

Enterprise Data Management - Fünf Schritte gegen das Bauchgefühl

Geschäftsentscheidungen sollten stets auf Basis korrekter und aktueller Daten getroffen werden. Doch dazu ist zunächst ein konsistentes Qualitätsniveau besagter Daten nötig – und zwar unternehmensweit.

Von Eric Ecker

Besonders bei erfolgreichen Unternehmern scheint es oft so, also ob sie wichtige Entscheidungen aus dem Bauch heraus treffen. Das mag in Einzelfällen stimmen, in der Praxis basieren Geschäftsentscheidungen aber auf harten Fakten. Und das ist gut so, denn ganz gleich ob betrieblich, analytisch oder strategisch – jede unternehmerische Entscheidung sollte gut überlegt sein, denn schließlich zieht sie fast immer Konsequenzen für das jeweilige Unternehmen nach sich.

So ist es für strategische Entscheidungen beispielsweise sehr wichtig, welche Kunden welche Produkte wie lange einsetzen oder wie viele Kunden den Kauf eines höherwertigen Produkts in Betracht ziehen. Zuverlässige Daten verraten dies bereits im Vorfeld. Falsche Daten wiederum führen dazu, dass in der Firma Fehler gemacht bzw. falsche Entscheidungen getroffen werden. So führen dann etwa Marketing-Aktionen ins Leere, Umstrukturierungen haben nicht den gewünschten Effekt, kurz, alle Beteiligten stochern im Nebel.

Das kostet Zeit, Geld und im harten Wettbewerb oft auch Kunden. Und im schlimmsten Fall wird aus dem Branchen-Primus, der nur auf sein Bauchgefühl setzt, dann schnell ein Auslaufmodell.

Dieser Artikel gibt einen Überblick darüber, wie Unternehmen am Besten mit ihren Daten verfahren, um so die Grundlage für gute Entscheidungen treffen zu können.

Jedes Unternehmen hat Datenqualitätsprobleme

Die schlechte Nachricht vorweg: Es gibt keinen Unternehmenstyp, der besonders anfällig ist für Datenqualitätsprobleme. Im Gegenteil: Alle Firmen haben damit zu kämpfen. Das zeigt sich natürlich besonders stark bei datenbezogenen Initiativen wie Enterprise Resource Planning (ERP) oder Customer Relationship Management (CRM). Die gute Nachricht: Es gibt erprobte Wege um diese Qualitätsprobleme in den Griff zu kriegen.

Gute Daten schlechte Daten

Es gibt viele Gründe für die explosionsartige Zunahme inkonsistenter oder nutzloser Daten. Angefangen von einfachen menschlichen Fehlern bis hin zu ungenügenden oder gar nicht vorhandenen unternehmensweiten Datenstandards, die alle Systeme, Geschäftsbereiche und gegebenenfalls Konzerngesellschaften erfassen. Um die Datenprobleme zu überwinden, benötigen Unternehmen eine technologische Grundlage mit der sie die Datenqualitätsprobleme aufdecken und beheben können. Doch Vorsicht: Datenqualitätsprobleme können nicht allein durch Technologie gelöst werden. Ebenso wichtig ist die Einbindung der richtigen Personen und Prozesse.

"Datenqualität und Data Governance erfordern die Einbindung von Menschen, Prozessen und Technologie."

Technologie ist nur ein Teil der Lösung, die zu konsolidierten, hochwertigen Unternehmensinformationen führt. Vergleichbar ist dabei etwa das Thema IT-Security. Auch in diesem Bereich schützen die besten technischen Systeme alleine nicht, wenn nicht auch Menschen und Prozesse dahinter stehen. Eine robuste Technologie, die jede Datenqualitätsinitiative unterstützt, ist also nur ein wichtiger Stützpfiler. Noch wichtiger für den Erfolg ist jedoch die richtige Methodik für den Einsatz dieser Technologie.

Das Unternehmen DataFlux, das als Lösungsanbieter für Datenqualität und -integrität seit Jahren Unternehmen zu diesem Thema berät, hat vor diesem Hintergrund eine eigene Methodik für Datenqualität und Data Governance in Unternehmen entwickelt.

Diese besteht aus fünf Komponenten oder Teilverfahren:

- Data Profiling
- Datenqualität
- Datenintegration
- Datenanreicherung
- Datenkontrolle

Diese Komponenten bieten eine Systematik, mit der sich der gesamte Prozess für eine verbesserte Datenqualität erfolgreich managen lässt. Nutzt man sie konsequent, profitiert das Unternehmen unter dem Strich von einem ganzheitlichen Überblick über alle relevanten Daten im gesamten Unternehmen, also z.B. in Bezug auf Daten zu Kunden, Produkten oder Lieferanten.

Data Profiling

Der erste Schritt ist das Data Profiling. Dieser Begriff steht für eine Ist-Analyse der vorhandenen Datenquellen. Ziel ist es, ein genaues Bild vom aktuellen Zustand der Daten zu erhalten.

Dabei werden im Einzelnen die folgenden drei Aspekte genauer betrachtet:

- Datenstruktur: Entspricht die Strukturierung der Daten den vorgesehenen Strukturen im Unternehmen? Passen die vorhandenen Daten zu den korrespondierenden Metadaten?
- Dateneinträge: Sind die Datenwerte vollständig, genau und eindeutig? Sind die Daten entsprechend der geltenden Regeln standardisiert?
- Datenbeziehungen: Weisen die Daten in allen Datensätzen die erforderliche Zuordnung der spezifizierten Schlüsselbeziehung auf? Bestehen abgeleitete Relationen über Tabellen und Datenbanken hinweg? Gibt es redundante Daten?

Ein Beispiel für die Ermittlung von Daten: Ergeben Stichproben bei einigen Einträgen in einer Datenbank, dass bestimmte Werte fehlen, dann ist das ein sicheres Indiz für die generelle Unvollständigkeit der Datenbank – und dahinter könnte System stecken. So könnte das bei einer Kundenadress-Datenbank darauf hindeuten, dass etwa die Postadressen nicht vollständig erfasst werden, etwa weil das Formular missverständlich oder fehlerhaft angelegt ist.

Nach dem Data Profiling erhalten Unternehmen auf diese Weise detaillierte und vor allem analytische Informationen über die Stärken und Schwächen ihrer Unternehmensdaten. Aus

diesen Kenntnissen lassen sich die weiteren Maßnahmen für eine Verbesserung der Datenqualität ableiten.

Die gewonnenen Informationen bilden den Ausgangspunkt für den zweiten Teil der Methodik, die Verbesserung der

Datenqualität

In diesem Schritt beginnt schon der Aufbau der Qualitätsdaten. Es werden hauptsächlich Fehler korrigiert, Informationen standardisiert und Daten unternehmensweit validiert.

Dabei empfehlen sich insbesondere die folgenden drei häufig angewandten Taktiken:

- **Datenstandardisierung** – Diese dient dazu, vielfältige „Permutationen“ von Daten herauszufiltern und zu korrigieren. Ein Beispiel: „ACME Manufacturing Corporation“ kann in derselben Datenquelle als „Acme Mftg Corp“, „ACME“ und „ACME Manufacturing“ auftauchen. Intelligentes Fuzzy Matching findet diese Varianten und macht eine einheitliche Nomenklatur daraus.
- **Musterstandardisierung** – Diese ermöglicht die Erstellung eines für alle Tabellen geltenden Datenmusters. Manche Datenbestandteile wie Telefonnummern oder Sozialversicherungsnummern sind in Musterabweichungen leicht zu finden. Spezielle Branchendaten, wie Produkt- oder Teilenummern können in Unternehmen jedoch unterschiedlichen Standards folgen. Die Musterstandardisierung bringt die Informationen aus nicht standardisierten Formaten in ein gewünschtes Standardformat.
- **Adressverifizierung**: Dieser Vorgang stellt sicher, dass Adressen gültig und nutzbar sind. So erkennt zum Beispiel die Software von DataFlux, wenn die angegebene Stadt nicht mit der Postleitzahl übereinstimmt. Die Software sucht dann selbständig nach postalischen Daten, findet die richtige Postleitzahl und ändert den Eintrag in eine korrekte Adresse.

Generell sollte die Entscheidung für eine der genannten Vorgehensweisen den jeweiligen Fachbereichen im Unternehmen überlassen bleiben, die für die spezifischen Daten verantwortlich sind. Wichtig ist dabei aber: Das Datenqualitätsverfahren verbessert die Qualität vorhandener Daten, es schafft jedoch nicht die Ursache des Datenproblems aus der Welt. Für eine nachhaltige Verbesserung der Datenqualität muss ein Unternehmen auch die Wurzel des Datenübels aufspüren, die erst zu den Fehlern führt. Danach gilt es, dass geeignete Gegenmaßnahmen ergriffen werden, um die Daten kontinuierlich auf einem höheren Qualitätsniveau zu halten. Das können beispielsweise Veränderungen in den Geschäftsprozessen sein. Ist der Datenqualitätsprozess abgeschlossen, verfügt die Firma bereits über die geeigneten Mittel, um die Daten aus jeder Quelle auf das angestrebte Qualitätsniveau zu bringen.

Der nächste notwendige Schritt besteht in der Zusammenführung der unterschiedlichen Quellen – in der Datenintegration.

Datenintegration

Steht eine Kundenadresse in mehreren Firmendatenbanken gleichzeitig, dann ist die Gefahr groß, dass sich dabei irgendwo ein Fehler eingeschlichen hat. Experten sprechen dabei von verschiedenen Datensätzen, die dasselbe Objekt repräsentieren. Die Datenintegration dient dazu, diese Daten aus verschiedenen Quellen zu optimieren. Abweichende Einträge in verschiedenen Datenbanken werden in diesem Schritt zu einem einzigen Stammdatensatz zusammengefasst. Beispielsweise könnte eine Firma zwei Datenquellen für ihre Produkte pflegen: die Originalprodukt Datenbank, extrahiert aus ihrem ERP-System in den USA, und

eine Produktdatenbank in Europa. Die Firma verkauft das gleiche Produkt in beiden Regionen, die Produktbezeichnungen in den Datenbanken sind jedoch unterschiedlich. Zusätzlich wurden die Produktnamen, Fabrikatsbezeichnungen und Produktbeschreibungen in jeder der Datenquellen von verschiedenen Mitarbeitern eingegeben.

Die erste Aufgabe der Datenintegration besteht jetzt darin, ein Produkt zu finden, das in beiden Quellen vorkommt (Linking). Die zweite Aufgabe: die Daten müssen in eine einheitliche Sicht gebracht werden. Beim Finden dieser mehrfach vorkommenden Objekte helfen Alleinstellungsmerkmale wie Produktidentifikationsnummern oder Steuernummern bei Kundendaten. Intelligentes Fuzzy Matching geht aber noch deutlich weiter und findet die Datensätze auch anhand von ähnlichen Produktbeschreibungen. Geht man zurück zum obigen Beispiel, dann wies die Datenbank in den USA eine Beschreibung, das Fabrikat und den Produktnamen aus. Zusätzlich gab es noch ein Feld mit verschiedenen Mustern. Die Datenbank in Europa enthielt nur die Produktbeschreibungen, diese aber in unterschiedlichen Mustern und Abkürzungen. Zur Konsolidierung der Einträge nutzte die Integrationstechnologie die folgenden Abläufe:

- Identifikation und Isolierung (Parsen) der produktspezifischen Attribute und der Fabrikatsbezeichnung aus der Beschreibung in der Datenbank in den USA
- Abgleich der Unterschiede in den Fabrikatsbezeichnungen
- Abgleich der Unterschiede in den Produktattributen (Kurzformen, Abkürzungen)
- Phonetische Anpassung der verglichenen Daten
- Erstellung eines Berichts zum Matching der Produkte

Nach der Datenintegration besitzt das Unternehmen jetzt die komplette Sicht über alle Daten zu einem Objekt. Die gleichen Prinzipien und Techniken können auf jeden Datentyp angewendet werden, um integrierte Stammdatensätze zu erstellen. Im nächsten Schritt soll der Wert der Daten durch Ergänzung um zusätzliche Informationen erhöht werden.

Datenanreicherung

Bei der Datenanreicherung werden die Datensätze ergänzt und zwar um den Informationswert zu erhöhen. Dies kann auf verschiedene Art und Weise erfolgen. So können beispielsweise die eigenen Daten durch Daten von Partnern oder anderen Firmen ergänzt und so genauere Informationen über Kunden aufgebaut werden – etwa zu ihrem Einkaufspotenzial oder ihrer Loyalität. In Form einer so genannten Geokodierung können inzwischen auch geografische Informationen integriert werden. Auf diese Weise lässt sich etwa feststellen, dass alle Häuser einer bestimmten Postleitzahl in Kombination mit einer bestimmten geografischen Region nach 1980 gebaut wurden und einen bestimmten Immobilienwert übersteigen. Daraus lassen sich dann wieder sehr spezifische Produktangebote für potenzielle Kunden in dieser Umgebung ableiten.

Datenanreicherung lässt sich aber nicht nur auf Adressdaten anwenden. Viele Arten von Daten, darunter Produkt-, Inventar- und Finanzdaten, können durch die Verbindung zu externen Quellen ergänzt und dadurch aufgewertet werden. So können internationale Warenwirtschaftsregister wie UNSPSC oder eCl@ss dazu dienen, die Produkte oder Lieferantendaten zu standardisieren und damit Einkauf und Verkauf zu rationalisieren. Die Integration dieser universellen Klassifizierung unterstützt die Genauigkeit der Lieferantendaten, was wiederum die Effizienz der Bestellungen erhöht und den Weg für reibungslose internationale Geschäfte ebnet.

Datenanreicherung unterstützt auch das von häufigen Neuerungen betroffene Compliance Management. Es ermöglicht den Vergleich von Kundenlisten oder Transaktionen mit externen Listen auf Übereinstimmung mit den gesetzlichen und industriellen Vorschriften. Derartige Listen können beispielsweise als Grundlage für die Bekämpfung von Geldwäsche oder Korruptionsdelikten dienen.

Datenkontrolle

Zu einem umfassenden Data Governance Programm gehört schlussendlich eine kontinuierliche Datenkontrolle. Nur so ist gewährleistet, dass der Datenqualitätserfolg aus den aufwändigen Korrektur- und Aufwertungsmaßnahmen nicht durch die allmähliche Rückkehr zu alten Gewohnheiten zunichte gemacht wird. Mit aktiver Datenkontrolle haben Unternehmen Kenntnis vom Zustand ihrer Daten und können die Ursachen für Qualitätsprobleme gezielt aufspüren und korrigieren. Datenkontrolle kann auf verschiedene Weise erfolgen:

Die einfachste Datenkontrolle automatisiert Data Profiling Reports, die in regelmäßigen Abständen Analysen der Daten in Bezug auf Ausnahmen und nicht standardisierte Daten liefern. Solche Reports identifizieren beispielsweise Auffälligkeiten wie einen zu großen Prozentsatz an außergewöhnlichen Daten, und können gegebenenfalls eine Mail oder eine Systemmeldung an zuvor bestimmte Adressaten oder Verantwortliche eines Bereichs schicken. Schneller und effektiver geht die Datenkontrolle vonstatten, wenn sie komplett über Geschäftsprozessregeln in Form von Web-Services realisiert wird. Mit dieser Methode können die gleichen Regeln, wie sie für die Bereinigung und Anreicherung der Daten benutzt wurden, dazu dienen, Daten in Echtzeit während der Eingabe oder auf ihrem Weg durch das Unternehmen zu überprüfen. Die Vorteile dieses Ansatzes umfassen die folgenden Möglichkeiten:

- Es können Data Governance Regeln aufgestellt und integriert werden. Damit ist gewährleistet, dass die unternehmensspezifischen Regeln für das Datenmanagement befolgt werden.
- Aufgabenkritische Prozesse werden erkennbar, indem Verstöße gegen oder Ausnahmen von Geschäftsregeln in einem eigenen Repository protokolliert werden. So lassen sich Trends verfolgen und die Prozesse entsprechend anpassen.
- Es können definierte „Ereignisse“ ausgelöst werden, die eine automatische Datenkorrektur auslösen oder vorab registrierte, ausgewählte Mitarbeiter (z.B. in Form von Mails) informieren und zu bestimmten „Aktionen“ auffordern – etwa dazu, die Daten in einem Repository zu protokollieren oder die korrigierten Daten in eine Datenbank einzutragen.

Zusammenfassung

Die Qualität jeder Geschäftsentscheidung ist nur so gut wie die Daten, die ihre Grundlage bilden. Ohne Daten, die unternehmensweit konsistent, akkurat und zuverlässig sind, kann ein Unternehmen leicht irreführende, falsche und deshalb geschäftsschädigende Schlussfolgerungen ziehen. Für eine erfolgreiche Verbesserung der Datenqualität sollten Unternehmen eine stimmige Methodik verwenden, die im Idealfall die folgenden fünf Bestandteile umfasst: Data Profiling, Datenqualität, Datenintegration, Datenanreicherung und Datenkontrolle.

Individuelle Geschäftsprozessregeln, die im Verlauf eines solchen Verfahrens entwickelt wurden, dienen anschließend als „Datenkontrolle“ zur Sicherung der neu gewonnenen

Datenqualität, denn Datengüte ist kein punktuell Ereignis – im Gegenteil: Unternehmensdaten müssen dauerhaft analysiert, optimiert und kontrolliert werden. Ziel jedes Unternehmens sollte es also sein, die Qualität seiner Datenbestände langfristig zu sichern. Die Erkenntnis für die Bedeutung dieser Thematik sowie die Bereitschaft, sich kontinuierlich damit auseinanderzusetzen, muss aus dem Unternehmen selbst kommen und betrifft nicht nur die Geschäftsführung, die IT oder einzelne Fachabteilungen, sondern stets alle Mitarbeiter. Ist dieses Commitment aber erst einmal gegeben, muss jedoch kein Unternehmen mehr das Rad neu erfinden – längst bietet der Markt ausgefeilte, erprobte und komfortable Kombinationen aus Beratung und Technologien an, die das Unternehmen bei dieser wichtigen Entwicklung unterstützen.