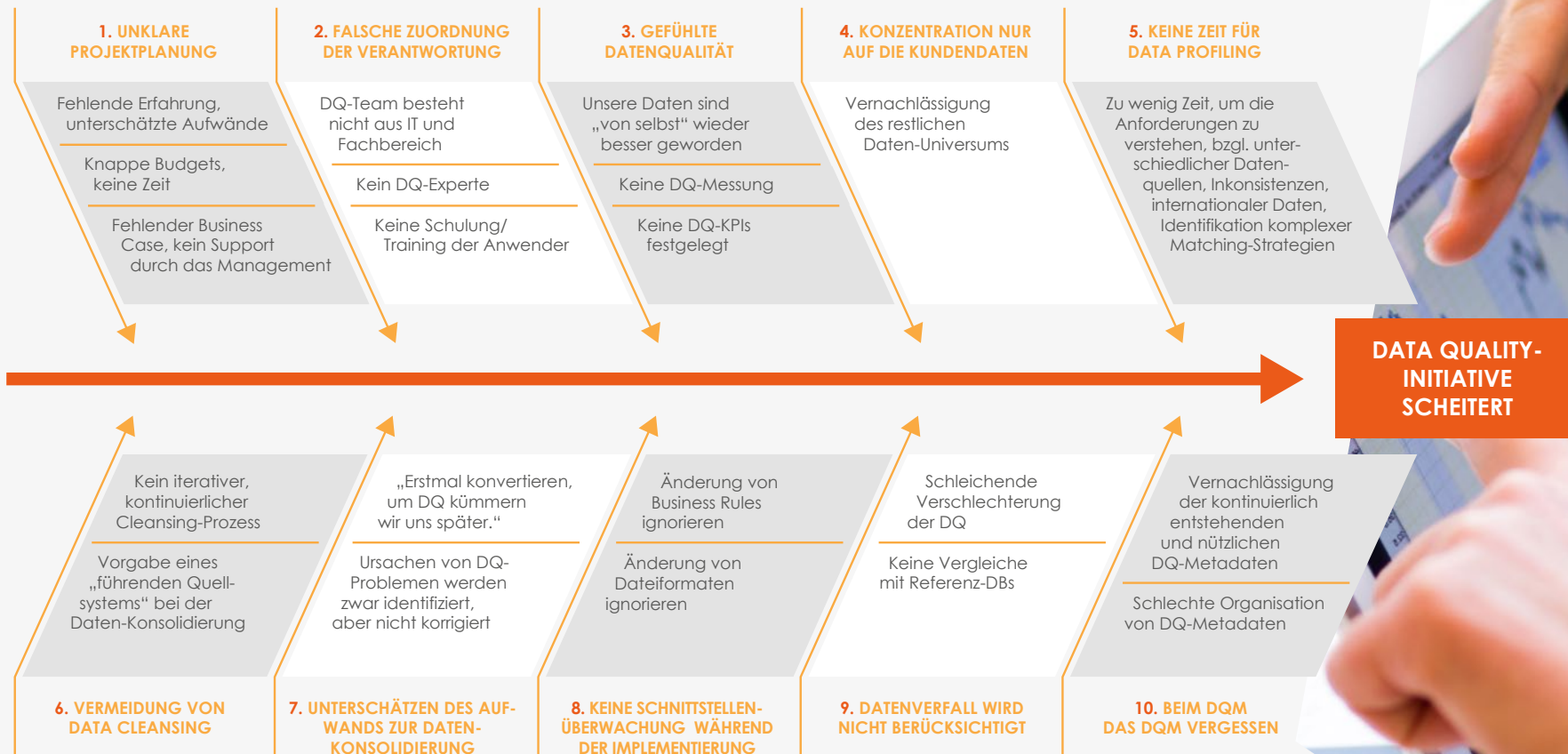


# 10 ZU VERMEIDENDE FEHLER IM DATA QUALITY MANAGEMENT



# 10 ZU VERMEIDENDE FEHLER IM DATA QUALITY MANAGEMENT

## 1. UNKLARE PROJEKTPLANUNG FÜR DAS DQM

In vielen Projekten soll das Thema Data Quality „nebenher“ mitgemacht werden. Es gibt kein eigenes Budget, keinen eigenen Business Case und keinen eigenen Projektplan – das Management ist nicht im Boot und die IT-Abteilung wird mit der Aufgabe allein gelassen.

### **Fehlende Erfahrung und knappe Budgets:**

Aufgrund enger Projektmeilensteine werden andere Themen vorrangig behandelt, das Thema Datenqualität startet zu spät und Best Practices werden nicht berücksichtigt.

Schon bei der Planung wird verkannt, wie komplex und zeitaufwändig die Aufgabenblöcke Data Profiling und Data Cleansing sind: Unterschätzt werden der Aufwand an Zeit und Investment für die Arbeit mit mehreren Datenquellen, inkonsistenten Datenformaten, internationalen Daten und auch der Aufwand, welcher sich bei der Identifikation, Planung, Test und Implementierung weiterführender Matching-Strategien ergibt.

## 2. FALSCH ZUORDNUNG DER VERANTWORTUNG: DIE ZUSAMMEN- SETZUNG DES DQM-TEAMS

„Wer ist verantwortlich für das Datenqualitätsmanagement?“ – doch keine klar definierte Gruppe verfügt über die entsprechende Expertise und Verantwortung. Selbst dann, wenn ein Unternehmen eine DQ-Abteilung etabliert, sind oft Mitarbeiter darin beschäftigt, die zwar Kenntnisse in allgemeiner IT und Datenverarbeitung, aber kein spezifisches DQ-Wissen haben. Und den Verantwortlichen für DQM-Initiativen fehlt oft die DQ-Erfahrung. **Diese personelle Ausstattung des DQ-Teams ist nicht ausreichend** (und nur selten gibt es eine DQ-Schulung oder ein DQ-Training).

Zwar erfordert Datenqualitätsmanagement ein gutes Maß an IT-Know-how. Doch **DQM ist keine One-Man-Show** für den IT-Bereich, sondern Unternehmensaufgabe: für ein DQM-Team braucht es immer sowohl IT-Spezialisten **als auch den Fachbereich**. Außerdem braucht das Team Datenqualitätsexperten – diejenigen, die Erfahrung aus erster Hand haben, von der Gestaltung und Implementierung bis hin zur Feinabstimmung von Datenqualitätsregeln und ihrer Kontrolle:

Die **DQ-Experten von INFOMOTION** helfen Ihnen dabei, Ihre DQ-Initiativen zum Erfolg zu führen. Ob es um die Sicherstellung hervorragender

Datenqualität für operative oder analytische Applikationen geht, um die erfolgreiche Durchführung von Systemmigrationen und -konsolidierungen, den Aufbau von Master Data Management-Systemen, den Aufbau eines Data Quality Warehouse, ob im lokalen Einsatz oder im Hybrid-Betrieb: **INFOMOTION bietet Ihnen für Ihre Aufgaben ein umfassendes und qualitativ hochwertiges Lösungsportfolio an.**

## 3. GEFÜHLTE DATENQUALITÄT – WENN DIE BEWERTUNG DER DATENQUALITÄT FEHLT

Zwar sind sich die meisten Unternehmen der Bedeutung von Datenqualität bewusst, können aber das Ausmaß der Probleme mit ihren aktuellen Daten nicht wirklich einschätzen. Ihr Wissen über DQ-Probleme beruht auf Gefühlen und Meinungen, nicht aber auf messbaren Fakten. Meist über- oder unterschätzen Unternehmen die Qualität ihrer Daten, selten erfassen sie die Auswirkungen der Datenqualität auf Geschäftsprozesse.

Hinzu kommt die trügerische Vermutung, dass sich die Datenqualität längst von allein verbessert hat, quasi als Nebenprodukt allgemeiner IT-Fortschritte: Denn in den vergangenen Jahren lag der Fokus von Datenverbesserungen auf modernen Datenbank-Technologien, besseren Informations-

# 10 ZU VERMEIDENDE FEHLER IM DATA QUALITY MANAGEMENT

systemen und anspruchsvollen Data Integration-Lösungen. Tatsächlich beeinträchtigen die meisten dieser IT-Prozesse die Datenqualität aber negativ: Neue System-Implementierungen und System-Aktualisierungen sind eine der Hauptursachen von DQ-Problemen. Schnittstellen zur Datenintegration erzeugen in Nu eine Unzahl von Fehlern. Und auch im generellen täglichen Betrieb ist die Datenverarbeitung nicht sicher vor Fehlern. Wenn nichts getan wird, verschlechtert sich die Datenqualität kontinuierlich bis zu dem Punkt, an dem die Daten zu einer riesigen Bürde werden.

Die Bewertung ist ein Eckpfeiler jedes DQM-Programms. Das Unternehmen erhält Kenntnis über den aktuellen Status, den Zustand der Daten, und es wird erkennbar, wie und wo Daten höherer Qualität verschiedene Prozesse unterstützen. Erst mit einer Bewertung kann ein Unternehmen abschätzen, was die bisherigen Datenprobleme gekostet haben.

**Was man nicht misst, kann man nur schwer bewerten:** Grundlegende Voraussetzung für den Erfolg von DQ-Initiativen ist eine genaue Kenntnis des Datenqualitäts-Zustandes der Businessdaten. Vollständigkeit, Korrektheit und Konsistenz sind nur einige der Kriterien, welche bei der Beurteilung der

Datenqualität herangezogen werden müssen. Daran scheitern viele BI-Projekte: Zwar werden Maßnahmen zur Verbesserung der Datenqualität eingerichtet; da es jedoch keine Methode zur Messung der DQ-Verbesserungen gibt, kann man Kosten und Nutzen nicht in Relation setzen.

Ein guter Weg, dieser Herausforderung zu begegnen, ist die Etablierung eines stetig aktiven Prozesses, welcher kontinuierlich die Datenqualität überwacht und vorhandene DQ-Niveaus bewertet und verbessert.

## 4. KONZENTRATION NUR AUF DIE KUNDENDATEN UND VERNACHLÄSSIGUNG DES RESTLICHEN DATEN-UNIVERSUMS

Bereits seit den 90er Jahren gibt es systematische DQM-Bemühungen aus Analyse, Abgleich, Standardisierung und Deduplizierung von Kundendaten. Und heute gibt es aktuelle Tools und Lösungen, die Unternehmen sehr hohe Erfolgsraten bei der Verbesserung ihrer Datenqualität ermöglichen. Viele Unternehmen haben diese Tools für Ihre Kundendaten bereits im Einsatz – **so ist heute insgesamt die Qualität von Kundendaten in Unternehmen auf höchstem Niveau.**

Dieser Erfolg gibt den Unternehmen ein gutes Gefühl bzgl. ihrer Bemühungen um eine hohe Datenqualität.

Doch gilt dies nicht für den Rest der Datenuniversums. Zwar sind Produkt- und Materialstammdaten, Finanz- oder Vertragsdaten zunehmend Gegenstand von Data Governance-Programmen und -Organisationen in den Unternehmen. Doch in viele andere Datenkategorien, wie HR, Finanzen, Product Orders und Sales, Kredite und Konten, Patienten oder Studenten, hat sich die Datenqualität kontinuierlich verschlechtert. Dabei sind diese Datenkategorien nicht weniger wichtig als Kundennamen und Adressen – und das Volumen dieser Daten ist viel größer. Allerdings ist die Struktur dieser Daten deutlich komplexer – eine „One-Tool-fits-all“-Lösung gibt es hier nicht. Mehr Aufwand und Expertise sind erforderlich.

Die bisherigen DQ-Tools sind hier bisher wenig hilfreich: Hier ist erst dann ein wesentlicher Fortschritt zu erwarten, wenn seitens Unternehmen die Nachfrage nach DQM-Programmen steigt, welche sich nicht nur auf Adressdaten fokussieren, sondern auch auf Daten aller anderen Bereiche sinnvoll anwendbar sind.

# 10 ZU VERMEIDENDE FEHLER IM DATA QUALITY MANAGEMENT

## 5. KEINE ZEIT FÜR DAS DATA PROFILING – KEINE ZEIT, UM DIE ANFORDERUNGEN ZU VERSTEHEN

Eine DQ-Maßnahme kann nur dann erfolgreich sein, wenn man den tatsächlichen Inhalt und die Struktur der Daten versteht. Meist stützen sich die Verantwortlichen auf die theoretischen Datendefinitionen und die Datenmodelle. Allerdings sind diese Informationen üblicherweise unvollständig, veraltet und inkorrekt. Die tatsächlichen Daten sehen nicht so aus, wie erwartet.

Doch auch aus Mangel an genügend Zeit und Budget zur Erfassung der Anforderungen (insbesondere für die Entwicklung und Validierung der Geschäftsregeln / Business Rules) wird der Ausgangszustand der Daten nicht vollständig verstanden.

Ein DQM-Programm sollte daher mit einem **umfangreichen Data Profiling** starten. Mit Data Profiling werden die Daten untersucht, um ihre tatsächliche Struktur sowie die tatsächlichen Abhängigkeiten zu verstehen. Erst auf Basis dieser Informationen können auch verlässliche Metadaten ermittelt werden.

[www.infomotion.de](http://www.infomotion.de)

## 6. VERMEIDUNG VON DATA CLEANSING: DAS PROBLEM MIT DEM „FÜHRENDEN QUELLSYSTEM“ BEI DER DATENKONSOLIDIERUNG

In einem Data Warehouse laufen Daten aus mehreren verschiedenen Betriebssystemen und Quellen zusammen. Nötig ist damit die Datenkonsolidierung (grundlegende Datenkonversion) – und fast immer überschneiden sich dabei die Daten aus den zu konsolidierten Quellsystemen: Es gibt zahlreiche Datenkonflikte, vom einfachen Duplikate bis hin zu überlappenden Themenbereichen und Datenhistorien.

### Falscher Ansatz: „The winner takes it all“:

Als pragmatischer Ansatz wird üblicherweise einfach eine Win / Lose-Matrix erstellt, die anzeigt, welches Quelldaten-Element im Konfliktfall abgeholt wird. Zielführend für eine solche Datenkonsolidierung wäre jedoch ein sauberes Data Cleansing: Der Datenbestand wird zunächst bereinigt, damit jeweils das tatsächlich relevante Quellattribut ermittelt und erst dann wird dieses weitergereicht.

Data Cleansing ist ein kontinuierlicher Prozess: Die Abstimmung und Konsolidierung erfolgt iterativ und kontinuierlich – und erfordert wiederholtes Testen, Optimieren und Abstimmen, um die gewünschten Ergebnisse zu erzielen.

## 7. UNTERSCHÄTZEN DES AUFWANDS FÜR DIE DATENKONSOLIDIERUNG

Ein Data Warehouse entsteht oft auf Basis von Datenkonvertierungen aus verschiedenen operativen Datenbanken und dieser Start ist meist etwas heftig und holprig. Der größte Teil des Implementierungsaufwands entsteht dabei in der Datenkonvertierung und nur selten läuft das reibungslos!

Grob gesehen hat ein Quellsystem drei Schichten: Datenbank, Business Rules und Benutzeroberfläche. Was Benutzer sehen, ist nicht das, was tatsächlich in der Datenbank gespeichert wird – insbesondere bei veralteten operativen Systemen. Bei der Datenkonvertierung konzentriert man sich aber nur auf die Datenstruktur, damit die Daten aus der alten DB in die neue DB passen. Dabei ist übliche Praxis: **„Was wir haben, haben wir – erstmal konvertieren, um die Datenqualität kümmern wir uns später“.**

Business Rules der Quellsysteme werden weder verstanden noch berücksichtigt. Häufig gibt es auch keine zuverlässigen Metadaten (mehr) über die Quelldatenbank. Während der Datenkonvertierung werden Ursachen von Datenqualitätsproblemen zwar identifiziert, aber nicht korrigiert. Einer hohen Datenqualität ist das nicht gerade zuträglich.

# 10 ZU VERMEIDENDE FEHLER IM DATA QUALITY MANAGEMENT

Die Datenqualität nach der Konvertierung steht in direktem Zusammenhang zu dem Aufwand, welcher für die Analyse und das Data Profiling sowie für die Ermittlung des tatsächlichen Dateninhalts investiert wird. Ein ideales Datenkonvertierungs-Projekt beginnt mit Data Profiling, umfassender DQ-Bewertung und Data Cleansing. Erst danach sollte es um Transformationsalgorithmen gehen.

## 8. KEINE SCHNITTSTELLEN- ÜBERWACHUNG: DATEIFORMATE UND / ODER BUSINESS RULES ÄNDERN SICH AUCH WÄHREND DER IMPLEMENTIERUNG

Ein Data Warehouse erhält Tag für Tag sowohl eine Unmenge von Lieferungen vom Batch-Feeds bis zu Echtzeit-Nachrichten aus mehreren Datenquellen und über viele verschiedene Schnittstellen – so manches DQ-Problem hat hier seinen Ursprung.

Durch interaktive Anwendungen können Datenbestände permanent verunreinigt werden, z. B. bei der Erfassung neuer, aber auch bei der Veränderung bereits vorhandener Datensätze. Sei es durch fehlerhafte Dateneingabe, einer Mehrfachspei-

cherung des gleichen Objekts, eine Umwidmung bestimmter vorgesehener Datenattribute u. v. m.

Erst im Laufe der Zeit sammeln sich diese Probleme signifikant an. Diese fallen lange nicht auf, denn eine schnellere Datenweitergabe und niedrigere Datenlatenzen stehen im Vordergrund

Die Lösung für eine Schnittstellenüberwachung besteht in der **Etablierung von Programmen**, die zwischen den Quell- und Zieldatenbanken operieren, und die die Aufgabe haben, die **Schnittstellendaten zu analysieren**, bevor sie geladen und verarbeitet werden. Individuelle Datenmonitore verwenden DQ-Regeln zur Prüfung von Datengenauigkeit und Integrität. Ihr Ziel ist es, alle potenziellen Datenfehler zu identifizieren. Höher entwickelte Prüfroutinen können komplexere Geschäftsregeln abbilden, Daten systemübergreifend zwischen Batches und Zieldatenbanken vergleichen und damit weitere potenzielle Probleme herausfinden. Manche Prüfroutinen enthalten eine Aggregations-Funktion und suchen nach unerwarteten Änderungen in Batch-Schnittstellen. Sie vergleichen verschiedene Aggregateigenschaften (z. B. Zählungen von Attributwerten) von einer Lieferung zur anderen. Sobald ein Wert außerhalb eines definierten Intervalls liegt, wird ein potenzielles Problem angezeigt.

## 9. DATENVERFALL WIRD NICHT BERÜCKSICHTIGT

Daten sind nur dann korrekt, wenn sie reale Objekte abbilden. Doch Daten, die ehemals akkurat waren, werden im Laufe der Zeit ungenau, auch wenn es im IT-System gar keine physische Änderungen gab. Personen verlegen ihren Wohnsitz, Unternehmen ändern ihre Rechtsform, Orte werden eingemeindet, ein Produkt wird umbenannt. Ob fehlerhaftes Verfahren zur Datenerhebung oder eine defekte Datenschnittstelle – irgendwann sind die **Daten und die reale Welt nicht mehr synchron**.

Zwar geht man von perfekten Datensammelungsprozessen aus – in Wirklichkeit aber verlaufen Änderungen an realen Objekten üblicherweise für Computer unbemerkt. Von außen wirkende Veränderungen der Datenobjekte führen trotz einer korrekten initialen Bereinigung und einer laufenden Qualitätssicherung schleichend zu einer Verschlechterung der Datenqualität.

Diesem Effekt muss durch periodisch durchzuführende Maßnahmen zur Wiederherstellung der gewünschten Datenqualität entgegengewirkt werden: Eine revolvierende DQ-Prüfung und ein Stichprobenvergleich mit Referenzdatenbeständen liefern Informationen über die Zerfallsrate und zeigen jene Kategorien von Daten, welche am schnellsten zum Verfall neigen.

# 10 ZU VERMEIDENDE FEHLER IM DATA QUALITY MANAGEMENT

Dieses Wissen kann zur Verbesserung des Verfahrens zur Datenerhebung und der Datenschnittstellen verwendet werden.

## 10. BEIM DQM DAS DQM VERGESSEN: SCHLECHTE ORGANISATION VON DQ-METADATEN

DQ-Initiativen produzieren enorme Mengen wertvoller Metadaten. Die DQ-Bewertung gibt Auskunft über bestehende DQ-Probleme und ihre Auswirkungen auf verschiedene Geschäftsprozesse.

Ein sehr verbreitetes Problem im DQ-Management ist jedoch eine unzureichende Architektur der DQ-Metadaten-Repositories. **Der Schuster hat die schlechtesten Schuhe:** Ausgerechnet die Projekte zur Bewertung der Datenqualität generieren routinemäßig unzählige unstrukturierte Fehlerberichte, ohne effektive Möglichkeit, die Ergebnisse zusammenzufassen und zu analysieren. Data Cleansing-Initiativen verfügen typischerweise nicht über Audit-Trail-Mechanismen, und ETL-Prozesse haben oft keine Data Lineage-Informationen. Dadurch wird der Wert von DQ-Initiativen stark verringert. Im Worst-Case-Szenario werden DQ-Projekte komplett aufgegeben.

[www.infomotion.de](http://www.infomotion.de)

Zur wirkungsvollen und nachhaltigen Implementierung einer DQ-Strategie muss die Qualität der gespeicherten Geschäftsdaten kontinuierlich überwacht und reportet werden. Ohne eine regelmäßige Kontrolle der Datenqualität besteht die Gefahr, dass ein einmal erreichter Qualitätslevel im Zeitablauf verloren geht. Data Cleansing ermittelt Ursachen von Fehlern und mögliche Fehlerbehandlungen. Schnittstellen-Monitoring identifiziert laufende DQ-Probleme und informiert über Data Lineage, ebenso wie Datenkonvertierung und -konsolidierung.

Die Lösung wäre mit Bordmitteln des Data Warehouse umsetzbar: Auf dessen Struktur könnte daneben ein umfassendes Data Quality-Warehouse etabliert werden – eine Sammlung von Tools für die Organisation und Analyse sämtlicher Metadaten, die für die DQ-Initiativen relevant sind oder von diesen erzeugt werden.

Diese Lösung ist komplex und verbindet Elemente des objektorientierten Metadaten-Repositories mit der analytischen Funktionalität eines Data Warehouse. Doch ohne ein passend konzipiertes Data Quality-Warehouse leidet jede DQ-Initiative unter genau der Krankheit, die sie eigentlich heilen will: einer schlechten Datenqualität.

SIE HABEN FRAGEN ZUM  
DATA QUALITY MANAGEMENT?

Kontaktieren Sie uns bezüglich Ihrer individuellen DQ-Lösung - von der Strategie bis zur Umsetzung. Unsere Experten beraten Sie gerne.

## IHR ANSPRECHPARTNER



**ANDRÉ TRAPPMANN**  
Principal Consultant

INFOMOTION GMBH  
Glockengießerwall 20  
20095 Hamburg  
T +49 69 56608-3437  
[anfrage@infomotion.de](mailto:anfrage@infomotion.de)

Die INFOMOTION GmbH ist das führende Beratungsunternehmen für Business Intelligence, Big Data und Digital Solutions im deutschsprachigen Raum. Das Leistungsportfolio reicht von der strategischen Beratung über die Konzeption, Implementierung und den Betrieb nachhaltiger Lösungen bis hin zur Schulung Ihrer Mitarbeiter.

**INFOMOTION**  
we love data