

Daten über Daten

■ Ballast oder Vorteile durch Metadaten

von Markus Nicklas

Transparenzvorgaben zum Datenbereitstellungsprozess – speziell im Bereich Risikomanagement wie durch die BCBS #239 formuliert – erfordern zunehmend, die fachliche wie auch technische Herleitung von Kennzahlen zentral zu dokumentieren. In diesem Kontext ist der Begriff der Metadaten (= „Beschreibung von Daten“) in aktuellen Fragestellungen rund um das Thema Datawarehousing und Business Intelligence kaum wegzudenken. Während sich die Hersteller von Software für Business Intelligence oder Datenintegration meist nur auf die IT-technischen Metadaten konzentrieren und sich ausschließlich auf den Wirkungskreis ihres ETL-Entwicklungsprozesses innerhalb der Datawarehouse-Schichtenarchitektur beschränken, geht ein vollständig integriertes Metadatenmanagement deutlich weiter. Es geht über Systemgrenzen hinaus und bietet sowohl eine Plattform für fachliche wie auch rein technische Informationsgewinnungsprozesse als auch eine gemeinsame Basis für den Fachbereich und die IT.

Technische Metadaten innerhalb der ETL-Entwicklung

Die Verwendung technischer Metadaten hat zum heutigen Zeitpunkt eine hohe Marktreife erlangt und wird von den etablierten ETL-Werkzeugen umfassend unterstützt. Das Grundprinzip hinter technischen Metadaten besteht darin, technischen Objekten, wie etwa einem Attribut in einer bestimmten Tabelle auf einem bestimmten Datenbankschema, eine verständliche Bezeichnung zu geben. Neben dieser Bezeichnung können weitere Metainformationen, wie etwa Beschreibung, Erstellungsdatum der Datenstruktur oder beliebige weitere Metadaten, gespeichert werden. Der Ausgangspunkt besteht jedoch immer in einem bereits existenten technischen Objekt, das nachträglich beschrieben wird.

Hierdurch ergeben sich – insbesondere im Umfeld der grafischen ETL-Entwicklung – zwei Vorteile:

1. Zunächst schafft die Visualisierung grafischer ETL-Jobketten durch die fachlich sprechende Bezeichnung von technischen Objekten eine erhöhte Lesbarkeit der ETL-Jobs für sämtliche Adressaten. Besonders Fachbereichsanwender profitieren von der Möglichkeit, ETL-Jobs „im Original“, d. h. beispielsweise im Entwicklungswerkzeug SAS DI-Studio (dargestellt in Abb. 1) betrachten zu können, ohne dabei in einem separaten Schritt technische Objektnamen für ein Verständnis des Ablaufs auf ihren fachlichen Ursprung zurückverfolgen zu müssen.



Abbildung 1: Exemplarische technische Metadaten in SAS-DI Studio

2. Weiterhin ermöglicht die metadatenbasierte ETL-Entwicklung eine Entkopplung von der physischen Datenebene. So kann etwa die Tabelle „Mitarbeiter“ heute auf einem anderen Datenbankserver gespeichert sein, als dies zukünftig der Fall sein wird. Eine Änderung des physischen Namen oder des Speicherortes erfordert dann lediglich eine einmalige Anpassung innerhalb der Metadaten. Im Gegensatz dazu wäre bei der Nichtverwendung von Metadaten eine Anpassung in jedem Programm- und Programmteil, in dem die Tabelle verwendet wird, notwendig.

Bereits die ausschließliche Verwendung technischer Metadaten schafft deutliche Vorteile innerhalb der DWH-Architektur. Eine weitere Wertsteigerung erfährt der Metadatenansatz jedoch durch die zusätzliche Erfassung von fachlich-konzeptionellen Metadaten.

Konzeptionelle Metadaten als Medium für die Entwurfsspezifikation

Während die ausschließliche Verwendung eines technischen Metadatenansatzes im Umfeld der ETL-Entwicklung immer ein bereits bestehendes physisches Bezugsobjekt fordert, werden konzeptionelle Metadaten losgelöst von einem solchen definiert.

Unterstützt werden somit die vor der Entwicklung liegenden Projektphasen, also konkret die Fach- und DV-Konzeption. Die zentrale Erfassung der konzeptionellen Metadaten auf einer Plattform wie dem MetaMaster von msgGillardon schafft umfassende neue Möglichkeiten gegenüber einem rein technisch getriebenen Ansatz. So können etwa Datenanforderungen oder Transformationsregeln spezifiziert und erst nachfolgend mit bestehenden technischen Metadaten in Verbindung gesetzt werden. Der Begriff der Metadaten erhält so seine ursprüngliche Bedeutung zurück: Anstelle einer fachlichen Nachdokumentation werden angeforderte Daten beschrieben, bevor sie im Rahmen der Entwicklung in Form von ETL-Strecken geladen und transformiert werden.

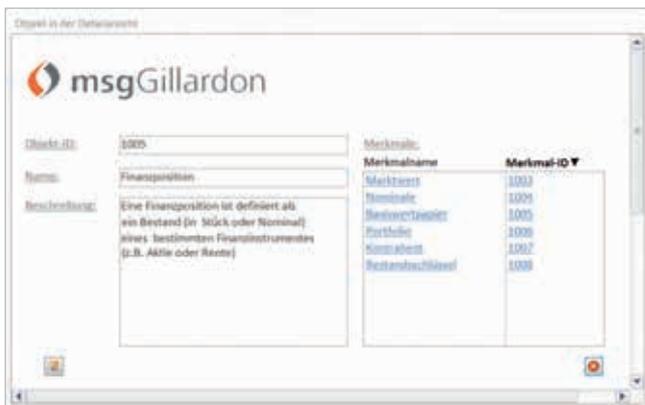


Abbildung 2: Fachliche Objektbeschreibung durch Merkmale

Die konzeptionelle Beschreibung kann dabei abhängig von gestellten Anforderungen und bereits erstellten Vorleistungen auf verschiedenen Meta-Ebenen erfolgen. Das Prinzip wird nachfolgend anhand eines durchgängigen Beispiels erläutert.

1. Ebene: Objekt- und Merkmalsebene

Auf dieser rein fachlichen Ebene werden Geschäftsobjekte und die für sie charakteristischen Merkmale beschrieben. Merkmale oder ganze Objekte können dabei bereits bestehend sein oder gewünschte fachliche Anforderungen darstellen.

Beispiel: Eine Finanzposition in einem Portfolio soll immer einen aktuellen Marktwert in Euro haben. Aus welchen tatsächlichen Quellen oder nach welchen Regeln die fachliche Anforderung, zu jeder Finanzposition einen Marktwert in Euro zu erhalten, später umgesetzt werden kann, muss jetzt noch nicht berücksichtigt werden.

2. Ebene: Verbindung zur Datengrundlage

Auf dieser Ebene wird als Nachdokumentation oder im Rahmen der Businessanalyse im Projekt spezifiziert, auf Basis welcher vorhandenen Daten Merkmale nach aktuellem Kenntnisstand gebildet werden können. Hierzu wird eine Verbindung zu der zur Verfügung stehenden Datengrundlage aufgebaut.

Wichtig: In Abgrenzung zu den eingangs beschriebenen technischen Metadaten in ETL-Entwicklungswerkzeugen werden die Datenquellen in diesem Schritt nicht „physisch“ referenziert. Es findet eine konzeptionelle Kopplung von „fachlicher Anforderung“ (Merkmal) zu einem „logischen Eintrag“ (Metadaten des Attributes) statt. Hierzu ist es notwendig, dem Institut oder dem Projekt prinzipiell zur Verfügung stehende Datenquellen in Form eines Quelldatenkataloges in einer Metadatenanwendung, wie dem Metadatentool MetaMaster, zu verwalten.

3. Ebene: Datenflussebene

Während auf der zuvor beschriebenen Ebene Verbindungen zwischen Merkmalen und Attributen nach „aktuellem Kenntnisstand“ und ohne Betrachtung von architektonischen Restriktionen her-

gestellt wurden, wird auf der Datenflussebene der Aufbau der tatsächlich umzusetzenden Schichtenarchitektur beschrieben.

Die zuvor angeforderten Daten sollen aus den Quellen in die DWH-Schicht transferiert werden. Fachliche Merkmale können optional ganz oder teilweise in Form von Berichtskennzahlen, Daten des DWHs oder der Ergebnisschicht aufgehen. Durch Datenflüsse wird beschrieben, welche Transformationen und Verbindungen zwischen (Quell-)Daten durchgeführt werden müssen, um ein Attribut der DWH-, Datamart oder Ergebnisschicht zu beladen.



Abbildung 3: Detailsansicht eines Datenflusses

Zu Attributen können zusätzlich Parameterlisten gepflegt werden. Diese Funktion schafft die Möglichkeit, sämtliche denkbaren Ausprägungen eines Attributs als Schlüsseltable an diesem zu verankern.

Neben der Erfassung von Metadaten unterstützt die Lösung MetaMaster ebenfalls bei deren Analyse. So kann auf Konzeptebene festgestellt werden, zwischen welchen Attributen der einzelnen Schichten innerhalb der BI-Architektur Wechselwirkungen bestehen. Die Impact Analyse auf ein DWH-Attribut, z. B. BESTAND_ID, zeigt alle in Verbindung stehenden Quellen und Abnehmer, inklusive der zwischen ihnen spezifizierten Datenflüsse an.

Abbildung 4 stellt exemplarisch dar, dass die Bestands-ID einer Finanzposition aus verschiedenen Quellsystemen gebildet wird. Weiterhin ist zu sehen, von welchen Zielsystemen aus auf die im DWH gespeicherten Bestandsdaten anhand der nun konsolidierten Bestands-ID zugegriffen wird.

4. Ebene: Spezifikationskopplung an tatsächliche Umsetzung

Abschließend kann, ebenfalls optional, die Kopplung zwischen spezifizierten Datenflüssen und tatsächlich erfolgter Umsetzung in einem ETL-Job vorgenommen werden. Das Metadatentool MetaMaster stellt hierzu eine generische Schnittstelle bereit, die von ETL-Entwicklungswerkzeugen wie dem SAS DI-Studio beliefert werden kann. Es erscheint dann innerhalb der Metadatenanwendung eine Beziehung zwischen dem beschriebenen konzeptionellen Datenfluss und dem realisierten ETL-Job.

Alleinstellungsmerkmale eines integrierten Metadatenmanagements

Die zuvor beschriebenen vier Ebenen werden in Business-Intelligence-Projekten auch dann durchlaufen, wenn kein zentrales Metadatenmanagement vorliegt:

- > Es werden fachliche Anforderungen erhoben.
- > Es wird geprüft, anhand welcher vorhandenen oder neu zu erhebenden Daten diese umgesetzt werden können.
- > Die Integration dieser Daten in die bestehende Architektur und Datenmodelle findet statt.
- > Es erfolgt eine technische Umsetzung der neuen Anforderung innerhalb der ETL.

Der Einsatz einer Anwendung für ein integriertes Metadatenmanagement bringt daher zunächst keine völlig neuen Vorgehensweisen im BI-Projektlauf mit sich. Signifikante Vorteile ergeben sich jedoch durch die für alle Beteiligten geschaffene Transparenz hinsichtlich der Zusammenhänge der einzelnen Schritte. Beteiligte sind sowohl Datenverantwortliche aus Fach- und IT-Bereichen, aber auch Projektmitarbeiter, vom Projektleiter über Business-

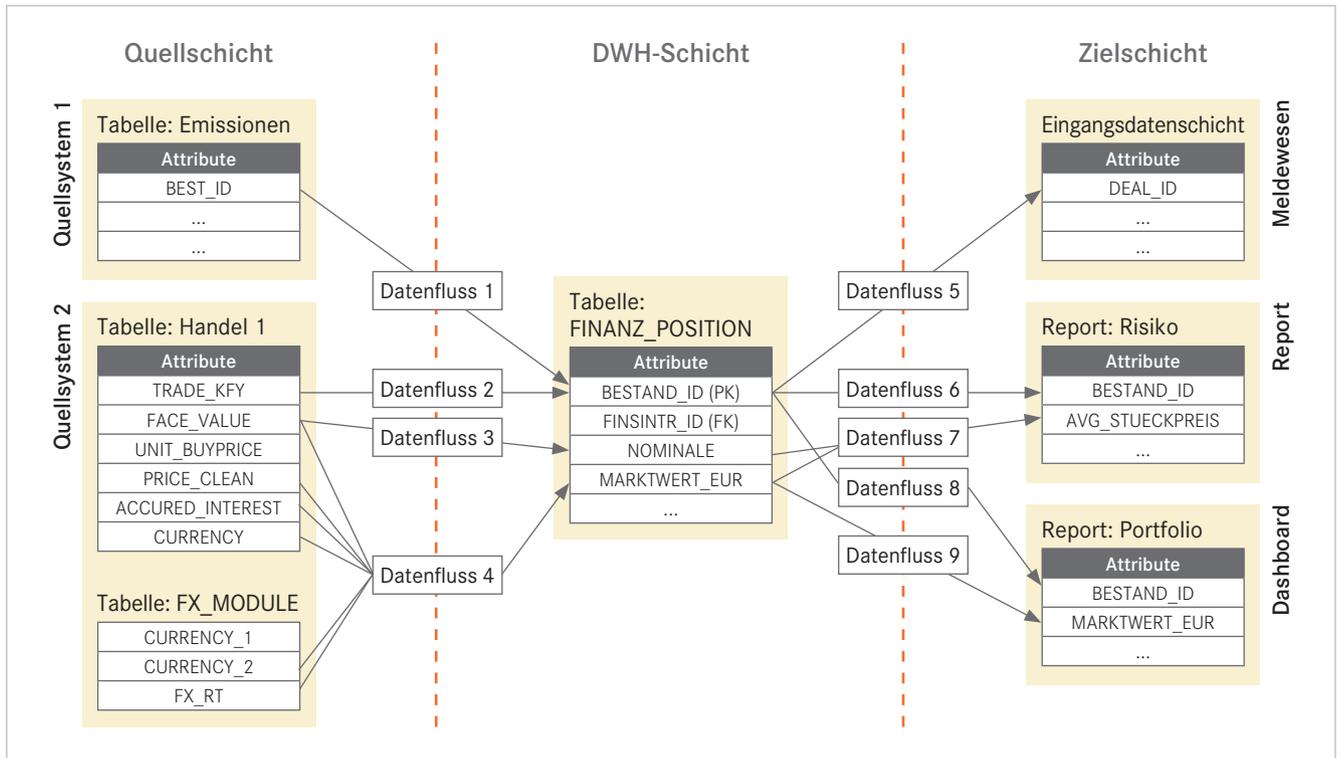


Abbildung 4: Ein vollständig integriertes Metadatenmanagement macht Wechselwirkungen innerhalb der Schichtenarchitektur transparent

analysten bis hin zum Entwickler. Neben diesen direkt Beteiligten steigt auch der Erfüllungsgrad neuer regulatorischer Anforderungen hinsichtlich der Transparenzvorgaben zum Datenbereitstellungsprozess, speziell im Bereich Risikomanagement, wie in den BCBS #239 formuliert.

Durch die zentrale Verwaltung sämtlicher Metadaten in einer Anwendung wie dem MetaMaster von msgGillardon, beziehungsweise seinem relationalen Datenmodell, verbunden mit einem Versionierungssystem, können Änderungen bei der Verwendung von Daten auch innerhalb des Zeitverlaufs stets nachvollzogen werden. Gerade bei Weiterentwicklungen und Änderungen an bestehenden DWH-Datenmodellen, Datamarts oder Quellsystemen können Auswirkungen auf verbundene Objekte an einer zentralen Stelle erkannt werden, ohne verteilt vorliegende und gegebenenfalls bereits archivierte Fach- und DV-Konzepte sowie

Umsetzungsprotokolle immer wieder mit großem individuellen Aufwand auf Abhängigkeiten zur angedachten Änderung analysieren zu müssen. Der ohnehin aufwendige Ausbau bestehender BI- und DWH-Systeme wird somit deutlich effizienter.

Autor



Markus Nicklas

Senior IT Consultant, CoC Financial Business Intelligence msgGillardon AG

- > +49 (0) 6196 / 7750 - 5453
- > markus.nicklas@msg-gillardon.de