

## SONDERDRUCK AUS BI-SPEKTRUM 4/09 FÜR



**Sven Braxein** ist Geschäftsführer der auf Qualitäts- und Testmanagement spezialisierten Unternehmensberatung TestGilde GmbH.  
E-Mail: sven.braxein@testgilde.de

**Dr. Marcus Dill** ist Geschäftsführer des BI-Analysten- und Beratungshauses mayato.  
E-Mail: marcus.dill@mayato.com

### Wider die vergessenen Fakten

## Qualität fängt mit den Daten an

*Der Einsatz von Business Intelligence gilt als Wettbewerbsvorteil, der über den langfristigen Erfolg eines Unternehmens entscheidet. Leider stehen BI-Projekte aber auch in dem Ruf, unkalkulierbar teuer und riskant zu sein. Nicht zu unrecht, verschlingt doch so manches Projekt deutlich mehr Geld als zunächst geplant, während sich gleichzeitig im Laufe des Projektes die eine oder andere Anwendung als nicht umsetzbar erweist. Lösungen werden häufig verspätet und nur für einen Bruchteil der ursprünglich geplanten Analysen und Anwender live gesetzt. Die Ursachen für die hohen Risiken von BI-Projekten sind vielfältig. Einen großen Anteil haben fachliche oder technische Lücken innerhalb der zugrunde liegenden operativen Prozesse sowie – oft damit einhergehend – eine mangelhafte Datenqualität. Häufig sind es auch vergessene Sachverhalte und verborgene Anforderungen, die BI-Projekte in Schieflage bringen, weil sie in der Konzeptionsphase übersehen werden. Viele dieser Probleme lassen sich gut durch einen frühen Realitäts-Check mit Hilfe gründlicher Datenanalysen vermeiden.*

### Theorie und Praxis von Unternehmensprozessen

Die Qualität der Ergebnisse eines Projektes hängt maßgeblich von der Qualität der verfügbaren Prozessin-

formationen und der Daten innerhalb der operativen Systeme ab. Metadaten zu Prozessen liegen in einem Unternehmen meist nur sehr rudimentär vor. Selbst wenn ARIS-Modelle oder Ähnliches existieren, sind diese im Normalfall von der Ebene der technischen Realität stark entkoppelt. Technische Prozessbeschreibungen ihrerseits spiegeln die Wirklichkeit nicht vollständig wider, 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 fast ausschließlich auf Dokumente zu Prozessen und Quellsystemen sowie auf Interviews und Workshops mit Spezialisten. Das sich auf diese Weise entwickelnde theoretische Verständnis von Zusammenhängen wird durch eine 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 Informationsqualität und damit ein geringes Risiko suggeriert.

Die Prozess- und Systemwelt eines Unternehmens beinhaltet eine Vielzahl von prozessualen oder technischen Workarounds und Sonderlösungen, die im Kontext des BI-Projekts nicht beachtet oder als irrelevant eingeschätzt werden. Die Auswirkungen dieser „vergessenen Fakten“ auf das BI-Projekt können jedoch erheblich sein: Datenmodelle müssen erweitert oder komplett umgestellt werden, da diese zwar den „95%-Fall“ sorgfältig berücksichtigen, nicht aber die vergessenen „Prozessausnahmen“. Mitunter ist die BI-Anwendung zunächst nicht verwendbar, da 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.

Vergessene Fakten sind oftmals auf die Historie von Prozessen und Daten oder auf die Auswirkungen von Einmalaktivitäten in der Vergangenheit bezogen. Zum Zeitpunkt der BI-Konzeption legt man hingegen typischerweise den Fokus auf den aktuellen Stand von Systemen und Prozessen. So können zum Beispiel eine Eingabeprüfung im Quellsystem erst vor kurzem eingebaut worden sein, jedoch eine Vielzahl von unvollständigen oder inkonsistenten Altdatensätzen vorliegen. Oder im Zuge der Migration oder Bereinigung von Datenbeständen wurden Felder mit Defaultwerten gefüllt, die aber fachlich nicht nutzbar sind.

Ein den vergessenen Fakten verwandtes Problem sind vergessene Anforderungen. Sie resultieren oft daher, dass Anwender Dinge für so selbstverständlich halten, dass sie diese nicht mehr explizit erwähnen. In anderen Fällen verbinden Anwender eigene Annahmen über

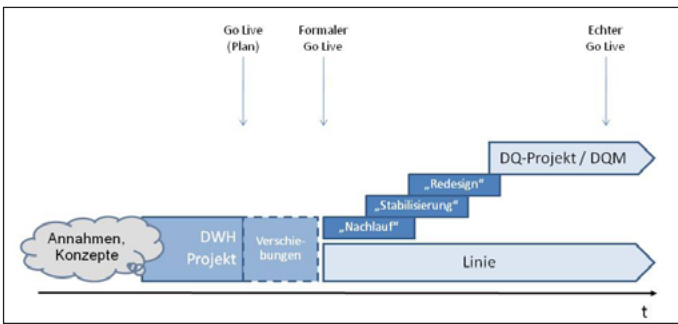


Abb. 1: Häufiges Ablaufmuster bei der Einführung eines Data Warehouse

Datenquellen und Lösungsdesign mit ihren Anforderungen und formulieren nicht die eigentlichen Anforderungen, sondern ihre Vorstellung von der Lösung. Beispielsweise kann der Anwender als Quelle für die von ihm gewünschten Daten System X angeben, nicht wissend, dass ein Teil der Informationen aus dem System Y stammt und von dort extrahiert werden muss. Oder der Anwender weiß, dass eine bestimmte Datenquelle bereits an das Data Warehouse angebunden ist, aber nicht, dass die Daten bestimmter Geschäftsbereiche (zum Beispiel bei einer Bank Geschäfte mit eigenen Mitarbeitern) bei der Extraktion aus Datenschutzgründen ausgefiltert werden.

## Folgen vergessener Fakten

Vergessene Fakten resultieren oft in konzeptionellen Problemen der entwickelten Lösung, die ihrerseits zu einer Verschiebung des „Go Live“ oder einer verlängerten und teuren Nachlaufphase führen können. In vielen Fällen wird diese Nachlaufphase genutzt, um Übergangslösungen zu etablieren, die mit einem erhöhten Aufwand für die Bewirtschaftung des Systems einher gehen und Anwender trotzdem nur eingeschränkt zufrieden stellen (s. Abbildung 1).

Zur Behebung der fachlichen Mängel werden Folgeprojekte aufgesetzt, die dann unter Überschriften wie „Stabilisierung“ oder „Re-Design“ das Datenmodell und die Beladungsprozesse überarbeiten, um so den vergessenen Fakten und Anforderungen gerecht zu werden. Lösungen, die in diesem Stadium gesucht werden, sind oftmals stark auf die BI-Seite konzentriert, das heißt man versucht, die BI-Lösung „irgendwie“ so zu gestalten, dass eventuelle Unzulänglichkeiten der Prozesse, Vorsysteme und Quellsystemanbindungen in den Analysen umgangen werden können. Dieses Kurieren der Symptome (in der BI-Umgebung) anstelle der eigentlichen Ursachen (in den Quellsystemen und operativen Prozessen) führt aber nicht immer zum Erfolg. Manchmal versteht man erst in die-

sem Stadium, dass operative Prozesse ganz grundsätzlich technisch oder fachlich überholt werden müssen, oder dass ein Datenqualitätsprojekt die Grundlage für ein systematisches Datenqualitätsmanagement (DQM) legen muss. Erst nach diesen weiteren Investitionen und oft Monate oder sogar Jahre nach dem geplanten „Go Live“ kann von einer echten Produktivnutzung durch eine größere Zahl an Benutzern und mit ernst zu nehmenden Anwendungen gesprochen werden.

Dieser ineffiziente und kaum zufriedenstellende Ablauf wird regelmäßig bei der Einführung eines

Data Warehouse (DWH) aber auch bei anderen BI-Projekten beobachtet. Nicht wenige Projekte scheitern dabei auch vollständig, da im Unternehmen der Glaube an den Erfolg des BI-Projekts mittlerweile schwer erschüttert ist sowie Mittel und Unterstützung für die erforderlichen Zusatzarbeiten und Prozessumstellungen fehlen. Viel schwerer als die Millionen, die ein BI-Projekt bis zu seinem teilweisen oder vollständigen Scheitern verschlungen haben mag, wiegt jedoch der Umstand, dass die Ziele des BI-Projekts nicht erreicht wurden: Das Unternehmen wird unverändert im Blindflug gesteuert und gesetzlichen Anforderungen nach wie vor nicht vollständig und effizient nachgekommen. Der Wettbewerbsdruck hat eher noch zugenommen und Wachstumschancen bleiben weiterhin ungenutzt. Zu den Projektkosten gesellen sich also massive wirtschaftliche Folgen durch das Nichterreichen von Projektzielen (s. Abbildung 2).

## Qualität fängt mit den Daten an

Erfolgreiche BI-Projekte erfordern eine Analyse der Datenqualität. Häufig erfolgt diese Datenanalyse jedoch erst im Rahmen einer Testphase, das heißt nachdem bereits viele Monate an Arbeit investiert worden sind. Eine Vielzahl der Probleme, unter denen BI-Initiativen leiden, lässt sich handhaben, indem bereits zu Beginn die Daten, deren Abbildung im Reporting Zweck des BI-Projekts ist, gründlich analysiert werden. Fakten zu Datenattributen, Objekten und Relationen ergänzen und validieren die bereits durch Dokumente und Experten gewonnenen

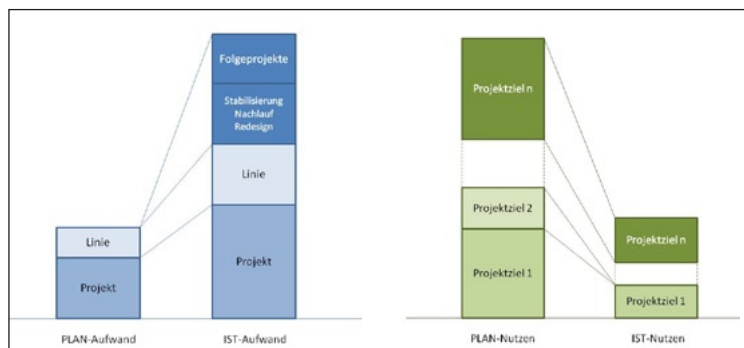


Abb. 2: PLAN und IST vieler BI-Initiativen

Metadaten und erleichtern so eine korrekte Konzeption und Modellierung. Eine frühe Prüfung der Datenqualität (zum Beispiel auf Dubletten und Fehleingaben) liefert zudem Hinweise auf die Notwendigkeit von DQ-Maßnahmen im Quellsystem beziehungsweise Prüfungen und Bereinigungen auf dem Beladungsweg innerhalb der BI-Applikation.

Frühe Datenanalysen eignen sich auch hervorragend zur Aufdeckung vergessener Fakten. Mögliche Fragestellungen dabei sind:

- ◆ Beinhalten die vermuteten Quellsysteme tatsächlich die benötigten Daten vollständig? Gibt es Prozessvarianten, für die Daten in weiteren Systemen vorliegen?
- ◆ Lassen sich die zu diesem Zeitpunkt bekannten Anforderungen an die spätere BI-Anwendung vollständig mit Hilfe dieser Daten abbilden?
- ◆ Gibt es in den Daten Hinweise darauf (zum Beispiel unerwartete Ausprägungen, Dummywerte, nicht gefüllte Pflichtfelder), dass außer den bekannten fachlichen und technischen Prozessen weitere Prozesse die betrachteten Daten beeinflussen oder beeinflusst haben?

Bereits mit einem Aufwand von wenigen Stunden für die Analyse einer Datenquelle (wenige Tage für ein komplettes DWH-Projekt) kann ein hoher Erkenntnisgewinn und eine Erhöhung der inhaltlichen Qualität der BI-beziehungsweise DWH-Konzeption erreicht werden. Die Ergebnisse der Analysen werden durch ein kleines Team (ein erfahrener BI-Modellierer beziehungsweise Datenanalyst, ein technischer Experte des Quellsystems sowie ein fachlicher Prozessspezialist) ausgewertet und Auffälligkeiten plausibilisiert.

Für kleine BI-Anwendungen und sehr einfache Datenstrukturen ist eine initiale Datenanalyse mit Hilfe von SQL und den ubiquitären Microsoftprodukten Excel und Access möglich. Gemeinhin ist dieses manuelle Vorgehen jedoch zu aufwändig, um damit umfassende statistische und Musteranalysen durchzuführen. Um die Auswertung und Interpretation der Daten möglichst effizient durchzuführen, bedient sich ein BI-Projekt vorzugsweise professioneller Data-Profiling-Werkzeuge (erhältlich als isolierte Spezialwerkzeuge, als Teil von DQ-Suiten oder innerhalb von ETL-Tools).

Neben ihrer Kernaufgabe, der automatischen Extraktion von Metadaten aus einer Datenbasis (zum Beispiel Eigenschaften von Attributen oder Tabellen, Abhängigkeiten und Relationen, enthaltene Regeln und Muster) bieten diese Werkzeuge vor allem auch hilfreiche Visualisierungsoptionen sowie die Möglichkeit zur Dokumentation und damit zur späteren Verwendung der Metadaten im Falle von Erweiterungen oder zum Zweck von DQ-Prüfungen.

Unabhängig vom verwendeten Tool sollte das Data Profiling möglichst nah an den Quelldaten durchgeführt werden, das heißt nicht mit bereits verarbeiteten Daten, in denen Sachverhalte durch

Mappings maskiert sind, und sich so der Entdeckung entziehen können. Ein Vorgehen über Stichproben birgt Nachteile, da in Stichproben je nach Art der Selektion oft nicht alle später im Produktivbetrieb auftretenden Konstellationen erscheinen. Ein Profiling der vollständigen Quellsystemdaten erhöht somit die Sicherheit, vergessene Fakten zu erkennen und in Konzepte zu integrieren, bevor sie zusätzliche Kosten, Verschiebungen oder reduzierte Analysemöglichkeiten verursachen.

## Datenanalysen als Basis für faktenbasierte Entscheidungen nutzen

BI-Projekte werden in vielen Unternehmen gemäß der dort für IT-Projekte üblichen Methodologie aufgesetzt. Typischerweise folgen diese Methodologien in der einen oder anderen Form dem klassischen Wasserfallmodell (Fachkonzept → DV-Konzept → Realisierung → Test → Go Live). Für BI-Projekte ist dieser Ansatz aufgrund der Abhängigkeit von zuliefernden Systemen und Prozessen und den darin liegenden Unsicherheiten und Risiken nur bedingt geeignet. Die Wahrscheinlichkeit, dass Qualitäts-, Budget- oder Terminziele nicht eingehalten werden können, ist daher in BI-Projekten stark erhöht.

Eine Möglichkeit, dieser Unsicherheit von BI-Projekten zu begegnen und damit Risiken zu umgehen, besteht in der Integration von Datenanalysen und darauf aufbauenden Entscheidungspunkten in den Projektablauf. Solche faktenbasierten Entscheidungspunkte sollten mindestens am Beginn des DWH-Projektes und begleitend zur fachlichen und technischen Konzeptionsphase liegen, wo sie wertvolle Dienste bei der unterstützenden Recherche und gezielten Validierung von Anforderungen und Annahmen leisten.

Am Beginn des Projekts sollten zunächst grundsätzliche Fragen beantwortet werden, wie sie im vorigen Abschnitt erläutert wurden. Dabei empfiehlt es sich, als Option auch den Abbruch beziehungsweise das Vorziehen anderer Aktivitäten vorzusehen. Eine Erkenntnis kann beispielsweise sein, dass die Voraussetzungen für eine vollständige Abbildung der Anforderungen nur gegeben

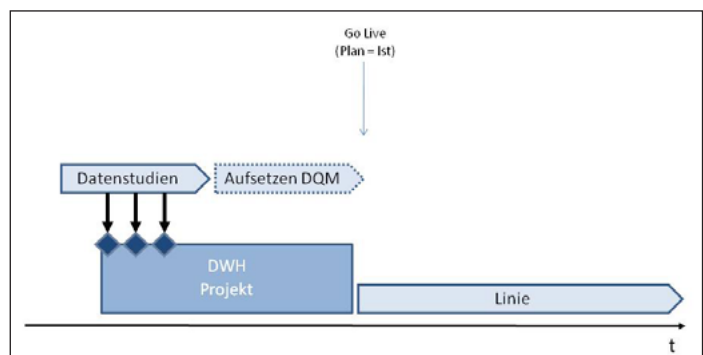


Abb. 3: Vorangestellte und begleitende Datenstudien vermindern Risiken und verkürzen den Zeitraum bis zum echten Go Live

sind, wenn zuvor bestimmte Prozesslücken auf operativer Seite geschlossen werden. Oder es wird festgestellt, dass ein DQ-Projekt die Basisqualität der Daten in den Quellsystemen sicherstellen muss (s. Abbildung 3).

Auch wenn das BI-Projekt im schlimmsten Fall gestoppt oder zurückgestellt werden muss, so ist zumindest der finanzielle Schaden durch hohe Projektkosten verhindert, gleichzeitig die Basis für weitere Unternehmensentscheidungen geschaffen und vor allem kein Kredit für die BI-Initiative verspielt worden.

Zusammenfassend kann festgestellt werden, dass sich viele Probleme von BI-Projekten lösen lassen, wenn sie rechtzeitig erkannt werden und das gewonnene Wissen um die Ist-Situation von Prozessen und Daten in Kon-

zeption und Umsetzung entsprechend genutzt wird. Das Risiko von Verschiebungen und Scope-Reduktionen wird auf diese Weise wesentlich vermindert. Projektziele können erreicht werden und es wird eine solide Basis für eine Vielzahl von analytischen Anwendungen gelegt.

**BI-SPEKTRUM** ist eine Fachpublikation des Verlags:  
SIGS DATACOM GmbH | Lindlaustraße 2c | 53842 Troisdorf  
Tel.: +49 (0) 22 41.2341-100 | Fax: +49 (0) 22 41.2341-199  
E-mail: [info@sig-datacom.de](mailto:info@sig-datacom.de)  
[www.bi-spektrum.de](http://www.bi-spektrum.de)

**SIGS DATACOM**  
FACHINFORMATIONEN FÜR IT-PROFESSIONALS