

## DQM - Ein modularer Lösungsansatz in fünf Stufen

AUGUST 2018 — VON CLEMES BOTT / THOMAS HERZ

Aufgrund der zunehmenden Integration der IT in die Geschäftsprozesse von Finanzinstituten sowie umfangreicher Compliance-Anforderungen nimmt die Bedeutung der Datenqualität erheblich zu. Einer aktuellen Erhebung von Experian und Gartner zufolge

- » sind 59% aller Datenqualitätsprobleme auf menschliche Fehler zurückzuführen
- » machen 42% der Unternehmen die Nutzung mehrerer Datenbanken für ihre Datenqualitätsprobleme verantwortlich
- » gefährden 75% der Unternehmen durchschnittlich 14% ihres Umsatzes aufgrund schlechter Datenqualität.

### Motivation

Die Erfassung von Daten läuft in der Regel über Benutzerschnittstellen oder automatisierte Geschäftslogiken vorgelagerter Systeme. Meist werden Daten dabei über Zwischenstationen portiert und transformiert. Data-Warehouse-Systeme sind häufig die „Endstation“ dieser Datenflüsse. Somit treten Datenqualitätsprobleme zumeist erst hier ans Licht. Das liegt daran, dass in Data-Warehouse-Systemen sämtliche Daten in verdichteter Form betrachtet werden. Werden Datenqualitätsprobleme nicht rechtzeitig erkannt und behandelt, führt dies zu Folgefehlern, die sich potenzieren und zu größeren Problemen aufschaukeln können. Es liegt somit auf der Hand, dass Finanzinstitute daran arbeiten müssen, systematisch die Qualität ihrer Daten zu verbessern und über den Aufbau vertrauenswürdiger Governance-Informationsmodelle das Asset ihrer Daten zu ‚monetarisieren‘.

### Verbreitete Ansätze

Datenqualitätsprobleme sind nicht neu. Institute haben nach unseren Projekterfahrungen mehrere Möglichkeiten gefunden damit umzugehen. Die am meisten vorzufindenden Ansätze sind:

- » Fest codierte Prüfoperationen  
Datenqualitätsprobleme in ablaufenden Business Prozessen sind häufig darauf zurückzuführen,

dass Prüfalgorithmen in bestehende Abläufe hart codiert eingebaut werden und dabei an unterschiedlichen Stellen nicht auf gleiche Definitionen zurückgegriffen wird.

- » Konvertierungen  
Im Rahmen von Datenaufbereitungen (Berichtswesen), -aggregationen und -konvertierungen geht häufig die Möglichkeit der Rückverfolgbarkeit der Herkunft und der an ihnen durchgeführten Operationen (Data Mining) verloren.
- » DQ-Prüfungen im ETL-Ansatz  
Im ETL-Verfahren implementierte DQ-Prüfung weisen häufig inkonsistente Qualitätsregeln auf. Speziell tritt dies bei Datenqualitätsprüfungen über Systemgrenzen hinweg auf. Dies führt zu fehleranfälligen manuellen Eingriffen in bestehende Prüfprozesse.
- » Einsatz von Desktoptools  
Auf Basis von Desktoptools durchgeführte manuelle DQ-Prüfungen weisen wie jede IDV ein Glaubwürdigkeitsproblem hinsichtlich der erzielten Ergebnisse auf, da sie mit hohem fehlerbehaftetem manuellen Aufwand einhergehen, um Daten zusammenzuführen, abzugleichen und zu berichten.
- » Workflowbasierter DQ-Ansatz  
Heutige Prozessabläufe sind zunehmend vielschichtig und komplex. Der Output eines Prozesses wird zum Input für Folgeprozesse. Auswirkungen von Datenqualitätsfehlern auf die gesamte Prozesskette lassen sich nur durch Berücksichtigung aller Abhängigkeiten identifizieren.
- » Manueller Korrekturansatz bei Datenpräsentation  
Häufig vorgenommene manuelle Datenkorrekturen und -adjustierungen zum Zeitpunkt der Datenaufbereitung für das Berichtswesen beeinflussen die Glaubwürdigkeit der präsentierten Informationen in erheblichem Umfang.

Alle vorgestellten Ansätze haben den Charakter von isoliert implementierten Workarounds und sind somit Initiativen, die einem konsequenten und ausgereiften Modell zur Verbesserung der Datenqualität nicht zuträglich sind.

## Master Data Management als strategisches Ziel

Mit zunehmender Komplexität der Datenlandschaften und der darin ablaufenden Prozesse wächst die Erkenntnis, dass ein konsistentes Datenqualitätsmanagement verankert werden muss.

### Modulares Vorgehensmodell



Ein ganzheitlicher Lösungsansatz sieht ein phasenorientiertes Vorgehen im Rahmen eines Reifegradmodells vor, um so je nach Status quo an die unterschiedlichen Reifegradstufen in Unternehmen anknüpfen zu können. Zur Ermittlung der individuellen Reifegradstufe eines Unternehmens sollte einer Datenqualitätsanalyse stets ein Value Assessment vorgeschaltet werden.

### 1. Value Assessment

Unter Einbeziehung aller betroffenen Business- und IT-Bereiche wird hier ein ganzheitliches Bild über die Qualität von Daten, Prozessen und einzusetzenden Technologien erhoben. Ein Business Case quantifiziert darüber hinaus die Chancen eines möglichen Datenqualitätsmanagementprojektes und fasst die fachlichen Ergebnisse in einer kundenindividuellen Roadmap zusammen.

Mit der Anbindung der Geschäftsführungssysteme beginnt der Data-Governance-Prozess. Die Datenextraktion und Prozesshomogenisierung inkl. der Anwendung des Regelwerkes bilden die Basis für die eigentliche DQ-Messung.

### 2. Data Analysis & Data Profiling

Nach Konnektierung der geschäftsführenden Systeme werden Daten performant in logisch korrekte, einheitliche Strukturen überführt. Auf diese homogenen Datenstrukturen kommt ein vorab definiertes Prüfregelwerk zur Anwendung. Dieses setzt sich aus einer Vielzahl von Analyseprofilen bestehend aus Minimal-/Maximalwerten, Durchschnittsgrößen, Konsistenz- & Plausibilitätsregeln

u. v. m zusammen.

**Der sich anschließende Bereinigungsprozess selbst bedarf eines automatisierten Ansatzes.**

### 3. Data Quality Improvement

Der Bereinigungsprozess (Data Cleansing) erfolgt auf Basis eines definierten Datenqualitätsregelwerkes, das sich im Kern aus anpassbaren, wiederverwendbaren Branchenstandards, Algorithmen und kundenspezifischen Vorgaben zusammensetzt. Das Ergebnis ist eine quantitative Verteilung von Fehlerbildern (Aussagen über Einzelfälle bzw. Fehlerbilder die im Verbund auftreten). Hieraus lassen sich wirkungsvolle Bereinigungsmaßnahmen ableiten. Schwellenwerte bzw. KPIs sind Kernelemente der Data-Governance-Strategie. Sie ermöglichen im Rahmen eines Scorings Aussagen über den aktuellen Grad der Datenqualität und liefern Hinweise für den Einsatz automatisierter Prüfverfahren in Abgrenzung zu manuellen Korrekturmöglichkeiten. Das Data Mastering wendet im Nachgang Regelwerke an, die Datensätze identifizieren, „matched & merged“ sowie Duplikate eliminieren.

**Saubere, bereinigte bzw. korrigierte Daten als Ergebnis des Improvement-Prozesses müssen schließlich wieder an die konsumierenden Systeme zurückgesendet werden.**

### 4. Data Inventory Correction & 5. DQ-Reporting

Die Ergebnisdaten aus der bereinigten Zielstruktur werden über bereits aufgebaute Datenverbindungen in die Quellsysteme zurückgeladen. Bei der Distribution kommen je nach Technologieanbieter unterschiedliche Integrationsszenarien (Batch, Real-time, ETL oder Big-Data-Integration) und Adapter Suiten zum Einsatz.

Das DQ-Reporting setzt auf der Messung vordefinierter KPIs auf. Aufbereitung und Kommunikation erfolgen über klassische BI-Visualisierungstools. Werden die Datenqualitätsmessungen periodisch ausgeführt, können Entwicklungen im Zeitverlauf dargestellt werden, um Verschiebungen in Fehlerbildern zu identifizieren.

**Mit der Datendistribution in die datenföhrenden Quellsysteme und dem analytischen DQ-Reporting wird der Regelkreis zur Bereinigung der Datenbestände geschlossen.**

Perspektivisch kann nur die Etablierung eines Master Data Managements in der Organisation den Geschäftswert nachhaltig steigern und langfristig Vertrauen in Daten wiederherstellen. Über den Bereinigungsansatz hinaus sind

dazu Daten zu modellieren, klare Governance-Prozesse zu etablieren und die Basis für erfolgreiche Audits, Datensicherheit etc. zu legen. Die Umsetzung folgt dabei einer Roadmap von Aufgabenpaketen, die projekthaft aufeinander aufbauend für alle Datenbereiche umzusetzen sind. Zu den wesentlichen Erfolgsfaktoren zählen insbesondere die organisatorischen Rahmenbedingungen (Rollen, Verantwortlichkeiten und Regeln für Datenentitäten) wie auch ein bereichsübergreifendes Verständnis für Datenqualität, das sich auf die Zusammenarbeit im operativen Geschäft unmittelbar auswirkt. Entsprechender Management-Support ist hierfür notwendig. Unterstützung erhalten die Mitarbeiter von dedizierten Stewards einer Data Factory (laufender Support, Profilingdienste, Beratungsleistungen und Monitoring). So lassen sich individuelle Prozesse zu gemeinsamen Datenprozessen zusammenführen.

## **Zusammenfassung:**

Die zunehmende Komplexität der Datenlandschaften und die darin ablaufenden Prozesse begründen heute die Notwendigkeit zur Implementierung eines konsistenten Datenqualitätsmanagements und dessen Weiterentwicklung in Richtung Master Data Management. In der Regel erfolgt der erste Entwicklungsschritt aus der Inaktivität mit dem Ziel, die aktuelle Datenqualität zu identifizieren und so die Basis für gezielte reparierende Eingriffe zu schaffen.

Workarounds, die sich daraus ergeben, haben zumeist nur reaktiven Charakter, behandeln Symptome, eliminieren Fehlerquellen aber nicht dauerhaft und systematisch. Dezidierte Ursachenforschung und Datenanalysen führen zu einem proaktiven Improvement-Prozess. Analyseprofile mit hinterlegten Regelwerken werden dazu auf extrahierte Daten angewendet.

Im Bereinigungsprozess selbst geht es um die Auswahl und den Einsatz der entsprechenden Werkzeuge und um den Abgleich der Datenqualität mit vorgegebenen Schwellenwerten. Ziel eines proaktiven, durchgängigen Ansatzes muss jedoch die Verankerung eines Master Data Management Hubs in der Organisation sein. Während der reine Plattformansatz die Datenqualität lediglich reaktiv verbessert, ermöglicht erst der proaktive Ansatz, einen 360°-Blick auf die Daten-Assets zu erlangen. Das dargestellte Vorgehen beschreibt einen Entwicklungsprozess, in dessen Vordergrund das Bewusstsein für den Wert der eigenen Daten und deren Qualität steht.

Wenn sich in der Organisation dieses allgemeine Bewusstsein für Datenqualität etabliert, ergeben sich Synergien auf mehreren Ebenen: die Etablierung hoher Qualitätsstandards für die Daten sowie entsprechender Prozesse und Technologien, die diese Qualität sicherstellen, reduziert nicht nur Kosten, sondern hebt gleichzeitig Potenziale im Umgang mit Kunden. Darüber hinaus existieren zahlreiche interne Stakeholder, die von hoher Qualität im Datenmaterial profitieren. So können Arbeitsabläufe erleichtert und etwaige manuelle Korrekturprozesse vermieden werden.

Auch im regulatorischen Umfeld ist mit erheblichen Erleichterungen zu rechnen: insbesondere im Umfeld des Meldewesens bzw. des Reportings ist ein etabliertes Master Data Management eine solide Grundlage, um auch künftigen Anforderungen gerecht zu werden.

Die Initiierung einer „Daten-Qualitätsoffensive“ liefert demnach in mehrfacher Hinsicht Gewinne. Zweifellos erscheint die zu bewältigende Aufgabe umfassend, doch sollte man sich hierbei nicht vom Ausmaß einer solchen Offensive irritieren lassen. Es kommt vielmehr darauf an, überhaupt zu beginnen.