



Fehlendes Datenqualitätsmanagement kostet ...

Angesichts der hohen Bedeutung von Daten und Informationen wird heutzutage in den Unternehmen diesem Wertschöpfungsfaktor im Sinne der Qualitätssicherung noch erstaunlich wenig Aufmerksamkeit geschenkt. Häufig bleibt es bei Lippenbekenntnissen oder isolierten, lokalen Lösungen in verschiedenen IT-Systemen oder organisatorischen Einheiten, die dem Anspruch eines durchgängigen, abgestimmten und erst dann den vollen Nutzen stiftenden Datenqualitätsmanagements (DQM) nicht gerecht werden. Immerhin lässt sich in jüngster Vergangenheit eine Renaissance des Themas beobachten, vielleicht u.a. durch Bewegungen und Fortschritte am (Software-)Markt verursacht. Ein wesentlicher Treiber sind auch regulatorische Vorgaben, die die Unternehmen im Rahmen des Compliance Managements dazu zwingen, Transparenz zu schaffen und Daten- und Prozessqualität zu verbessern.

Nach wie vor leiden „Infrastrukturmaßnahmen“ wie das Datenqualitätsmanagement unter dem Problem des Nachweises der Wirtschaftlichkeit. Zwar gibt es Aussagen namhafter Analysten, die beeindruckende Zahlen von aufgrund schlechter Datenqualität verursachten Kosten und gescheiterten Projekten in den Raum stellen. Angesichts der Allgemeinheit und fehlender direkter Übertragbarkeit solcher Thesen lässt sich das Management im eigenen Unternehmen oft nicht sofort von der Notwendigkeit eines DQM überzeugen. So bleiben nur konkrete Wirtschaftlichkeitsbetrachtungen, idealerweise anhand zuvor festgelegter Business Cases. Aber auch hier kann man nicht immer auf das beliebte Beispiel von den leicht messbaren Einsparungen bei Marketingaktionen durch korrekte und zielgruppengerechte Adressierung potentieller Kunden zurückgreifen. Stattdessen müssten häufig indirekte Nutzenpotenzialen erfasst werden, die sich nur schwer quantifizieren lassen. Letztlich sollte man sich aber immer vor Augen halten: Nicht Datenqualität kostet, sondern fehlende Datenqualität.

Datenqualitätsprobleme im Data Warehouse haben ihren Ursprung häufig in den zugrundeliegenden operativen Systemen. Die damit verbundenen organisatorischen und technischen Grenzen schränken aus Sicht der Data-Warehouse-Abteilungen die Möglichkeit eines sogenannten proaktiven Datenqualitätsmanagements meist ein. Denn a priori die Entstehung von Fehlern zu vermeiden, wäre langfristig einer permanenten nachträglichen Datenbereinigung im Rahmen der Datenbewirtschaftung des Data Warehouses vorzuziehen. Erschwerend kommt hinzu, dass die Datenqualität aus Sicht der operativen Systeme durchaus ausreichend sein kann, im Data Warehouse bei Integration mit anderen Daten oder auf-

grund höherer Ansprüche (etwa durch regulatorische Vorgaben) aber nicht mehr. Wenn dann im Rahmen des Data Warehousing Korrekturmaßnahmen unumgänglich werden, kann heute zunehmend auf die Funktionalität von Datenintegrationswerkzeugen zurückgegriffen werden. Gerade die namhaften ETL-Hersteller ergänzen derzeit ihre Toolsuiten durch gezielte Zukäufe um Lösungen für Data Profiling und Data Cleansing. Damit wird ein integriertes Konzept von Datenbewirtschaftung, Datenqualitätsmanagement und auch Metadatenmanagement mit entsprechender technischer Umsetzung erleichtert.

Wie fast immer in der IT, ist aber Technik nur die halbe Miete. Folgt man dem gängigen Verständnis von Datenqualität als „der Gesamtheit aller Eigenschaften von Daten hinsichtlich der Fähigkeit, die Anforderungen des Anwenders zu erfüllen“ (in Anlehnung an die DIN/ISO-Norm 9004), dann wird auch klar, dass Datenqualität nicht durch allgemeingültige Kennzahlen bestimmt und gesteuert werden kann, sondern immer einen unternehmens- und anwendungsspezifischen Kontextbezug aufweist. Folglich muss auch im Data Warehousing – wenn man von standardisierten Qualitätschecks absieht, die aber eher auf syntaktische Korrektheit prüfen und nur eingeschränkt auf Semantik (bspw. bei der Gültigkeit von Adress-Sätzen) abzielen - bei zunehmender Fachlichkeit das Business einbezogen werden. Es wird benötigt zur Definition von Prüfregelein, die fehlerhafte Datensätze aussteuern oder Festlegung von Kennzahlen, die die Datenqualität messen. Die Komplexität und der Aufwand für das Aufstellen solcher fachlicher Regeln sollte nicht unterschätzt werden.

Fehlen die organisatorischen, politischen oder technischen Möglichkeiten zur Verbesserung der Datenqualität, sei es proaktiv oder nachträglich, so ist zumindest die Messung und Kommunikation von Qualitätskennziffern sinnvoll. Denn sie helfen, die Korrektheit, Aktualität und Relevanz von Reports und Kennzahlen aus dem Data Warehouse richtig einzuschätzen. Im Sinne der 80/20-Regel müssen (und können) auch nicht notwendigerweise immer alle Daten einem hundertprozentigen Qualitätsanspruch genügen – dann ist aber Transparenz über den Qualitätsstand der Daten gegenüber den Anwendern unerlässlich. Übrigens erweist sich immer wieder die offene Kommunikation von Qualitätsdefiziten im Unternehmen als ein geeignetes Instrumentarium zur Bewusstseinsbildung in den betroffenen Organisationseinheiten und letztlich zu einer sukzessiven Verbesserung der Datenqualität.

Aufgrund der geschilderten Aktualität stellt das Qualitätsmanagement auch eines der Kernthemen im Kompetenzzentrum Enterprise Information Warehousing (CC EIW) des Instituts für Wirtschaftsinformatik der Universität St. Gallen dar. Hier wird in Kooperation mit Großunternehmen die Professionalisierung des Data Warehousing und die Steigerung des Geschäftsnutzens der Informationslogistik unter Einbezug aktueller fachlicher Treiber vorangetrieben.

Nähere Informationen sind unter <http://eiw.iwi.unisg.ch> erhältlich.

Dr. Barbara Dinter

Projektleiterin
Kompetenzzentrum Enterprise Information Warehousing
Institut für Wirtschaftsinformatik
Universität St. Gallen
E-Mail: barbara.dinter@unisg.ch