

Executive Outline

Datenmanagement und Verbesserung der Datenqualität

Mangelnde Datenqualität in Unternehmen stellt für diese ein Problem da, denn in den Daten steckt Wissen, dass die Unternehmen im Wettbewerb voranbringt. Vor diesem Hintergrund stellt sich die Frage wie das Datenmanagement und die Verbesserung der Datenqualität organisiert werden muss. Nur mit hochwertigen Daten können hochwertige Analysen und Prognosen erstellt, und damit Wettbewerbsvorteile erzielt werden.

Fehlerquellen

Fehler in den Daten lassen sich kaum vermeiden. Es können folgende Fehlerquellen festgestellt werden: Fehler bei der Dateneingabe, Eingabe im falschen Feld, Mängel die durch externe Daten importiert werden, sowie falsche Kodierungen und das Vermischen von korrekten und falschen Daten, wodurch ganze Datensätze wertlos werden können. Aber auch die Zusammenführung verschiedener Datenquellen in einem Data Warehouse kann zu fehlerhaften Informationen führen. Dabei kann es zur Dublettenbildung und Inkonsistenzen kommen.

Unzureichende Praxis

Das Ziel ist es Regeln und Kriterien zu entwickeln, wie die Fehler behoben werden können und so die Datenqualität erheblich verbessert werden kann. Die Realität im Unternehmen sieht noch ganz anders aus. In der Praxis wird oft das große Datenqualitäts-Projekt von der IT Abteilung gestartet, um die Datenqualität unternehmensweit zu optimieren. Dieser Top-down-Ansatz verlagert das Problem der Datenqualität auf einer anwenderfernen Ebene. Dies führt oft zur unflexiblen, intransparenten Lösungen und ist meist langwierig. Eine Lösung des Datenqualitätsproblems, die von vielen Unternehmen bevorzugt wird, ist die, dass die Mitarbeitern in den Fachabteilungen ihren eigenen Weg zur Problembehebung suchen und erstellen dann Provisorien, meist mittels ihnen vertrauten Office-Tools. Diese Lösung führt zur Intransparenz und bietet nur Einzellösungen an, wodurch lediglich in Einzelfällen die Daten an der Realität angepasst werden, die Qualität der Daten wird aber dadurch nicht unternehmensweit für alle verbessert. Diese Vorgehensweise verbirgt die Gefahr, dass Reports nicht den tatsächlichen, sondern einer individuell angepassten Datenbasis entsprechen.

Die entscheidende Frage muss sein „Wer ist für die Güte der Daten verantwortlich“.

Ein Ansatzpunkt der Zusammenarbeit der Fach- und IT-Abteilungen ist das Validieren der Daten mit Hilfe des so genannten Profiling: Stammdaten werden in der Regel ein einziges Mal erfasst. Solange die wesentlichen Informationen stimmen, besteht auf dem ersten Blick kein Grund diese Daten zu bereinigen, denn für die operativen Prozesse sind sie gültig. Für Analysen und Prognosen reicht diese minimale Datenqualität nicht aus. Daher muss jeder Fachbereich die Güte der Daten überprüfen und Korrekturen vornehmen. Beim Profiling wird nach Auffälligkeiten gesucht, dabei kommen Regeln zur Anwendung, die bereits in einer Software enthalten und vorinstalliert sind, die der Anwender bei der Qualitätssicherung selbst erstellt. Innerhalb der Datenintegrationsprozesse werden diese Regeln angewendet und ausgeführt. Dazu erhält jedes Feld in einer Transformation eine Rolle zugeordnet. Die rollenspezifischen Funktionen zerlegen oder standardisieren die Inhalte, Da es für die meisten Sachverhalte eine Vielzahl von Darstellungsmöglichkeiten gibt, müssen Modelle definiert werden, die die Variationen in ein einheitliches Format überführen, ein Beispiel dafür sind abteilungsspezifische Definitionen von Begriffen, die angeglichen werden müssen.

Je leistungsstarker die Funktionen für das Erstellen von Validierungs- und Korrekturregeln in einer Software zur Verbesserung der Datenqualität sind, umso flexibler lassen sich diese Regeln handhaben. Für diese Variablen gibt es vorgefertigte Regeln. Ein flexibles System zur Datenbereinigung lässt auch weitere Regeln zu, die der Anwender selbst erstellt.

Executive Outline

Damit sind alle Mitarbeiter aufgerufen an die Verbesserung der Datenqualität mitzuwirken. Eine zentrale Stelle, meist in der IT Abteilung angesiedelt sammelt, pflegt und verwaltet die Regeln und stellt sie unternehmensweit zur Verfügung. So profitiert jede Abteilung vom Wissen der anderen und trägt gleichzeitig zur Wertschöpfung bei. Dieses Wissen wird durch die Regeln langfristig im Unternehmen bleiben, zugleich bleibt auch die Konsistenz der Daten garantiert, es wird einen Wildwuchs der Definitionen und Vorgaben verhindert.

Business Intelligence als Wettbewerbsvorteil

Der Einsatz von Business Intelligence gilt als Wettbewerbsvorteil, der über den langfristigen Unternehmenserfolg entscheidet. Oft sind aber BI-Projekte teuer und riskant: Im Laufe des Projektes stellt sich heraus, dass manche Anwendung sich als nicht umsetzbar erweist. Lösungen werden häufig verspätet und nicht für alle geplanten Analysen und Anwender live gesetzt.

Die Ursachen für die hohen Risiken der BI-Projekte sind vielfältig. Einen großen Anteil haben fachliche oder technische Lücken innerhalb der zugrundeliegenden operativen Prozesse sowie oft damit verbunden, eine mangelhafte Datenqualität. Oft werden aber auch in der Konzeptionsphase Sachverhalte und verborgene Anforderungen vergessen. Viele dieser Probleme lassen sich gut durch einen frühen Realitäts-Check mit Hilfe gründlicher Datenanalysen vermeiden.

Die Qualität der Ergebnisse eines Projektes hängt maßgeblich von der Qualität der verfügbaren Prozessinformationen und der Daten innerhalb der operativen Systeme ab. Metadaten zu Prozessen liegen in einem Unternehmen meist nur rudimentär vor. Technische Prozessbeschreibungen spiegeln die Wirklichkeit nicht vollständig wieder, wenn diese nicht aktuell gehalten werden oder keine relevanten Workarounds, Historien und Sonderlösungen enthalten. Auch der Rückgriff auf das umfangreiche Wissen von technischen und fachlichen Experten führt zu keinem vollständigen, aktuellen und korrekten Bild von Systemen und Prozessen. Dennoch verlassen sich viele BI-Projekte auf Dokumentationen zu Prozessen und Systemen sowie auf Interviews und Workshops mit Spezialisten. Das auf dieser Weise entwickelte theoretische Verständnis von Zusammenhängen wird durch einer Reihe expliziter und impliziter Annahmen zu einem vermeintlich vollständigen und umsetzbaren Bild. Theorie und Annahmen werden unzureichend hinterfragt und sind daher mit einer hohen Fehlerwahrscheinlichkeit behaftet, auch wenn die Papierform eine gute Qualität aufweist und dadurch ein geringes Risiko suggeriert.

Die Auswirkungen „vergessenen Fakten“ auf das BI-Projekt können erheblich sein. Datenmodelle müssen erweitert oder komplett umgestellt werden, da diese zwar die meisten Fälle sorgfältig berücksichtigen, aber nicht die vergessenen „Prozessausnahmen“. Mitunter ist die BI-Anwendung nicht verwendbar, wenn zu spät erkannt wird, dass das angebundene Quellsystem für bestimmte Kundengruppen oder Transaktionsarten nicht die gewünschten Daten beinhaltet, da für diese Fälle die Verarbeitung in speziellen Systemen erfolgt.

Empfehlung für die Praxis

Profiling als Schlüssel zur Datenqualität:

Die Datenqualität kann durch eine sog. Data Profiling überprüft werden. Dabei kommen Regeln zur Anwendung, die entweder in einer Software enthalten und vorinstalliert sind, oder die der Anwender bei der Qualitätssicherung selbst erstellt. Dies kann durch den Einsatz spezieller Tools wie beispielsweise die von Informatica oder Hyperion. Aber auch die Prüfung der Datenqualität mittels SQL und Excel kann im Einzelfall eine sinnvolle und kostengünstige Alternative sein.