

# COMPUTERWOCHE

Sonderdruck aus COMPUTERWOCHE ONLINE vom 11. März 2014



Wie qualitativ hochwertig sind Ihre Daten?

Foto: violetkaipa, Fotolia.com

## Datenqualität neu denken

**Eine schlechte Datenqualität kann sich heute kein Unternehmen mehr leisten. Für eine individuelle Kundenansprache braucht es korrekte Informationen – womit viele Verantwortliche noch immer zu kämpfen haben.**

Von Holger Stelz

Für die meisten Unternehmen waren Daten aller Art bisher entweder ein Nebenprodukt ihrer Geschäftstätigkeit oder ein notwendiges Übel, das in erster Linie dort zu bewältigen war, wo die Daten anfielen - meist in den jeweiligen Abteilungen. Im Zuge der Big-Data-Debatte rückt nun allerdings der geschäftliche Nutzen von Daten zunehmend in den Fokus des Interesses. Unternehmen stehen vor der Herausforderung, das wachsende Datenaufkommen technisch zu bewältigen und in ihren Geschäftsprozessen effizient zu nutzen. Dazu gehört auch, den Erfolg eines datengetriebenen Geschäfts messbar zu machen.

In diesem Zuge fördert der Bedeutungswandel von Daten allerdings oft ein großes Defizit zu Tage: die schlechte Datengrundlage. Diese wurde in einzelnen Abteilungen bisher entweder zähneknirschend hingenommen oder wollten auch gar nicht wahrgenommen werden. Insbesondere im Kontakt mit den Kunden wird der Einfluss von korrekten und vollständigen Daten auf den Geschäftserfolg schnell sichtbar. Kunden erwarten von Unternehmen, dass diese über die Geschäftsbeziehung an jedem Kontaktpunkt vollständig und aktuell informiert sind. Insbesondere kundenzentrierte Unternehmen sind stark gefordert.

In der Praxis lässt sich diese Erwartung wegen Mängeln im Datenbestand und beim Datenmanagement allerdings oft nicht

erfüllen. Und auch nach innen hat mangelhafte Datenqualität schnell harte Auswirkungen, die sich in Euro und Cent beziffern lassen: Dem Vertrieb entgehen Cross- und Up-Selling-Potenziale, das Marketing verbrennt durch falsche Zielgruppensegmentierung unnötig Geld und die Buchhaltung erstellt womöglich fehlerhafte Rechnungen, die Kunden verärgern. Diese und weitere Szenarien könnten sich jedoch durch einen einheitlichen, vollständigen und aktuellen Informationsstand über die Kunden vermeiden lassen.

### Was ist Datenqualitäts-Management?

Vieles hängt also an der Datenqualität. Es gilt sie zu verbessern und laufend zu managen. Das ist keine Einmalaufgabe, denn fast alle Daten in Unternehmen, Kundendaten zumal, unterliegen ständigen Veränderungen. Ziel muss es deshalb sein, sicherzustellen, dass Kundeninformationen konstant einheitlich, vollständig und aktuell vorliegen. Dennoch verbessern Unternehmen ihre Datenqualität meist nur phasenweise, weil etwa ein neues Projekt einen Anlass dazu bietet (und entsprechende Budgets zur Verfügung stehen). Im Anschluss wird die Datenqualität aber meist wieder schlechter. Das liegt in der Natur der Sache, denn Daten ändern sich durch neue Gegebenheiten, beispielsweise durch die Änderung der Mobilfunknummer oder der Adresse.

Ist die Rede von einem Management der Datenqualität, handelt es sich um eine Vorgehensweise, die die Datenqualität über den gesamten Lebenszyklus der Daten hinweg sicherstellt – von der Erfassung über ihre Speicherung und Nutzung bis zu ihrer Archivierung und Löschung. Zum Einsatz kommt dabei gemeinhin der Regelkreisansatz („closed loop“) aus dem Total Quality Management. Zu Beginn werden die Kundendaten bereits während der Datenerfassung mittels Datenqualitäts-Services geprüft. Fehlerhafte Kundendaten, die nicht automatisch bereinigt werden können, werden in einer Zwischendatenbank gespeichert und ein Bericht oder Alert an die Eingabestelle geschickt, so dass diese korrigierend eingreifen kann.

Mit diesem Kreislauf lassen sich Kundendaten kontinuierlich während der Erfassung und Bearbeitung überprüfen. Werden über diese Prozesse regelmäßig Berichte verfasst (etwa über ein Data Quality Dashboard), können Anwender die Leistung des geschlossenen Regelkreises für das Datenqualitäts-Management messen (Performance Management) und der Prozess kontinuierlich verbessern. Das Ergebnis ist eine nahezu konstante Datenqualität auf hohem Niveau.

Datenqualität hört an dieser Stelle aber nicht auf. Unternehmen sind überwiegend so strukturiert, dass das Datenqualitäts-

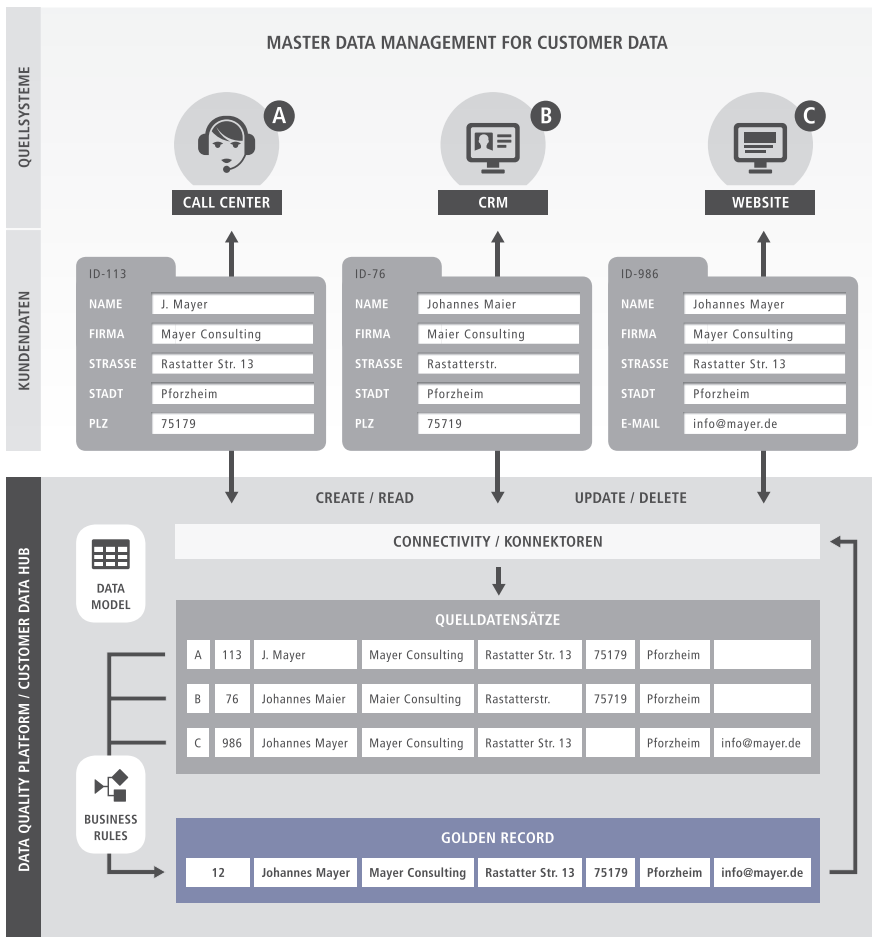
Management einen übermäßigen Aufwand erzeugt, weil die Datenhoheit meist bei den Abteilungen liegt. So können unterschiedliche Abteilungen oder neu erschlossene Geschäftsfelder nicht auf alle Kundendaten im Unternehmen zugreifen. Die Datentöpfe passen nicht zusammen. Das Management der Datenqualität bleibt in solchen Konstellationen auf voneinander getrennte Systemsilos beschränkt. Diese beherbergen zwar eine Menge an Kundendaten, die durch Zusammenführung mit unternehmensweit verfügbaren Daten in ihrer Qualität verbessert und angereichert werden könnten. Doch de facto verursachen die bestehenden Strukturen und Prozesse hohe Kosten durch Redundanzen.

Und was noch schwerer wiegt: Unternehmen verspielen das große Potenzial, das in ihren Datenbanken liegt – nämlich die Chance auf eine einheitliche Sicht auf ihre Kunden. Die Realität ist ernüchternd, denn den Unternehmen fehlt es an Überblick und ihr Management kann sich kaum auf die Daten als Basis für Entscheidungen und Maßnahmen verlassen. Fehlentscheidungen und Fehlinvestitionen können die teure Folge sein. Die Notwendigkeit für Datenqualität ist offensichtlich. Damit der Erfolg eines umfassenden Datenqualitäts-Managements aber tatsächlich auf die tägliche Arbeit der Mitarbeiter und den Geschäftserfolg des gesamten Unternehmens durchschlägt, ist ein Master Data Management notwendig.

**Silos aufbrechen: Der Golden Record**

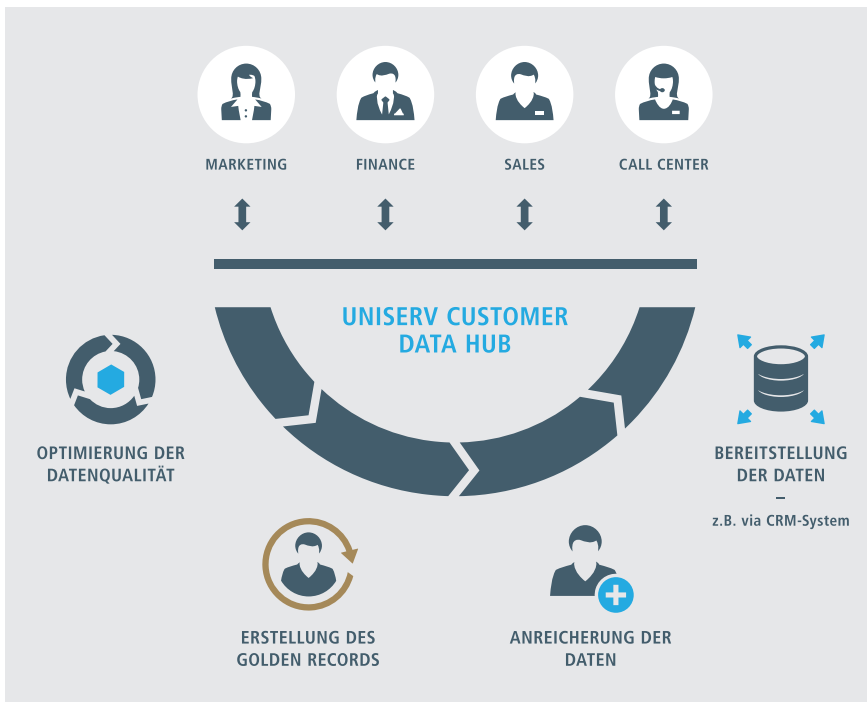
Unternehmen müssen die Datensilos auflösen, indem sie die Verwaltung ihrer Kundendaten zentralisieren. Ziel ist der „Golden Record“ – gewissermaßen die Mutter aller Datensätze, der einzig wahre Datensatz auf den sich alle operativen Systeme beziehen sollen. Technisch gibt es verschiedene Ansätze, wie Unternehmen zum Golden Record gelangen können. Der aufwändigste ist die vollständige Ablösung der einzelnen abteilungsspezifischen Datensilos durch eine zentrale Datenbank. Die Komplexität solcher Projekte ist allerdings enorm – nicht wenige scheitern aufgrund widerstrebbender Interessen im Unternehmen und unerwarteter technischer Probleme.

Andere Lösungen setzen auf eine Abstraktionsebene, die den Golden Record gleichsam virtuell erstellt und den einzelnen Anwendungen nur Verknüpfungen zu dem Quelldaten anbietet. Das Verfahren ist zwar agiler als die Komplettintegration und schafft den einheitlichen Blick auf den Kunden, ändert aber nur wenig an der Ausgangslage mit vielen Datensilos und schwankender Datenqualität.



Mit Master Data Management zu besseren Daten.

Foto: Uniserv



Bereits während des Eingabevorgangs können Daten auf festgelegte Qualitätskriterien hin überprüft werden.

Foto: Uniserv

gleich, ob dieser Kunde bereits in der Datenbank vorhanden ist. Damit vermeiden Unternehmen Dubletten – eines der wohl häufigsten Probleme innerhalb von Kundendatenbanken. Um den Kunden gegebenenfalls einem vorhandenen Stammdatensatz zuzuordnen, müssen Datensätze mit anderen Schreibweisen, etwa Johannes Mayer oder J. Maier, abgeglichen und überprüft werden.

Andere Eingabefehler lassen sich auf ähnliche Weise durch automatisierte Vorgänge vermeiden: Vergisst der Mitarbeiter etwa bei der Eingabe von neuen Bankdaten (IBAN) eine Ziffer, wird er darüber informiert, bevor er zum nächsten Datensatz geht. Das gleiche gilt für festgelegte Regeln, auch für Rechtschreibung und Grammatik. Standardisierte Informationen wie Postleitzahlen können auf ihre Korrektheit überprüft werden, indem sie mit länderspezifischen Vorgaben abgeglichen werden (das sogenannte Matching).

Der dritte – hybride – Ansatz kombiniert die Agilität der Datenvirtualisierung mit einer transaktionalen Komponente, die qualitätsgesicherte Daten zentral vorhält, diese aber auch bei Bedarf in die Quellsysteme zurückschreiben kann. Der Golden Record erfüllt bei diesem „Hybrid Master Data Management“-Verfahren (Hybrid MDM) seine Funktion als einheitlicher, vollständiger und aktueller Referenzpunkt an zentraler Stelle. Entsprechend kann er auch im Zuge von Data Governance Initiativen als Referenz für Compliance-konforme Prozesse dienen. Die Abteilungen, welche die Quellsysteme betreiben, haben aber die Wahl, ob sie die qualitätsgesicherten Daten zurückschreiben wollen oder nicht. Entsprechend höher sind die Erfolgsaussichten von Hybrid-MDM-Projekten.

Datenqualität unternehmensweit zu begreifen, unterscheidet sich grundlegend von herkömmlichen Datenqualitätsprojekten. Diese nehmen die meist abteilungsspezifischen Einfallstore für neue Daten ins Visier. Entsprechend fragmentiert bleiben die Erfolge. Der zentrale Ansatz und die Herangehensweise im Rahmen eines Master Data Managements sind eher dazu geeignet, das „Datenkapital“ für das betreffende Unternehmen deutlich zu mehren, weil die höhere Qualität allen Datennutzern zur Verfügung steht.

### Datenqualitäts-Management in der Praxis

So lassen sich Daten auf festgelegte Qualitätskriterien schon während der Eingabe eines Datensatzes, also in Echtzeit, automatisch prüfen: Wird beispielsweise ein neuer Kunde namens Johannes Mayer angelegt, überprüft ein automatischer Prozess

### Schritt für Schritt zu besseren Daten

Unabhängig von ihrer Struktur und Organisation, sollten Unternehmen bei der Überprüfung ihrer Datenqualität Schritt für Schritt vorgehen. Zunächst müssen zwei Bereiche, die eng miteinander verflochten sind, bewertet werden: Die Datenqualität und die damit verbundenen Prozesse. Da Daten in Geschäftsprozessen genutzt werden, beschreiben Prozesse ihren Verwendungszweck und geben das notwendige Datenformat vor. Liegen etwa qualifizierte Leads in einer Excel-Tabelle (Format) vor und sollen für Mailings eingesetzt werden (Prozess), muss das Format entsprechend angepasst werden. Dazu gehört, dass Datenfelder ergänzt, angereichert und so kodiert werden, dass sie für das Mailing-Tool importfähig sind. Dementsprechend hat jeder Prozess Einfluss auf die Datenqualität. In der Praxis ist der Verwendungszweck von Daten und somit der weitere Prozess vielen Mitarbeitern häufig gar nicht bekannt; der Mangel an Bewusstsein dafür sorgt allein schon für Fehler.

#### 1. Symptome und Ursachen erkennen

Anhaltspunkte einer jeden Datenqualitätsbewertung sind Symptome, die analysiert werden müssen. Beklagen Mitarbeiter beispielsweise einen hohen manuellen Rechercheaufwand oder Nacharbeit bei der Erstellung von Kundenlisten, müssen diese Klagen als Symptome erkannt werden. Die Ursache kann in der fehlenden Aktualität von Daten liegen oder in fehlendem Vertrauen der Mitarbeiter in die Daten. Während nun also von un-

zureichender Datenqualität auszugehen ist, muss auch überprüft werden, ob bestehende Prozesse darauf Einfluss nehmen.

Ein anderes Beispiel ist das mehrfache Auftauchen von Kunden und Geschäftspartnern in datenführenden Systemen (Dubletten). Die Analyse der Prozesse kann hier etwa ergeben, dass Mitarbeiter vor dem Anlegen neuer Kunden nicht überprüfen, ob diese schon im System vorhanden sind.

Um Symptomen und Ursachen auf die Spur zu kommen, sind möglichst Mitarbeiter, die regelmäßig mit den Daten arbeiten, in den Prozess mit einzubeziehen. Aus den jeweiligen Symptomen und Ursachen ergeben sich die zu analysierenden Datenobjekte und -prozesse. Zudem können und sollten die Verantwortlichen daraus ableiten, wie Unternehmensstammdaten aussehen müssen, damit sie alle damit arbeitenden Mitarbeiter effizient nutzen können.

## 2. Profiling

Nun gilt es, das „Bauchgefühl“ und Erfahrungswissen der Mitarbeiter zum Status der Datenqualität durch Analyseergebnisse zu belegen. Ziel ist es, den Status Quo der Daten aufzuzeigen und eine fundierte Aussage über Vollständigkeit, Korrektheit und Redundanz der Daten zu treffen. Das Ergebnis dieses Profiling sollte folgende Fragen beantworten:

- Welche Informationen fehlen?
- Wo tauchen verdächtige Ausreißer auf?
- Wo passt das Format nicht zur Bedeutung?
- Wo sind zwei oder mehrere Attribute nicht konsistent?
- Wo werden vorgegebene Regeln verletzt?
- In welchem Kontext treten Fehler auf?
- Wie stellt sich ein Fehler in unterschiedlichen Datensegmenten dar (verschiedene Regionen, unterschiedliche Erfassungszeiträume etc.)?
- Wie verändert sich das Auftreten von Fehlern im Lauf der Zeit?

Um diese Fragen zu beantworten, empfiehlt sich der Einsatz von Software, die automatisiert nach Lücken, Fehlern und Zusammenhängen innerhalb eines Datenbestands sucht, diese zusammenträgt und damit die Grundlage für das Analyseergebnis liefert.

Das Profiling ist ein entscheidender Schritt – übergehen Unternehmen diesen Punkt, was häufig der Fall ist, treffen sie ihre Maßnahmen zur Verbesserung der Datenqualität auf gut Glück. Denn wenn die Verantwortlichen ihre Datengrundlage nicht eindeutig kennen, veranlassen sie Maßnahmen zur Verbesserung der Datenqualität nach Bauchgefühl, gehen aber nicht allen Ursachen nach. Als Folge bleiben ihre Daten unzuverlässig und Mitarbeiter müssen weiterhin holprige Prozesse und schlechte Datenqualität hinnehmen.

## 3. Maßnahmenkatalog erstellen

Auf Basis des Profiling werden Maßnahmen festgelegt, die die Daten- und Prozessqualität auf höchstes Niveau bringen sollen. Das kann sein, den Verwendungszweck von Daten sowie die Anforderungen an sie festzuschreiben oder einen Real-Time-Adress-Check einzuführen. Möchte man Dubletten in den Griff bekommen, könnte eine Echtzeit-Dublettensuche die Lösung sein. Die Bandbreite der möglichen Maßnahmen ist – genauso wie die der Anforderungen an Daten und Prozesse – groß und hängt immer vom jeweiligen Fall ab.

## 4. Bereinigen, Prozesse anpassen

In jedem Fall muss das Unternehmen seinen untersuchten Datenbestand bereinigen, also fehlerhafte Daten entfernen und korrigieren. Dies kann beispielsweise die automatisierte Überprüfung von Telefonnummern sein – eine wichtige Maßnahme und Unterstützung für CRM-Projekte, den Helpdesk, Beschwerdemanagement oder andere Aufgaben des Kundenkontaktsmanagements. Dort werden Dubletten automatisch beseitigt. Ein aktuelles Beispiel ist die Konvertierung von Kontodaten nach alter Systematik (Kontonummer und Bankleitzahl) in das neue SEPA-Format mit IBAN und BIC.

Neben der Bereinigung der Daten müssen häufig Prozesse wie auch Systemlandschaften angepasst werden. Sind alle Bereiche entsprechend justiert, hat die Datenqualität im Idealfall das maximal erreichbare Niveau. Ob es konstant bleibt, hängt indes vom weiteren Daten- und Qualitäts-Management im gesamten Unternehmen ab. (ba/sh)

## Sonderdruck aus COMPUTERWOCHE ONLINE vom 11. März 2014 für

### Uniserv GmbH

Rastatter Straße 13  
75179 Pforzheim

Telefon: +49 7231 936-0  
Telefax: +49 7231 936-3002  
E-Mail: [info@uniserv.com](mailto:info@uniserv.com)



Uniserv ist der größte spezialisierte Anbieter von Lösungen für das Data Management in Europa. Data Management vereint Datenqualitätssicherung und Datenintegration zu einem ganzheitlichen Ansatz. Kundendaten stehen dabei im Mittelpunkt von Initiativen für Datenqualität, Datenmigration, Data Warehousing sowie Master Data Management, beispielsweise im Umfeld von CRM-Anwendungen, eBusiness, Direct- und Database-Marketing, CDI/MDM-Anwendungen und Business Intelligence. Mit mehreren Tausend Installationen weltweit bedient Uniserv die Erwartungshaltung einer ganzheitlichen Lösung für alle Geschäfts- und Kundendaten über den gesamten Datenlebenszyklus hinweg. Am Stammsitz in Pforzheim sowie in den Niederlassungen in Paris, Frankreich, und Amsterdam, Niederlande, beschäftigt das Unternehmen 133 Mitarbeiter und zählt branchenübergreifend und international zahlreiche renommierte Unternehmen wie beispielsweise Allianz, Deutsche Bank, eBay, EDEKA, E.ON, France Telecom, Lufthansa, Otto, Siemens, Time Warner sowie TUI und VOLKSWAGEN zu seinen Kunden. Mehr Informationen unter [www.uniserv.com](http://www.uniserv.com).