

Datenqualitätsmanagement

# Hohe Datenqualität erfordert Mix aus Software und Best Practices

Der Aufwand für die Qualitätssicherung der Daten in Business-Intelligence- und Big-Data-Projekten wird oft unterschätzt. Ein Mix aus Software-Tools und Best Practices ist erforderlich.



**So lange** mit der Datenqualität alles gut läuft, wird sie als selbstverständlich hingegenommen. Erst wenn Probleme auftreten, kommt es zu einem lauten Aufschrei. Die Folgen falscher Analysen sind für Unternehmen unabsehbar. Schon wenige unsaubere Inhalte in den Kundendaten können zu hohen Zusatzkosten und zur Verärgerung bei der angestammten Klientel führen. Unterm Strich steht ein Image- und Umsatzverlust.

Wird die Datenqualität nicht frühzeitig in die Planung von Business-Intelligence- oder Big-Data-Projekten einbezogen, dann lassen sich laut Oraylis die Fehler meist nur mit hohem Aufwand identifizieren und beseitigen. Datenqualität ist ein komplexes Thema und die Ursachen von Qualitätsmängeln können äußerst vielfältig sein. Sie reichen von Defiziten beim Datenimport über Systemausfälle bis hin zu falschen Ein-

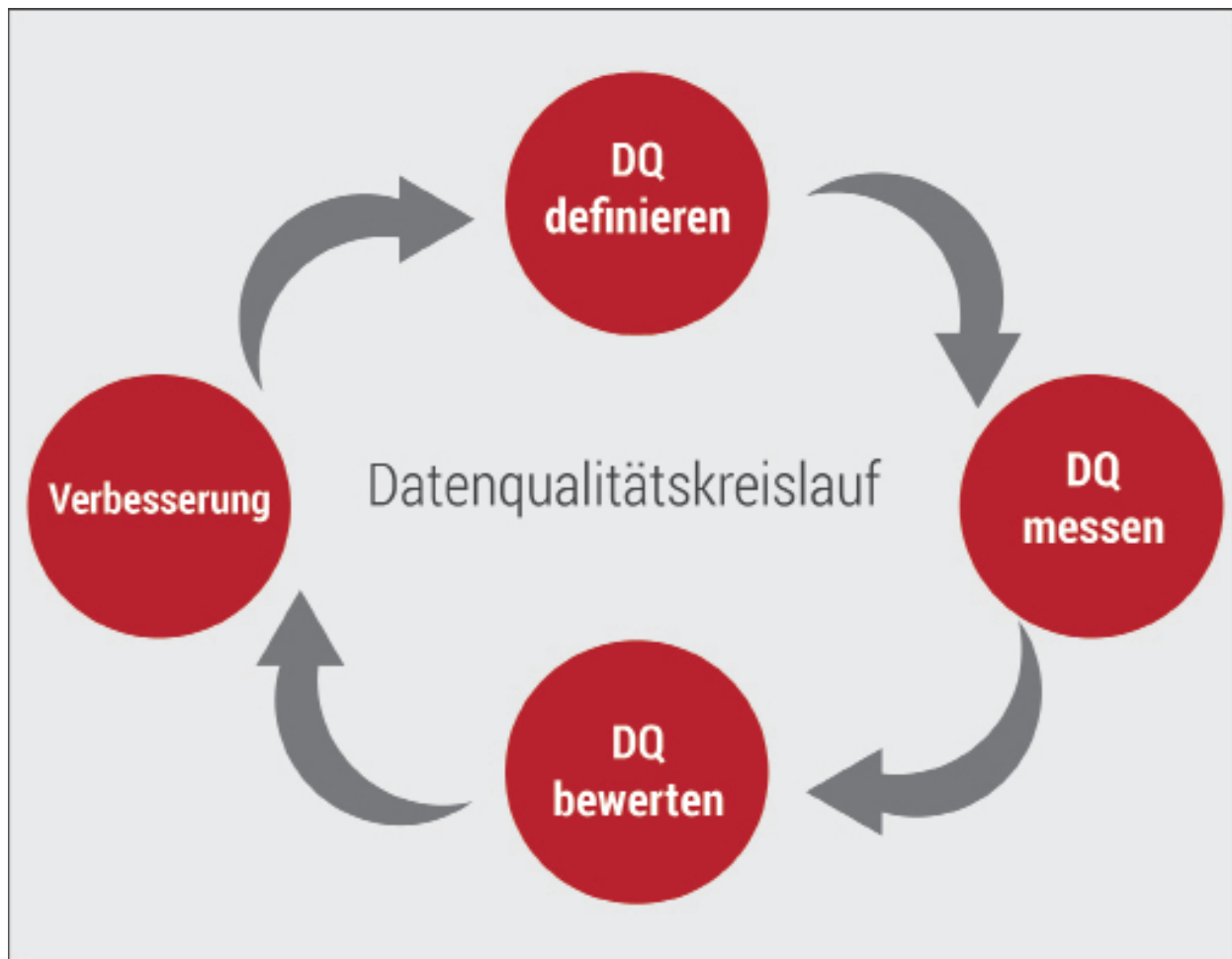
gaben. Um die Frage zu beantworten, wie die Verantwortlichen etwaigen Fehlern im Datenbestand vorbeugen oder diese zumindest in kürzester Zeit beheben können, sei es sinnvoll, zunächst den Begriff der Datenqualität zu untersuchen.

#### **Die Definition von Qualität fällt vielfach individuell aus**

„Datenqualität ist ein höchst subjektiver Begriff, der in jedem Projekt



Foto: Fotolia



Datenqualitätsmanagement stellt laut Oraylis einen fortlaufenden Prozess dar, in dessen Kontext die Qualitätsmaßstäbe immer wieder aufs Neue definiert, geprüft und optimiert werden.

Quelle: Oraylis GmbH

anders definiert wird“, erläutert Thomas Strehlow, Geschäftsführer der BI- und Big-Data-Experten Oraylis. „Entsprechend allgemein fällt auch die Definition aus. Wir sprechen in diesem Kontext von der Gesamtheit aller Eigenschaften eines Datenbestandes, mit denen die Anforderungen des Anwenders erfüllt werden.“ Daten müssten zunächst einmal konsistent sein – also widerspruchsfrei und ohne Doubletten. Sie sollten zudem vollständig vorliegen und aus validen beziehungsweise glaubwürdigen Quellen stammen. Ebenso wichtig sei die Genauigkeit und Granularität etwa mit Blick auf das Format sowie die Anzahl der Nachkommastellen. Schließlich müssten die Daten ak-

tuell sein, also entsprechend der Erwartung pünktlich bereitgestellt werden.

Um diese allgemeine Begriffsbestimmung herum setze jedes BI- oder Big-Data-Projekt seine eigenen Schwerpunkte bezüglich der erforderlichen Datenqualität. Sowohl beim Aufbau als auch bei der Optimierung einer Lösung sollten die Projektbeteiligten im ersten Schritt exakt festhalten, was sie unter Datenqualität verstehen. „Wenn ein klarer Maßstab existiert, lässt sich Qualität messen“, erläutert Strehlow. Fehler können dann frühzeitig identifiziert und proaktiv behoben werden.“ Dazu gehöre auch, dass der Qualitätsbegriff im Zuge laufender Verbesserungen immer

wieder auf den Prüfstand gestellt und entsprechend modifiziert werde. Hundertprozentig korrekt seien Daten niemals.

Oraylis betrachtet die Datenqualität als einen fortlaufenden Prozess, den es im Rahmen eines Projektmanagements organisatorisch zu verankern gilt. Das Projektmanagement fungiere als Leitstelle, die sämtliche Aufgaben rundum das Thema koordiniert. Es bringe interne wie externe Beteiligte an einen Tisch und definiere Projektrollen, Stakeholder, Anforderungen sowie Zuständigkeiten. Zudem werde festgelegt welche konkreten Maßnahmen in welcher Abfolge vorzunehmen seien.

Schließlich sei das Projektmanagement für regelmäßige Audits

zuständig, aus denen künftige Vorgehensweisen abgeleitet würden.

### Software kann Daten automatisiert prüfen

Beliebte Werkzeuge des Datenqualitätsmanagements seien Software-Lösungen, welche die Datenqualität an bestimmten Stufen des Prozesses automatisiert prüften. Diese Tools könnten bereits beim Import und der Migration von Daten aus den Quellsystemen ansetzen. „Die Quellsysteme können Daten beispielsweise verspätet, unvollständig, in unerwarteten Formaten oder gar nicht liefern“, erläutert Strehlow. „Dann bietet sich eine laufende Analyse auf Basis entsprechender Software an.“ Dabei würden die gelieferten Daten anhand zuvor definierter Vorgaben abgeglichen. Träten Datensätze auf, die den Regeln nicht entsprechen, dann werde der Prozess frühzeitig abgebrochen und die Verantwortlichen würden unmittelbar informiert. Auf die gleiche Weise ließen sich auch Mängel im Bereich der Implementierung identifizieren, etwa wenn es um Verarbeitungsfehler bei der Beladung geht.

Einzelne Lösungen könnten in diesem Kontext ein breites Aufgabenspektrum abdecken: „Mit BI Quality haben wir ein Werkzeug entwickelt, das eine umfassende Qualitätsprüfung sowohl während als auch nach der Bewirtschaftung ermöglicht“, berichtet Strehlow. „Fehler werden in jeder Phase erkannt und können unmittelbar behoben werden.“ Abhängig von der jeweiligen Umgebung empfehle sich neben derartigen Prüfwerkzeugen der Einsatz weiterer Software: „Kein Tool kann alleine ein umfassendes Datenqualitätsmanagement gewährleisten.“ Meist sind immer mehrere Lösungen erforderlich, etwa für die tiefgehende Analyse von Datenzusammenhängen, manuelle Korrekturen, die Verwaltung von Masterdaten oder eine Doublettenanalyse.“ Den am besten Mix passenden müs-

## Der Experte



Foto: Oraylis GmbH

Thomas Strehlow ist Geschäftsführer der Oraylis GmbH und Experte für die Implementierung von Business-Intelligence-Systemen und Big-Data-Lösungen.

se jedes Unternehmen für sich finden. Alleine mit der richtigen Tool-Mischung sind Unternehmen laut Oraylis bezüglich ihrer Datenqualität noch nicht auf der sicheren Seite. Gerade im Bereich der Implementierung gebe es Situationen, in denen ausschließlich Best Practices und umfangreiches Spezialwissen weiterhelfen würden. So könne beispielsweise ein erfahrener Experte Qualitätsmängeln vorbeugen, die aus fehlerhaften Logiken oder Altsystemen resultieren. Beim Aufbau eines neuen BI-Systems ließen sich auf diese Weise auch Fehler auf der ETL-Strecke (Extraktion, Transformation, Laden) ausschließen.

### Best-Practices und Fachwissen sind unverzichtbar

Manchmal lasse sich eine hohe Datenqualität nur durch den intensiven Austausch zwischen allen Beteiligten im Rahmen des Projektmanagements sicher stellen. Ein typisches Beispiel dafür seien schnelle Änderungen in den Geschäftsprozessen, die in den Quellsystemen als Workaround abgebildet würden. Zukäufe von Unternehmen und deren Infrastruktur gehörten ebenfalls dazu. Auch falsch definierte Schnittstellen oder Missverständnisse zwischen Anforderung und Abnahme machten

eine tiefgreifende Kommunikation zwischen den Projektbeteiligten erforderlich.

Grundsätzliche Modifikationen an vorhandenen IT-Systemen erforderten stets eine enge Abstimmung, um das Datenmodell mit Blick auf die referentielle Integrität, Datentypen und Constraints anzupassen. Falsche Eingaben der Anwender nach Veränderungen in den entsprechenden Maske ließen sich beispielsweise durch eine Eingabvalidierung auf der Basis von Referenzdaten oder eine Doublettenuche beim Anlegen neuer Einträge vermeiden.

Insgesamt sei die Datenqualität eine anspruchsvolle Aufgabenstellung, die bei der Planung von Business-Intelligence- und Big-Data-Projekten frühzeitig Beachtung finden sollte. Auf Basis eines gezielten Projektmanagements werde idealerweise ein kontinuierlicher Prozess zu etabliert, der eine ständige Messung, Bewertung und Justierung der Datenqualität beinhalte. Dabei sollten sämtliche Protagonisten des Datenmanagements einbezogen werde, zum Beispiel Entwickler, Administratoren und Anwender. Speziell mit Blick auf die erforderlichen Best Practices könne es zudem sinnvoll sein, externe Beratungskompetenzen hinzu zu ziehen.

Hinsichtlich der Umsetzung rät Strehlow, dass die Unternehmen klein anfangen: „Die Verantwortlichen sollten sich zunächst auf das Wesentliche konzentrieren und erst im weiteren Verlauf vertiefende Maßnahmen vornehmen. Wichtig ist, immer die Relation von Kosten und Nutzen im Blick zu behalten.“ Zum Erfolg führe die passende Mischung aus Software-Lösungen und konkreten Erfahrungswerten sowie fachlichen Kenntnisse zu Zahlen: „Niemand sollte sich alleine auf das OK einer Software verlassen. Der gesunde Menschenverstand bleibt eine wichtige Instanz bei der qualitativen Bewertung von Daten und Analyseergebnissen.“ *if*