

Über das Datenmanagement

Zweck – Gewähren neue Analyseverfahren die Gewinnung feiner aufgelöster Resultate, ergeben sich neue Probleme sowohl in der Verarbeitungsgeschwindigkeit als auch in der Qualitätsbewertung¹ der neu gewonnenen Meßwerte oder daraus abgeleiteter Informationen. Dies gilt ebenso für moderne IV-Systeme.

Im Zuge dieser Entwicklung veränderte sich auch die Rolle der EDV (heute: IT) in den Unternehmen. War das traditionelle "Rechenzentrum" zunächst eine mit angesehenen Spezialisten versehene Anlaufstelle für datentechnische Probleme aller Art, so vollzog sich bald der Wandel zu einem eher zum Hausbetrieb gehörenden Dienstleister, bis der datenverarbeitende Bereich zur heutigen gleichberechtigten, produktiven Institution wurde, die ein unternehmenswichtiges betriebswirtschaftliches Gut zu bearbeiten hat².

Die Rolle der IT in den Unternehmen durchläuft einen bis heute nicht beendeten Prozeß von der einstmaligen "rohen Datenverarbeitung" bis zur heutigen "Bereitstellung" relevanter Daten. Diese Tätigkeit reicht sehr nahe an das Betätigungsfeld einer (investigativen) Redaktion³ heran. Als Zweck des – nun so genannten – *Datenmanagements*, kann also zusammenfassend⁴ die optimale Nutzung der für das Unternehmen relevanten verfügbaren Daten gefordert werden.

Aufgaben – Aus oben Gesagtem ergibt sich fast zwangsläufig eine Palette von Aufgaben, die individuell auf die Firmensituation oder -größe anzupassen sind. Kernbereiche sind jedoch auf jeden Fall:

- Abschöpfen bestehender und Erschließung neuer Datenquellen (Rohdaten)
- Bewertung der aufgabenbezogenen Signifikanz der Daten (Vorfilterung)
- Bewertung der Datenpriorität und -qualität ausgewählter Daten in Abhängigkeit der Aufgabenbezogenheit (Siebstufen)
- Aufbereitung der Daten in auswertbare Form und ggf. Schaffung eines darauf abgestimmten produktiven Applikationsystems (Redaktion)

Ein besonderer Focus des Datenmanagements ist dabei unverzichtbar der Aspekt der *Datenqualität*. Die einzelnen Komponenten⁵ hierfür sind:

Aktualität	Ständige Erneuerung sich realiter verändernder Daten
Authentizität	Wirtschaftlich vertretbares Abbild der Realität, Korrektheit
Genauigkeit	Wirtschaftlich vertretbare Auflösungsschärfe der Daten
Integrität	Erhalt der Datenoriginalität ⁶
Konsistenz	Widerspruchsfreiheit und Redundanzarmut
Validität	Gültigkeit der Daten in bezogenen Kontexten ⁷
Vollständigkeit	Keine signifikanten Lücken im Datenbestand

1 So ergeben sich beispielsweise durch die vermehrte Auswertung immer größerer Zeiträume im Zusammenhang mit der Sterblichkeit mit Krebserkrankungen zwar neue, veränderte statistisch korrelierte Daten; es ist jedoch ohne zusätzliche Erforschung der Kausalzusammenhänge nicht ein automatischer Zusammenhang oder gar eine Zunahme in diesem Bereich ablesbar, da frühere Erhebungen weder in dieser feinen Auflösung noch in dieser zeitlichen Tiefe erfolgten, wodurch kein direkter Vergleich zulässig ist. Vgl. Interview mit Professor Robert Weinberg, Cambridge MA, im Deutschlandfunk am 04.01.2006, <http://www.dradio.de/dlf/sendungen/forschak/454796/>. Dasselbe gilt auch für betriebswirtschaftliche oder andere Zusammenhänge.

2 Wobei nicht in allen Firmen der letzte Schritt vollzogen worden ist. Es gibt sogar noch Firmen und Behörden, in denen die erste Interpretation vorherrscht.

3 Redaktion: lat.: redigere = etwas in einen Zustand bringen

4 Abweichend von den Definitionen in Krčmar, H.: Informationsmanagement, 3. Aufl., 2003, S.86 und Heinrich, L., Lehner, F.: Informationsmanagement, 8. Aufl., 2005, S.223

5 Alle genannten Punkte bedingen sich größtenteils gegenseitig und bilden je nach Interpretation sogar gerichtete Abhängigkeiten. Hier wurde jedoch o.B.d.A. eine flache Listenstruktur gewählt. Auch wurde auf die diffizile Varianz der aus dem amerikanischen übersetzten Begriffe keine Rücksicht genommen (z.B. "validity", was nicht gut mit "Validität" übersetzbar ist).

6 Abweichend von der Definition in Heinrich, L., Lehner, F., S.223: Durch die Definition der "Konsistenz" in genanntem Werk ist die gesamte Widerspruchsfreiheit bereits abgedeckt, wohingegen die versehentliche oder böswillige Änderung, der Austausch oder die Löschung von Daten unbeachtet bleibt. Dieser Punkt gehört jedoch unverzichtbar mit zu den Aufgaben des Datenmanagements.

7 Z.B. gesetzliche Grenzwerte, Einhaltung physikalischer Gesetze, betriebliche Prokura, etc.

Es ist an dieser Stelle schwer zu unterscheiden, wo reines Datenmanagement sich gegen das Arbeitsgebiet der *Datensicherheit*⁸ abgrenzt. Betrachtet man beispielsweise die einzelnen genannten Punkte zur Datenqualität, so ergeben sich stets Aspekte, die auch durch böswillige Manipulation beeinträchtigt werden können. Somit ergibt sich in diesem Kontext die Datensicherheit in jeglicher Ausprägung sogar als weiterer Aspekt der Datenqualität.

Da es nicht immer ausreicht, nur Daten zur Verfügung zu stellen, ist ein weiterer wichtiger Focus die (Be-) Schaffung und Bereitstellung einer geeigneten Anwendung. Dazu gehören neben einer geeigneten Datenbank auch flexible und brauchbare (Frontend-)Anwendungen. Zu den hierbei wichtigsten Punkten⁹ gehören:

- Kreieren eines effizienten und skalierbaren Datenmodells, Implementation
- Ständige Administration (Wartung, Pflege, Kontrolle), auch des Datenmodells
- Festlegung der Zugriffsrechte ("Wem gehören die Daten?")
- Sichere und flexible Benutzungsschnittstelle
- Exportierbarkeit der Resultatdaten¹⁰

Besondere Erwähnung verdient das *Reengineering* der Daten. Der Datenfluß und -bestand für die Geschäftsprozesse wird optimal gesteuert, so daß weder Unterversorgung (Fehlen von benötigten Daten) noch Überversorgung (Bereitstellen nicht [mehr] benötigter Daten) auftreten. Dabei schließen sich die beiden Zustände nicht aus, es werden im Gegenteil bei schlecht gepflegten Datenbeständen oft beide Zustände gleichzeitig eintreten. Die optimale Steuerung versucht dabei ebenfalls unnötige Redundanzen zu eliminieren. Datenbezogene Redundanzen werden durch geeignete Modellierung, volumenbezogene Redundanzen durch Kompression beseitigt.

Aus dem Gesagten läßt sich leicht schließen, daß die bisherige historische Einordnung in den hauptsächlich operativen Bereich sich inzwischen auch und vor allem in den administrativen Bereich verschoben hat.

Zuständigkeiten – Da die Aufgabestellungen vielerlei Tätigkeiten sowohl aus dem technischen als auch betriebswirtschaftlichen Bereich umfassen¹¹, ist die Gestaltung und Einteilung der Zuständigkeiten eine der schwierigeren Entscheidungen. Abhängig von der Größe der Firma kann dieser Bereich einer Person oder mehreren Personen oder sogar Abteilungen zugeordnet sein. Verschiedene Ansätze werden in der Literatur diskutiert, allerdings gibt es einen Grundkonsens in notwendigen Arbeitsgebieten:

1. *Datenadministrator*: Vorbereitender Bereich, zuständig für Datenmodell, Normierung und Methodik. Ist die Modellierung nicht optimal, werden stets (sonst vermeidbare) Kosten anfallen, da durch ständige Änderungen oder sogar Neumodellierungen (öfter aber durch Benutzung veralteter Datenmodelle) der Betrieb dauerhaft beeinträchtigt werden kann. In der Regel werden solche Kosten gar nicht oder nur schwer entdeckt.
2. *Datenbankadministrator*: Zuständig für Implementation und Speicher- und Laufverhalten. Vor allem die Effizienzgewinnung bei Datenbankzugriffen ist von erheblichem Effekt. Sie beeinflusst die Kosten pro Arbeitsplatz¹² und bei externer Anwendung, z.B. im Auskunftsbereich, sind längere Wartezeiten auf zu suchende Daten kundenabstoßend.
3. *Anwendungssystemadministrator*: Die Rolle in diesem Bereich ist nicht so scharf zu fassen. Während Heinrich¹³ hier eher den Aspekt der Daten- und Dateienverwaltung und Konvertierung der Dateiformate heraushebt, liest man in den aktuellen Stellenbeschreibungen eher die Schwerpunkte konfigurieren, betreiben und pflegen von Unternehmensanwendungen, oft auch beschrieben als Schnittstelle zur Softwareentwicklung.

8 Hierbei ist die *externe* Datensicherheit gemeint. Damit werden böswillige, von Menschen verursachte Manipulationen wie Veränderung, Löschung oder Falscheintrag bezeichnet. *Interne* Datensicherheit beschäftigt sich dagegen mit rein technisch bedingten Anfälligkeiten wie Festplattenausfall, Netzstörungen oder andere Systemstörungen (Betriebssicherheit).

9 Vgl. Krcmar, S.86ff

10 Dieser Punkt findet sich selten in der Literatur, ist jedoch ein wichtiger Aspekt, da die Resultate von Datenrecherchen stets in anderem Kontext (und nicht immer bereits automatisiert) weiterverwendet werden müssen (z.B. in Berichten, Vorträgen, Publikationen, zur Weiterleitung in andere Abteilungen, zum Import in andere Anwendungen, etc.)

11 Vgl. Heinrich, L.: Informationsmanagement, 7. Auflage, 2002, S.230

12 Für einen einzelnen Arbeitsplatz ist es in einem gewissen Bereich nicht sehr relevant, ob auf ein Resultat zwei oder fünf Sekunden gewartet wird. In größeren Firmen kann man das aber schnell auf die Anzahl der betroffenen Arbeitsplätze pro Jahr hochrechnen und auf signifikante Summen stoßen. Selbst in kleinere Firmen kann sich das auf den Unternehmensumsatz auswirken (beispielsweise durch wachsenden Unmut der Mitarbeiter und deren dadurch resultierende steigende Demotivation).

13 Vgl. Heinrich, L., S.230

4. *Datenkatalogsystemadministrator*: Hier ist eher der amerikanische Name für "Datenkatalog" geläufig ("data dictionary", DDIC). Vor allem im Bereich großer Datenbanksysteme (z.B. SAP) spielt dieserart Administration eine große Rolle. Hier werden die Komponenten der Datenbank festgelegt und gepflegt.

Der *Datenmanager* steht dabei als Koordinator dieser Tätigkeiten administrativ und ohne viel operative Tätigkeit im Zentrum aller genannten Tätigkeitsbereiche. Die Liste nun erweiternd, werden noch zwei weitere, unverzichtbare Rollen eingeführt:

5. *Datensicherheitsadministrator*: Wie bereits erwähnt, ist dieser Bereich untrennbar mit dem Datenmanagement verbunden. Hier werden Sicherungsmechanismen zur internen Datensicherheit (z.B. redundante Datensicherungsverfahren) und Methoden zur externen Datensicherheit (z.B. Abwehrmaßnahmen v.a. gegen böswillige Datenmanipulationen) eingerichtet und gepflegt.
6. *Datenschutzbeauftragter*: Gesetzlich¹⁴ vorgeschrieben, muß der Datenschutzbeauftragte eng mit den anderen Stellen des Datenmanagements zusammenarbeiten, soweit es um personenbezogene Daten geht. Dies trifft in vielen Fällen zu, so daß diese Rolle mit zu den Tätigkeiten des Datenmanagements gezählt werden muß. Denn hier werden die gesetzlichen Vorgaben zur Modellierung und Vorhaltung sowie Verarbeitung und Bereitstellung festgelegt, was signifikante Einflüsse auf die Arbeit der anderen Stellen hat.

Verankerung – Die Einordnung des Datenmanagements mit seinen beschriebenen Tätigkeiten wird üblicherweise in einer IT-Abteilung stattfinden. Dies sollte jedoch streng von einer Hardware-IT-Administration separiert werden, denn das Datenmanagement hat nur bedingt mit Hardware-Aspekten zu tun, auch wenn beide gelegentlich zusammenarbeiten müssen. Durch den strategischen Aspekt des Datenmanagements und auch wegen spezieller Rollen (z.B. Datenschutz) ist es empfehlenswert, beispielsweise eine Stabstelle der Unternehmensleitung zu schaffen, so daß Resultate und Empfehlungen direkt kommuniziert oder Entscheidungen direkt umgesetzt werden können.

Methoden – Die Datenmodellierung ist eine wesentliche Methode des Datenmanagements in der Analysephase. Die Vorgehensweise und das Chen'sche ER-Modellierungsverfahren sind hinlänglich bekannt. Darüberhinaus ersetzen mehrdimensionale Datenmodelle¹⁵ die bisherigen, für den Endanwender etwas hölzern vorgefertigten Auswertungen, so daß der Endanwender in vertrauten Softwareumgebungen selbständig Schwerpunkte bestimmen kann und so eigene Auswertungen erhält¹⁶. Auf weitere Methoden wird an dieser Stelle nicht eingegangen; hier wird auf die entsprechende Literatur verwiesen¹⁷.

Glossar

<i>Datenmanagement</i>	Notwendige "Vermaischung" der Daten in der Vorstufe zur Destillation von hochprozentiger Information. Anders ausgedrückt: Vorbereitung, Aufbereitung und Bereitstellung von unternehmensrelevanten Daten zur Informationsgewinnung als wirtschaftlichem Gut.
<i>Datenmodell</i>	Darstellung einer für den zu betrachtenden unternehmerischen Aspekt ausgewählten Teilmenge der realiter vorkommenden Zusammenhänge in einem verabredeten Formalismus zur vereinfachten Situationsanalyse oder zur Implementation einer auf diesem Modell beruhenden Datenbankanwendung.
<i>Datenquelle</i>	Meßbares Ereignis, welches durch ständige Auswertung erlaubt, Daten im Kontext einer Situation oder eines Phänomens zu erheben.
<i>Signifikante Daten</i>	Daten, die innerhalb eines ausgewählten Kontextes Reaktionen seitens der erhebenden Stelle erfordern oder die zum Beweis eines Zustandes benötigt werden.
<i>Bereitstellung von Daten</i>	Aufgrund eines Modells ausgewählte und nach Signifikanz vorgefilterte Daten, die in einen für die abrufende Stelle konsumfähigen Zustand überführt werden.

14 Vgl. §4f Bundesdatenschutzgesetz, http://bundesrecht.juris.de/bdsg_1990/BJNR029550990BJNE006501301.html

15 Ein ganz guter Einstieg findet sich unter <http://de.wikipedia.org/wiki/MOLAP>, (Stand Jan 2006)

16 Vergleichbar mit den einfacheren Pivot-Tabellen in MS-Excel

17 Vgl. Heinrich, L. oder Krömer, H.