

Datenmanagement und Datenqualität

Erschienen in: Direkt Marketing 07/2001

Durch die Einführung von CRM in Unternehmen bekommen auch die unternehmenseigenen Daten zum Kunden eine neue Bedeutung. Vorbei sind die Zeiten, in denen Datenmanagement nur die richtige Adresse im Mailing zu benutzen bedeutete. Die heute angestrebte individuelle Kundenkommunikation benötigt mehr und besser qualifizierte Informationen zum Kunden und Kundenverhalten. Das Datenmanagement bildet dabei die Grundlage, um mit intelligenten Methoden z.B. des Data Minings dieses umfassende Wissen bereitzustellen und den Kunden in differenzierte Zielgruppen einzuteilen. Daraus resultierend ist eine optimale Kommunikation zum Kunden aufzubauen.

Aber häufig stehen den betroffenen Abteilungen gar nicht die nötigen Informationen zum Kunden- und Firmenverhalten zur Verfügung, um optimal auf die derzeitigen und kommenden Anforderungen/Wünsche der Kunden eingehen zu können. Leider ist es heute so, dass in den Firmen zwar viele Informationen gesammelt worden sind, aber diese Informationen sind in unterschiedlichen Bereichen und Abteilungen gespeichert und dadurch für andere Abteilungen bzw. unternehmensübergreifend nicht oder nur schwer nutzbar. Ein weiteres Problem stellt die Tatsache dar, dass viele Informationen aufgrund der Art und Weise, wie sie jetzt gespeichert werden, zum Teil nur schwer analysierbar sind. Das liegt zum einen an der Datenhaltung, die zumeist für die Erledigung des operationalen Tagesgeschäftes optimiert wurde. Zum anderen liegt es häufig auch daran, dass Informationen z.B. als Textfelder vorliegen. Ein Beispiel dafür ist die Speicherung von Beschwerden.

Es geht also darum, die unterschiedlichen Techniken zur Informationsspeicherung, -verwaltung und Analysen im Unternehmen so einzusetzen, dass Kunden optimal bedient werden können.

Die Basis dieser komplexen Systeme bildet häufig eine Kundendatenbank oder ein unternehmensweites Datawarehouse, das alle relevanten Daten für die Analysen und Selektionen bereitstellt.

Um rasch Nutzen aus einem unternehmensweiten Datawarehouse ziehen zu können, zerteilt man das Datawarehousegesamtpjekt in einzelne Phasen. Diese Phasen orientieren sich an der Bereichs- bzw. Informationsstruktur eines Unternehmens, wobei schon in Phase 1 der Blick auf das Gesamte nicht vergessen wird bzw. die Informationen, die alle Bereiche eines Unternehmens gleichzeitig benötigen, werden schon in Phase 1 realisiert, während in späteren Phasen nur noch die bereichstypischen Informationen ergänzt werden. Konkret heißt das man startet zum Beispiel mit dem Bereich Marketing und entwickelt eine kleine Lösung, die sich vor allem am Informationsbedürfnis dieses Bereiches orientiert. D.h. die Informationen und Prozesse rund um den Kunden und seine Kauftransaktionen werden sehr ausführlich abgebildet (z.B. Alter, Geschlecht, erhaltene Werbemittel/ Besuche, aktuelle Kaufdaten, historische Kaufdaten, usw.), während zum Beispiel Detailinformationen zum Produkt wie besondere Herstellungs- oder Lieferdaten, soweit sie vom Marketing nicht benötigt werden (z.B. Lieferantenartikelnummer, Lagerplatz, Gewicht, Volumen, u.ä.) nicht abgebildet werden. Gleichzeitig werden an solchen Stellen im Datawarehouse Schnittstellen eingebaut, die eine spätere Erweiterung zum Beispiel in Richtung Einkauf, Produktion oder Logistik ermöglichen, um den speziellen Informationsbedarf dieser Bereiche zu einem späteren Zeitpunkt ins Datawarehouse einbinden zu können. Durch diese Vorgehensweise wird sicher gestellt, dass das Projekt in jeder Phase überschaubar und bezahlbar bleibt. Zudem können schon mit Abschluss der ersten Phase Informationen und Gewinn aus dem Datawarehouse gewonnen werden.

Damit ein Datawarehouse seine Aufgaben erfüllen kann, ist es besonders wichtig, dass man schon im Vorfeld des Datawarehouseprojektes sich sehr genaue Gedanken über die Nutzung macht. Noch besser ist es, wenn sich der Unternehmensvorstand mit der strategischen Ausrichtung und den strategischen Erwartungen an das Datawarehouse selber auseinandersetzt. Die Fachabteilungen müssen von Anfang an in das gesamte Projekt mit einbezogen werden.

Der Erfolg eines Datawarehouse misst sich vor allem an der Zufriedenheit seiner Benutzer.

Doch leider steht viel zu oft die Auswahl der Produkte im Vordergrund der Projekte. Von einer koordinierten, abteilungsübergreifenden Planung unter Mitarbeit der Fachabteilungen/Anwender sind viele Datawarehouse Projekte weit entfernt.

Es empfiehlt sich deshalb neben der Hard- und Softwareauswahl folgendes schrittweises Vorgehen zusammen mit den Projektmitgliedern der Fachabteilungen, um ein Datawarehouse zu erstellen.

- Erarbeiten der strategischen Ziele des Datawarehouses
- Klärung des Informationsbedarf aller späteren Anwender zur Bewältigung ihres Tagesgeschäftes und zur Umsetzung der strategischen Ziele
- Analyse der bestehenden und zukünftigen Geschäftsprozesse des Unternehmens und der Datenbestände des operativen EDV-Systems und Analyse möglicher externer Daten
- Definition der Datenqualität für das Datawarehouse
- Erarbeiten und Implementierung des Datenmodelles und Definition der Metadaten (Metadaten sind vergleichbar mit einem sehr ausführlichen Inhalts-/Quellenverzeichnis des Datawarehouses). Dieses Datenmodell muss ausbaubar, flexibel und den Geschäftsprozessen angepasst sein
- Aufbereitung und Übernahme der Daten aus den operativen Systemen ins Datawarehouse für den Prototypen des Datawarehouse
- Testen des Prototyp-Datawarehouse durch die Fachabteilungen und gegebenenfalls Überarbeitung und Weiterentwicklung
- Aufbereitung und Übernahme der Daten aus den operativen Systemen ins Datawarehouse für das Echt-System
- Tuning des Datawarehouses für die jetzt bekannten hauptsächlichen Abfragen/Analysen
- Nach Inbetriebnahme des Datawarehouse muss eine konstante und konsequente Anpassung des Datenwarehouses an die sich entwickelnden Geschäftsprozesse und Aufgaben/Analysen erfolgen

Aus allen teilweise sehr unterschiedlichen operativen Systemen eines Unternehmens werden die relevanten Daten für das Datawarehouse auf ihre Qualität geprüft und dann gegebenenfalls aufbereitet in das Datawarehouse geladen. Im Gegensatz zu den operativen Systemen spielt die Abbildung von Vergangenheitsdaten und externen Daten im Datawarehouse eine große Rolle.

Aber das Sammeln alleine hilft nur bedingt weiter, denn die Datenqualität des Datawarehouse bestimmt nachhaltig seinen Nutzen. Fehler in den Datenbeständen wie z. B. Rechnungen ohne Kundenstammdaten (der Kundenstammdatensatz wurde gelöscht, da z.B. der Kunde seit x Jahren inaktiv ist, die Rechnung muss aber noch auf Grund der gesetzlichen Bestimmungen archiviert werden), Intradoubletten, unvollständige und inkonsistente Daten können die Aussagekraft der späteren Analysen in Frage stellen. Deswegen ist es notwendig in diesen Phasen besonders die anderen Fachabteilungen mit einzubeziehen und einen Qualitätsstandard zu erarbeiten, der nicht nur bei der Erstbefüllung sondern auch im späteren Betrieb des Datawarehouses Bestand hat.

In dieser Phase des Projektes ergibt sich deshalb häufig der Effekt, dass es eine bereinigende Rückkoppelung aus dem Projekt in Richtung operative Systeme gibt. Das kann z.B. so aussehen, dass für die weitgehend dublettenfreie Kundenneuanlage eine Verbesserung des Matchcodesystems durchgeführt wird, endlich die Daten der Interessenten oder Katalogretouren auch permanent gespeichert werden oder aber auch dass man die Erfassungsmasken so gestaltet, dass Eingabefehler reduziert werden.

Besonders das Thema Intradoubletten, der Umgang mit inaktiven Kunden und Interessenten birgt hier Schwierigkeiten. Im operativen Geschäft sind z.B. Doubletten häufig kein Problem. Das gleiche gilt für inaktive Kunden. Beides hat keinen oder nur wenig Einfluss auf den laufenden Betrieb. Deshalb haben viele Firmen dieses in der Vergangenheit häufig nicht weiter beachtet. Für die Bewertung eines Kunden z.B. in Bezug auf die Selektion eines Mailings kann es aber sehr wohl wichtig sein möglichst viele Informationen zu einem inaktiven Kunden zu haben beziehungsweise Detailinformationen zum Interessenten in die Bewertung mit einbeziehen zu können. Der mehrfache Versand von ein und demselben Mailing an den gleichen Kunden kostet nur unnötiges Geld und verärgert den Kunden. Man sollte also ein solches Projekt auch nutzen, um die bestehenden Prozesse des operativen Geschäftes zu untersuchen und gegebenenfalls zu verbessern.

Auch das Thema Nutzung von externen Daten zur Anreicherung birgt neben der datenschutzrechtlichen Seite auch auf der Anwendungsseite eine Menge von Tücken. Das beginnt mit der Anreicherung selbst. Auf welchen Ebenen können die Daten den eigenen Beständen zugespielt werden, auf PLZ- Niveau, über den Gemeindegemeinschaftsschlüssel oder über Adressabgleich, wobei besonders die Anreicherung über den Adress-/Doublettenabgleich besonders problematisch ist, da over- bzw. underkills einen nicht zu unterschätzenden Einfluss auf die Aussagekraft der späteren Analysen haben. Hier ist zu bedenken, dass für die unterschiedlichsten Datenbestände abhängig von Nutzungszweck häufig ganz andere Parameter im Doublettenabgleich eingestellt werden müssen. Da man nicht davon ausgehen kann, dass zu jedem internen Kundensatz auch externe Daten vorhanden sind, muss für das Thema der fehlenden Werte eine umsetzbare Lösung gefunden werden. Ein weiterer häufig verdrängter Aspekt ist die Tatsache, dass bei externen Daten die Art und Weise der Erhebung sowie die Aktualität von den eigenen Daten abweicht, so dass auf Basis von beiden Datenbeständen getroffene Aussagen immer auch vor diesem Hintergrund gesehen und interpretiert werden müssen.

Die Befüllung des Datawarehouses erfolgt zum erstenmal während des Projekts als Initialbefüllung, das heißt es werden alle für die Datenbank als relevant definierten Daten mit dem Stand XY in die Datenbank geladen. Während des Ladens wird ein Teil der Daten noch gemäß der Definitionen transformiert bzw. aggregiert z.B. aus der Anrede wird eine Variable Geschlecht erstellt.

Darüber hinaus werden Routinen implementiert, die die regelmäßige Befüllung bzw. das Updaten sicherstellen. Je nach Einsatzgebiet entscheiden sich die Firmen meistens für einen täglichen oder wöchentlichen Rhythmus.

Neben den Selektionen sind die verschiedensten Analysen die Haupteinsatzgebiete der Datenbank.

Bei den Analysen sollte man grundsätzlich zwischen den mehr kreativen forschenden Analysen des Data Mining unterscheiden und den standardisierbaren, regelmäßig wiederkehrenden Analysen wie z.B. Werbeerfolgskontrollen, Tagesstatistiken oder typische Fragestellungen des Berichtswesen unterscheiden. Diese Analysen erfolgen häufig mit sogenannten OLAP-Tools.

So stellt es kein Problem dar, die Anzahl oder das Durchschnittsalter neuer Kunden oder Interessenten auszugeben. Während die Nutzung der Daten zur Beantwortung entscheidungsrelevanter Fragen einen großen strategischen Vorteil darstellt: Beispiele für solche Fragen sind:

- Welchen Kunden sollte wann welches Angebot unterbreitet werden?
- Welche Kunden sind gefährdet?
- Wie hoch ist das Cross-Selling-Potential für ein neues Produkt?
- Welcher Lifetime-Profit lässt sich mit welchem Kunden erzielen?
- Wie lassen sich Top-Interessenten mit hohen Lifetime-Values gewinnen?
- Welcher Umsatz wird im nächsten Jahr erzielt?

Kernfrage ist: Warum blieben gerade diese, für das Management entscheidende, Fragen häufig unbeantwortet? Betrachtet man den Charakter der Fragestellungen, so liegt die Antwort auf die aufgezählten Fragen nicht in einem einzelnen Datenfeld oder einem Kundenmerkmal, sondern in der richtigen Kombination unterschiedlicher Kundeninformationen. So kann bspw. die Angebotsaffinität eines Kunden von einer Vielzahl von Merkmalen wie Alter, Geschlecht, Familienstand, demographischen Typologien, bisher gekauften Produkten, gezeigtem Produktinteresse, Zahlungsmoral und einer Reihe weiterer Eigenschaften abhängen.

An dieser Stelle setzt das Data Mining an.

Beim Data Mining unterscheidet man grundsätzlich zwei Vorgehensweisen:

- Validieren von Hypothesen auf den Daten
- Entdecken und Entwickeln von bisher unbekanntem Muster/Regeln in den Daten

Während für die erste Vorgehensweise schon vorab eine entsprechende Hypothese entwickelt werden muss, versucht man in der zweiten Vorgehensweise neue Muster/Regeln in den Daten mit Hilfe der geeigneten Verfahren zu finden. Wobei natürlich grundsätzlich gilt, nur Informationen, die den Analyseverfahren zur Verfügung gestellt werden, können auch bei der Entwicklung der neuen Regeln berücksichtigt werden. Als Beispiel: Kein Verfahren kann einen irgendwie gearteten Zusammenhang zwischen Geschlecht und Kaufverhalten feststellen, wenn in dem zu analysierenden Datenbestand keine Variable für Geschlecht, Anrede oder die Vornamen der Kunden abgelegt sind, oder der gesamte Datenbestand nur Kunden eines Geschlechts enthält.

Technisch ist man heute soweit, dass man Kundenverhalten genau vorhersagen kann und somit die Grundlage für die Aktionen schafft, um bereits bestehende Kunden zu halten und neue zu gewinnen.

In Kürze:

- Um die Kunden optimal zu bedienen und den CRM-Gedanken mit Leben zu füllen, sollten alle zur Verfügung stehenden Informationen zum Kunden genutzt werden.
- Die internen und externen Informationen zum Kunden sollten zusätzlich zum operativen Geschäft so gespeichert werden, dass sie optimal für die Analysen zur Verfügung stehen. (Kundendatenbank, Datawarehouse)
- Während des Aufbaus eines solchen Datawarehouses muss die jeweilige Fachabteilung mit einbezogen werden und die Chance genutzt werden eventuell die Abläufe mit Tagesgeschäft zu optimieren.

- Der Erfolg eines Datawarehouse misst sich vor allem an der Zufriedenheit seiner Benutzer. Ein Datawarehouse muss sich mit dem Geschäftsprozessen weiterentwickeln.
- Moderne Analysemethoden (Data Mining) und ein aussagefähiges Berichtswesen (OLAP), sowie ausreichende und qualitativ hochwertige Daten sind der Schlüssel zum erfolgreichen Customer Relation Management.