierten Datenformat zu erfolgen. Die bereitgestellten Informationen in den Anwendungen bilden somit die reale Welt im vordefinierten Format ab.

I.2.2.3 Konsistenz

Die Konsistenz bzw. die Inkonsistenz von Daten stellt gerade im Umfeld von Business-Intelligence-Projekten oftmals ein massives Problem hinsichtlich der Wahrnehmung schlechter Datenqualität durch den Anwender dar. Die Inkonsistenz selbst kann dabei praktisch an jeder Stelle im Rahmen der Datenverarbeitung, beginnend bei den Quellsystemen, entstanden sein.

Konsistenz in diesem Zusammenhang bezieht sich dabei in erster Linie auf die Übereinstimmung bzw. auf die vorhandene Möglichkeit zum Abgleich von Daten aus verschiedenen analytischen und operativen Anwendungen. Dieses Kriterium zielt auf die Widerspruchsfreiheit der Daten aus unterschiedlichen Anwendungen. Inkonsistente Daten führen zu inkonsistenten Berichten und zu einem Glaubwürdigkeitsproblem für das gesamte Business-Intelligence-Projekt.

I.2.2.4 Zuverlässigkeit

Das Kriterium Zuverlässigkeit beschreibt die ausreichende Verfügbarkeit von Informationen, um die ursprüngliche Herkunft und die verschiedenen Transformationen der Daten rückverfolgen und nachvollziehen zu können. Dies beginnt bereits bei der initialen Erfassung der Daten und geht hin bis zur Erstellung der Berichte aus den BI-Anwendungen⁴² heraus.

Zur Vertrauensbildung trägt die sogenannte Reconciliation bei. Reconciliation generell beschreibt den Abgleich von Daten aus unterschiedlichen Datenquellen. Im Speziellen ist mit diesem Begriff oftmals der Abgleich von Detaildaten aus dem Data Warehouse mit den Daten aus dem Hauptbuch des Buchhaltungssystems gemeint.

⁴² Die Begriffe „BI-Anwendung“ und „analytische Anwendung“ werden synonym verwendet.

I.2.2.5 Relevanz

Das Datenqualitätskriterium Relevanz beschreibt, inwieweit das Informationsangebot der Anwendungen mit dem Informationsbedarf der Anwender übereinstimmt. Eine 100-prozentige Qualität der bereitgestellten Daten ist aus Sicht der Anwender unzureichend, wenn die bereitgestellten Daten generell nicht benötigt werden oder andere Daten vonnöten sind.

Die Relevanz beschreibt somit die generelle Verfügbarkeit von Inhalten, welche der Anwender benötigt, um seine Aufgaben vollständig erfüllen zu können. Dies führt zu der Frage, ob z. B. alle wesentlichen Datenbereiche in den Business-Intelligence-Systemen integriert sind, um die Anforderungen der Anwender erfüllen zu können.

I.2.2.6 Vollständigkeit

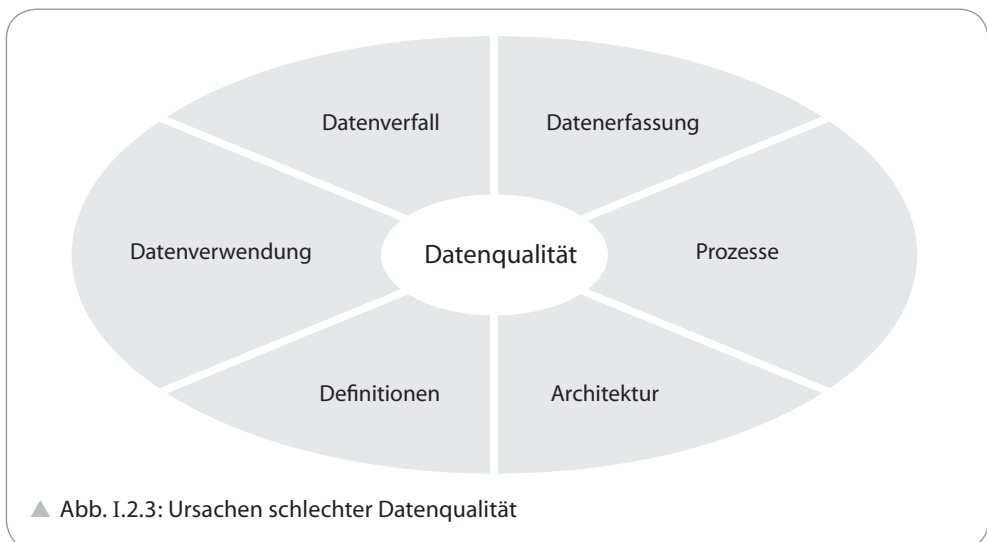
Die Vollständigkeit der Daten hat wiederum zwei Ausprägungen. Einerseits kennzeichnet dieses Kriterium die Vollständigkeit der übermittelten Daten zwischen den Anwendungen, das bedeutet, es sind im Zuge der Transformationen keine Daten verloren gegangen.

Andererseits bezeichnet das Datenqualitätskriterium Vollständigkeit die vollständige Befüllung der angelieferten Attribute mit Daten, d.h. die einzelnen Attribute beinhalten keine NULL-Werte.

I.2.3 Ursachen schlechter Datenqualität

Ebenso wie die Ausprägungen sind auch die Ursachen schlechter Datenqualität grundsätzlich sehr vielfältig. Gerade im Rahmen von Business-Intelligence-Projekten ist es wesentlich, die Ursachen zu verstehen und rechtzeitig zu adressieren. Davon ist in vielen Fällen der letztendliche Erfolg dieser Projekte abhängig.

Abbildung I.2.3 gibt in sechs Kategorien zusammengefasst einen ersten Überblick über die Ursachen schlechter Datenqualität, deren Auswirkungen insbesondere in Business-Intelligence-Projekten oder -Anwendungen zutage treten.



I.2.3.1 Datenerfassung

Die initiale Erfassung der Daten ist eine der wesentlichsten Quellen für das Auftreten schlechter Datenqualität. Die verschiedenen Ursachen für Fehler bei der Erfassung der Daten sind vielfältig. Dies beginnt bei einfachen Benutzerfehlern und geht hin bis zu fehlenden Ressourcen, um eine korrekte Datenerfassung in unterschiedlichen, meist hochkomplexen Anwendungen garantieren zu können.

Benutzerfehler unterteilen sich zum Beispiel in falsche Eingaben, fehlendes Know-how für eine korrekte Bedienung der Anwendungen, falsche Interpretationen der zu erfassenden Daten sowie die unvollständige Erfassung der Daten aufgrund fehlender Informationen.

Eine der grundlegenden Ursachen für das Entstehen schlechter Datenqualität bereits bei der Datenerfassung ist das fehlende Problembewusstsein für die Auswirkungen mangelhafter Datenqualität sowie die fehlende Incentivierung für eine korrekte Erfassung von Daten.

Es wird oftmals ausschließlich die Anzahl der erfassten Daten und die Geschwindigkeit der Erfassung gemessen, ohne auf die Qualität zu achten. Dies führt in weiterer Folge dazu, dass keine Zeit für eventuell notwendige Rückfragen bei Unklarheiten im Rahmen der Erfassung zur Verfügung steht.

Mangelhaftes Design der Eingabemasken stellt eine weitere Quelle für Probleme bei der Erfassung der Daten dar. Die Absicherung einzelner Felder ist nicht immer ausreichend, die Möglichkeiten zur Einschränkung der Erfassung in den Masken bleiben ungenutzt. Zudem erfolgt das Design solcher Masken meist aus Sicht eines einzigen Unternehmensbereiches, etwa durch den Vertrieb.

Dies hat zur Folge, dass sich z. B. abgesicherte Muss-Felder in den Eingabemasken vor allem an den Bedürfnissen des Vertriebs orientieren und vorhandene Kann-Felder nicht weiter eingeschränkt werden. Aus Sicht des externen Rechnungswesens haben diese Kann-Felder möglicherweise eine ganz andere Bedeutung und führen in weiterer Folge zu Datenqualitätsproblemen im Rahmen der Bilanzerstellung.

Ein potenzieller Ressourcenmangel im Zuge der Datenerfassung zieht in vielen Fällen ebenfalls eine mangelhafte Datenqualität nach sich, falls beispielsweise über längere Zeiträume keine bzw. nur eine eingeschränkte Datenerfassung erfolgt. Dies hat Auswirkungen auf die Zeitnähe der Daten in den verschiedenen Business-Intelligence-Anwendungen. Eine unzureichend aktuelle Datenerfassung hat unter Umständen falsche Meldungen an die Aufsichtsbehörden zur Folge oder kann zu Irritationen in den Kundenbeziehungen führen.

Fehlerhafte externe Daten sind ebenfalls oftmals einer der Gründe für schlechte Datenqualität in den Business-Intelligence-Anwendungen. Externe Daten können dabei z. B. zugekaufte Adressdaten sein.

Die Schlussfolgerung daraus ist, dass bereits bei der Erfassung der Daten die Bedürfnisse des Gesamtunternehmens in die Überlegungen zur Sicherstellung der Datenqualität einfließen müssen, um so aus einer ganzheitlichen Sichtweise heraus potenzielle Maßnahmen definieren zu können.

I.2.3.2 Prozesse

Diese Kategorie beinhaltet die Auswirkungen auf die Datenqualität, die sich durch die Komplexität der Abläufe in den Unternehmen bzw. durch generell mangelhaft aufgesetzte Prozesse im Zuge der Datenverarbeitung, ausgehend von der Erfassung der Daten bis hin zur Erstellung der Berichte, ergeben.

Fehlende oder unvollständig aufgesetzte Prozesse führen zu fehlerhaften Datenerfassungen. Fehlende Prüfprozesse im Sinne einer weiteren Absicherung hinsichtlich Falscheingaben verstärken dies zusätzlich.

Ergänzend zu den Erfassungsprozessen sind fehlende Feedbackprozesse ebenfalls problematisch, da Möglichkeiten zur standardisierten Revidierung falscher Eingaben fehlen.

Verstärkt werden diese Probleme zusätzlich noch durch eventuelle Brüche in den verwendeten Medien, z. B. durch eine Unterbrechung der automatisierten Datenverarbeitung durch manuelle Ergänzungen oder auch durch eine teilweise manuelle Erfassung innerhalb der Kette von der Datenerfassung bis zur Datenverwendung.

Abteilungsübergreifende Prozesse mit einem Übergang der Verantwortung im Rahmen der Datenbereitstellung führen ebenfalls zu Abstimmungsproblemen und ziehen potenziell eine Verschlechterung der Datenqualität nach sich.

Überschreitungen von Unternehmensgrenzen, wie sie beispielsweise durch ein Outsourcing der Datenerfassung hervorgerufen wurden, stellen eine besondere Herausforderung an die Sicherstellung der Datenqualität dar. Diese Konstrukte sind meist ausschließlich auf Kostenersparnis und die Maximierung der erfassten Datenmengen bzw. Minimierung der Erfassungszeiten aufgebaut. Die für eine Sicherstellung der Datenqualität erforderlichen Ressourcen und deren zeitliche Verfügbarkeit spielen dabei häufig nur eine sekundäre Rolle.

Fehlende, standardisierte Abstimmungsprozesse zwischen den Daten auf verschiedenen Stufen der Verarbeitungskette führen dazu, dass die Erkennung potenzieller Fehler erst sehr spät stattfindet. Daraus resultiert eine Verkürzung der Reaktionszeiten bei auftretenden Datenqualitätsproblemen. Eine in den Verarbeitungsprozess integrierte Reconciliation im Sinne eines Abgleichs des Data Warehouse mit den Buchhaltungssystemen ermöglicht z. B. ein frühzeitiges Reagieren auf Abweichungen in wesentlichen Bereichen.

I.2.3.3 Architektur

Das Design der Datenarchitektur eines Unternehmens ist eine wesentliche Aufgabe zur Sicherstellung der Datenqualität. Unter den Begriff Datenarchitektur fallen hier neben der Vielfalt der verwendeten Technologien und der bereitgestellten Daten auch deren Zugänglichkeit bzw. Unzugänglichkeit durch die Anwender.

Datenarchitektur beginnt schon bei den operativen Systemen, die als Datenquellen für die durchzuführenden Aufgaben ausgelegt sein müssen. Eine konsistente Konzeption der Datenintegration der BI-Anwendungslandschaft beginnt bereits bei den ETL⁴³-Prozessen bzw. den generellen Datenflüssen zwischen diesen Systemen. Inhaltlich klar abgegrenzte analytische

⁴³ ETL = Extraktion, Transformation, Laden.

Data Marts sind dabei ebenso eine wesentliche Voraussetzung für eine konsistente Datenarchitektur von Unternehmen.

Das Thema Datenarchitektur im Umfeld von Business-Intelligence-Projekten wird in Kapitel I.5 im Detail behandelt. In diesem Kapitel erfolgt zunächst eine erste Sensibilisierung hinsichtlich der Datenarchitektur.

In vielen Unternehmen sind in den vergangenen Jahren die operativen Anwendungslandschaften kontinuierlich gewachsen und um zusätzliche Funktionalitäten erweitert worden, die ursprünglich nicht vorgesehen waren. Das Ergebnis dieser nachträglichen Modifikationen ist oftmals eine heterogene operative Anwendungslandschaft mit redundanten, inkonsistenten Datenhaltungen sowie redundanten Transformationen zwischen den verschiedenen Anwendungen.

Systemumstellungen, Datenmigrationen und Fehler in den Anwendungen selbst führen bereits in den operativen Systemen zu einer Verschlechterung der Datenqualität. Die auf dieser Stufe entstehenden Probleme vervielfachen sich im Zuge der Datenverarbeitung bis hin in die BI-Anwendungen und führen zu einem erhöhten Aufwand für die Eindämmung der Auswirkungen.

Innerhalb der BI-Anwendungslandschaft sind die Problembereiche analog den operativen Anwendungen. Dazu kommen noch inkonsistente Transformationen der Daten aus den operativen Systemen in die verschiedenen analytischen Systeme. Dies zieht Inkonsistenzen in den Berichten und einen erhöhten Abstimmungsaufwand zwischen den Anwendungen nach sich.

Zusammenfassend führen Datenflüsse zwischen verschiedenen Anwendungen unweigerlich zu Problemen, hervorgerufen durch unterschiedliche Felddefinitionen und mehrfache Transformation der Daten. Im Zuge der Betrachtung von Datenqualitätsproblemen ist daher immer ein Augenmerk auf die Datenflüsse zwischen den Anwendungen zu legen. Das Design der Datenarchitektur sowie die Nachvollziehbarkeit von Transformationen in den Datenflüssen sind wesentliche Voraussetzungen für die nachhaltige Sicherstellung der Datenqualität.

I.2.3.4 Definitionen

Ein weiterer wesentlicher Punkt hinsichtlich schlechter Datenqualität sind inkonsistente fachliche Beschreibungen der Daten. Dies beinhaltet unterschiedliche Definitionen von Feldformaten, verschiedene Inhalte von grundsätzlich identischen Feldern in verschiedenen Anwendungen sowie oftmals unvereinbare Sichtweisen auf das Unternehmen durch die einzelnen Unternehmensbereiche.

Bedingt durch heterogene Anwendungslandschaften verstärken sich die Auswirkungen dieser unterschiedlichen Definitionen auf die Qualität der Daten mit jeder weiteren Transformation zwischen den Anwendungen. Diese Kategorie hat hinsichtlich einer konsistenten Business-Intelligence-Architektur höchste Relevanz. Gerade spezifische Data Marts in den einzelnen Fachbereichen der Unternehmen bergen die Gefahr inkonsistenter Definitionen in sich.

Die Konzeption vieler Anwendungen erfolgt aus Sicht eines einzelnen Unternehmensbereiches, ohne Berücksichtigung der Bedürfnisse des Gesamtunternehmens. Ein wesentlicher Grund dafür ist oftmals die beschränkte Verfügbarkeit von Ressourcen, aber auch fehlendes Problem-

bewusstsein oder im schlimmsten Fall einfach nur Ignoranz. Dies führt in weiterer Folge unweigerlich zu erhöhten Abstimmungserfordernissen aufgrund inkonsistenter Berichte.

Unterschiedliche Sichtweisen auf ein Unternehmen sind zum Beispiel voneinander abweichende Definitionen der Produktkataloge eines Unternehmens oder auch inkonsistente Definitionen der Kundensegmentierungen bzw. eine fehlende einheitliche Definition des Umsatzbegriffs.

Viele subjektiv durch die Anwender wahrgenommene Datenqualitätsprobleme lassen sich auf mangelhafte Kommunikation innerhalb der Unternehmen zurückführen. Oftmals ist es schon ausreichend, entsprechende abteilungsübergreifende Kommunikationsforen in den Unternehmen einzurichten, um eine Reihe von Datenqualitätsproblemen auszumerzen.

Auf organisatorische Aspekte des Datenqualitätsmanagements wird detailliert in Kapitel I.4 eingegangen.

I.2.3.5 Datenverwendung

Die inkorrekte Nutzung von Anwendungen führt bei den Empfängern von Daten aus diesen Anwendungen ebenfalls zum subjektiven Empfinden einer schlechten Datenqualität. Die inkorrekte Nutzung durch den Anwender äußert sich dabei vornehmlich in zwei verschiedenen Ausprägungen.

Einerseits ist dies die inkonsistente Durchführung von Datenkorrekturen in verschiedenen Applikationen bzw. in unterschiedlichen analytischen Anwendungen. Dazu kommt, dass diese Korrekturen in den seltensten Fällen in den operativen Systemen nachgezogen werden. Generell sind Datenkorrekturen immer in den Quellenanwendungen durchzuführen. In Ausnahmefällen, z.B. im Zuge der Bilanzerstellung, kann es aus zeitlichen Gründen vorkommen, dass Korrekturen zwischenzeitlich auch innerhalb der BI-Anwendungslandschaft durchgeführt werden müssen. Dies muss in jedem Fall nach vordefinierten Regeln sowie nach Abstimmung mit der internen Revision stattfinden.

Andererseits fällt darunter auch das fehlende Know-how für die Interpretation der Daten oder auch das mangelnde Wissen um eine korrekte Erstellung von Berichten.

I.2.3.6 Datenverfall

Das Problem des Datenverfalls tritt speziell in Business-Intelligence-Anwendungen verstärkt auf, da diese Anwendungen für viele Anwender oftmals die einzigen Datenquellen darstellen.

Im Laufe der Zeit kommt es zu einem unweigerlichen Verfall mancher Daten. Beispiele dafür sind in erster Linie Adress- und Telefondaten, die sich im Verlauf der Zeit naturgemäß ändern. Ein weiteres Beispiel für den Verfall von Daten ist die Zugehörigkeit von Personen zu bestimmten Berufsgruppen, die sich ebenfalls regelmäßig ändern kann.

Der Datenverfall führt unweigerlich zur Ungültigkeit von Daten und somit in weiterer Folge zu einem Datenqualitätsproblem. Es ist daher notwendig, entsprechende Prozesse aufzusetzen, um einerseits gefährdete Daten zu erkennen und andererseits diese Daten laufend aktuell zu halten. Der Datenverfall wird durch das Datenqualitätskriterium Zeitnähe abgebildet.

I.2.3.7 Fazit

Zusammenfassend sind das Fehlen klar definierter Verantwortungen und im Falle mangelhafter Datenqualität zu ergreifender Maßnahmen neben inkonsistenten Datenarchitekturen die wesentlichsten Ursachen für schlechte Datenqualität. Daher sind in jedem Business-Intelligence-Projekt die Kette der Datenverarbeitung und der zugehörigen Verantwortungen zu untersuchen und mögliche Probleme proaktiv aufzuzeigen.

Weiterhin ist über alle Kategorien hinweg das Thema der nur in den seltensten Fällen ausreichend vorhandenen Ressourcen und Budgetmittel relevant. Oftmals sind die Datenqualitätsprobleme im Unternehmen bekannt, allerdings werden keine Maßnahmen zu deren Bereinigung ergriffen.

Dies ist in vielen Fällen auf ein fehlendes Bewusstsein für die Auswirkungen schlechter Datenqualität bei den Entscheidungsträgern zurückzuführen. Die Kosten-Nutzen-Darstellung von Maßnahmen zur Sicherung der Datenqualität lässt sich zudem nicht immer ausreichend quantitativ belegen (siehe Kapitel I.3).

I.2.4 Beispiel: Finanzdienstleister

Abschließend soll beispielhaft ein Business-Intelligence-Projekt angeführt werden, das bei einem Finanzdienstleister durchgeführt wurde. Im Zuge dieses Projektes erfolgten eine Erhebung von Datenqualitätsproblemen sowie deren Kategorisierung nach Ursachen.

Die Ergebnisse dieser Erhebung wurden verwendet, um in weiterer Folge die Auswirkungen der Datenqualität auf das Business-Intelligence-Projekt frühzeitig abschätzen und darauf aufbauend entsprechende Maßnahmen planen zu können.

I.2.4.1 Ausgangssituation

Ausgangssituation in diesem Projekt war die Implementierung eines unternehmensweiten Data Warehouse und eine damit einhergehende Konsolidierung der über das gesamte Unternehmen verteilten analytischen Data Marts.

Die Erhebung der Datenqualitätsprobleme war in diesem Zusammenhang ein Begleitprojekt mit zwei Zielsetzungen: einerseits die Reduktion der Projektrisiken und andererseits die Erstellung eines Maßnahmenkatalogs, um den anschließenden regulären Betrieb der Anwendung auf qualitativ hohem Niveau sicherstellen zu können.

I.2.4.2 Vorgehensweise

Im Rahmen des Projekts erfolgte eine qualitative Erhebung mittels persönlicher Interviews hinsichtlich subjektiv wahrgenommener Datenqualitätsprobleme und eine Analyse der Ursachen. Dabei wurden die künftigen potenziellen Anwender im gesamten Unternehmen befragt.

Die so erhobenen Probleme wurden konsolidiert und zu sechs Kategorien zusammengefasst. Die Kategorien entsprechen dabei den in Abschnitt I.2.3 dargestellten Ursachen für schlechte Datenqualität. In Summe wurden insgesamt circa einhundert subjektive Einzelprobleme kategorisiert.

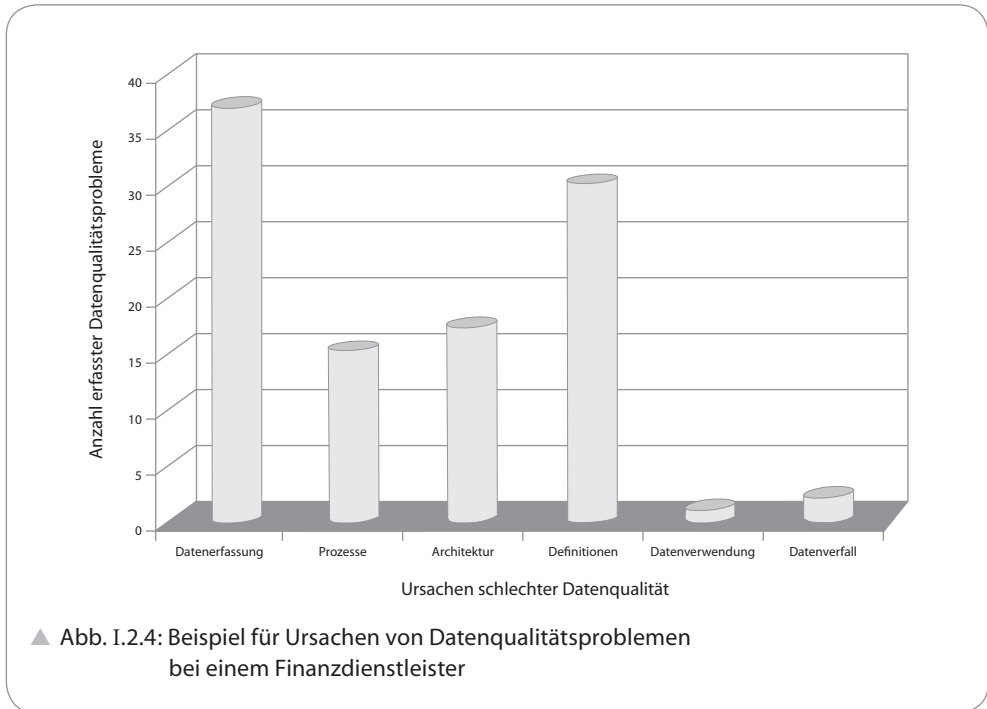
I.2.4.3 Ergebnis

Abbildung I.2.4 stellt die Ergebnisse dieser subjektiven Befragung und der anschließenden Kategorisierung der Datenqualitätsprobleme grafisch dar.

Auf der x-Achse sind die sechs Problemkategorien bzw. die Ursachen schlechter Datenqualität aufgetragen, auf der y-Achse die Anzahl der einzelnen genannten Datenqualitätsprobleme je Ursache.

Aus diesem Beispiel ist klar ersichtlich, dass die Erfassung der Daten aus Sicht der Anwender die mit Abstand größten Probleme verursacht. Dies war durchaus zu erwarten. Etwas überraschend in diesem Beispiel mag erscheinen, dass die fehlenden oder falschen Definitionen bereits an zweiter Stelle der wahrgenommenen Probleme stehen, während die Prozesse und die Architektur erst mit etwas Abstand folgen.

Die Probleme, die durch die Datenverwendung und den zeitlichen Verfall der Daten hervorgerufen werden, sind aus Sicht der Anwender in diesem Unternehmen anzahlmäßig nur von geringer Bedeutung.



Die konkreten Auswirkungen der einzelnen Datenqualitätsprobleme auf die Compliance mit regulatorischen Anforderungen oder auf andere Risiken sind im Einzelfall unabhängig von der Anzahl der Probleme je Kategorie zu betrachten. Eventuell notwendige Maßnahmen müssen einzeln analysiert, entschieden, mit Ressourcen versehen und umgesetzt werden.

I.2.4.4 Erkenntnisse

Wie oben erwähnt, ist aus Abbildung I.2.4 ersichtlich, dass der Großteil der Datenqualitätsprobleme bereits aus der Erfassung der Daten resultiert. Dies war durchaus zu erwarten und zeigt, dass Maßnahmen zur Sicherstellung der Datenqualität möglichst frühzeitig aufzusetzen sind.

Die hohe Anzahl von Problemen in der Kategorie Definitionen lässt sich im Wesentlichen auf die Tatsache zurückführen, dass die Ursache der subjektiven Wahrnehmung schlechter Datenqualität in einem Unternehmensbereich oftmals eine bewusste Entscheidung eines anderen Unternehmensbereiches war.

Das sogenannte Datenqualitätsproblem besteht somit in vielen Fällen in einer mangelhaften Kommunikation zwischen den Unternehmensbereichen. Eine Vielzahl der Probleme, die in die Kategorie Definitionen fallen, lässt sich oftmals auf fehlende Kommunikationsforen innerhalb des Unternehmens zurückführen. Diese Erkenntnis ist gerade in Hinblick auf Business-Intelligence-Projekte mit fachbereichsspezifischen Applikationen wesentlich.

Das wichtigste Ergebnis der Untersuchung war allerdings die Erkenntnis, dass die Verantwortung für eine Sicherstellung der Datenqualität auf allen Ebenen der Datenverarbeitung fehlte. Die erste Maßnahme in diesem Projekt war somit der Aufbau einer entsprechenden Organisation sowie die Zuweisung klar definierter Verantwortungen.

I.2.5 Zusammenfassung

Zusammenfassend ist festzuhalten, dass die Ursachen für schlechte Datenqualität vielfältig sind. Schlechte Datenqualität ist eine subjektive Wahrnehmung aus Sicht der jeweiligen Anwender. Um die Datenqualität aktiv zu managen und nachhaltige Verbesserungen herbeiführen zu können, sind die Ursachen schlechter Datenqualität im Einzelfall zu analysieren und geeignete Maßnahmen zu definieren. Aktives Datenqualitätsmanagement bedeutet klar zugewiesene Verantwortungen für die Sicherstellung der Datenqualität auf allen Ebenen der Datenverarbeitung sowie die Bereitstellung entsprechender Budgetmittel und Ressourcen, um auch tatsächlich Verbesserungen herbeiführen zu können.