

Im Datendschungel

Metadatenmanagement – Thema gelöst oder doch noch Handlungsbedarf?

von Markus Nicklas

An Artikeln mit Beschreibungen und Erfahrungsberichten zum Thema Metadaten und Metadatenmanagement herrscht kein Mangel – zumeist im Kontext von Business Intelligence oder Data Governance –, und auch in der NEWS 01/2014¹ haben wir bereits auf die innovativen Einsatzmöglichkeiten von Metadaten zur Unterstützung der Businessanalyse und Spezifikation in Datawarehouse-Projekten aufmerksam gemacht. Zudem bietet der Markt eine Vielzahl von Standardlösungen, die zumindest damit werben, ein umfassendes und konsistentes Metadatenmanagement zu ermöglichen.

Das Thema, so will es scheinen, wurde mittlerweile hinreichend umgesetzt. Dies spiegelt auch die Ersteinschätzung des Umsetzungsstands zum Thema Metadatenmanagement in den Instituten wider. Eine tiefer gehende Befragung im Rahmen unserer aktuellen, in Kooperation mit dem Handelsblatt erstellten Studie **banking insight. BCBS 239 – Überregulierung oder Impuls für ein richtungsweisendes Risikomanagement?** zeichnet jedoch – zumindest teilweise – ein anderes Bild: Je intensiver die Beschäftigung mit dem Thema in den jeweiligen Instituten erfolgte, desto kritischer fiel die Selbsteinschätzung aus.



Metadaten beschreiben nicht nur Daten

Der Begriff „Metadaten“ wird häufig als „Daten über Daten“, also Beschreibungen zu Daten, definiert. Diese Definition, so prägnant sie auch ist, gibt jedoch nur einen Teil der Reichweite des Themas Metadaten wieder. Neben den Beschreibungen von Tabellen und Attributen durch Informationen wie etwa technische Formate, fachliche Bedeutung, Wertebereiche oder mögliche Ausprägungen müssen für ein vollständiges Metadatenmanagement zwingend auch Beschreibungen zu Datenflüssen erfolgen.

Datenflüsse – Metadaten „zwischen Daten“

Datenflüsse entstehen immer dann, wenn Daten von einer Architekturebene in eine andere Architekturebene transferiert werden. Hierbei handelt es sich häufig nicht um einfache Eins-zu-eins-Übernahmen. Stattdessen erfolgen Filterungen, Aggregationen und Transformationen oder Kombinationen daraus.

- > **Filterung:** Datensätze mit bestimmten Ausprägungen werden nicht weiter berücksichtigt, zum Beispiel werden Objektsicherheiten mit Länderschlüssel <> Deutschland nicht weiterverarbeitet.
- > **Aggregation:** Mehrere Datensätze mit gleichem Schlüssel werden zu einem neuen Datensatz verdichtet, zum Beispiel werden alle Buchungen auf ein Konto zum Saldo summiert.
- > **Transformation:** Die durchzuführende Geschäftslogik beschreibt, wie aus Eingangsdaten ein Ergebnisdatum erzeugt wird. Zum Beispiel erfolgt abhängig von der Geschäftsart ein Ab- oder Aufschlag auf den anzusetzenden Geschäftswert.

Filterungen und Aggregationen werden dabei, wie bereits aus den Beispielen hervorgeht, auf Datensatzebene beschrieben. Das heißt: Zur Befüllung einer Zieldatentabelle in einer höher liegenden Architekturschicht werden zunächst die benötigten Quelldatentabellen identifiziert und im Anschluss daran Filterkriterien für beizubehaltende oder nicht weiter zu verwendende Datensätze fachlich beschrieben und technisch spezifiziert. Nachfolgend wird der fachliche Schlüssel gekennzeichnet, über den gegebenenfalls Aggregationen durchgeführt werden müssen. Für die zu aggregierenden Informationen werden Aggregationsfunktionen wie zum Beispiel SUMME, ANZAHL, MIN, MAX, Durchschnitt oder Median definiert.

Auf Feld- beziehungsweise Attributebene werden dann die granularsten Transformationsvorschriften beschrieben. Ein Datenfluss auf Attributebene besteht daher immer aus einem Zielattribut und kann mehrere eingehende Quellattribute besitzen. Diese eingehenden Quellattribute können über einfache (z. B. Bildung einer Prüfziffer per Quersumme) oder komplexe Funktionen und Logiken (z. B. Errechnung LGD) zur Zielinformation verarbeitet werden.

Die vollständige Metadatenerfassung zu bestehenden oder geplanten Datenflüssen ermöglicht sogenannte Data-Lineage-Analysen, die unter anderem einen erheblichen Anteil zur Erfüllung der BCBS-239-Anforderungen leisten können.

Vollständige Data Lineage

Wurden die Datenflüsse zwischen sämtlichen Architekturebenen vollständig erfasst, kann mit einer Abfrage (im msgGillardon-Tool MetaMaster „Impact Analyse“) eines beliebigen Attributs dessen Data Lineage nachverfolgt werden. Abbildung 1 visualisiert die Bedeutung von Data Lineage: Angezeigt werden sämtliche mit einem Attribut über Datenflüsse verbundenen weiteren Attribute. Somit kann auf Knopfdruck ein Geflecht von Wechselwirkungen über Architekturschichten hinweg dargestellt werden – eine wesentliche Fähigkeit, wenn es darum geht, die in BCBS 239 formulierten Anforderungen zur Risikoaggregation beziehungsweise den Aggregations- und Ermittlungsverfahren transparent zu machen.

1 Daten über Daten – Ballast oder Vorteile durch Metadaten, Markus Nicklas, NEWS 01/2014.

► Top-Story

Ausgehend von einer beliebigen Kennzahl im Gesamtrisikobereich kann zurückverfolgt werden, aus welchen Eingangsdaten diese gebildet wurden. Bei den Eingangsdaten zu einer einzelnen hoch verdichteten Kennzahl handelt es sich in der Regel ebenfalls um bereits ermittelte Werte. Auch zu diesen zuvor ermittelten Werten werden deren originäre Eingangsdaten angezeigt. Dieses Verfahren wird so lange iteriert, bis die tatsächlichen Basisinformationen im Regelfall also in der Architekturschicht der operativen juristischen Systeme erreicht sind. Die so sichtbar gemachte Kette von Abhängigkeiten übersteigt den Informationsgehalt eines „einfachen Metadatenansatzes“ mit ausschließliche fachlich ergänzten Feld- oder Kennzahlbeschreibungen an Umfang und Aussagekraft deutlich.

Vorteile auch außerhalb regulatorischer Anwendungsszenarien

Neben der exemplarisch für BCBS 239 erläuterten Vorgehensweise, um die Eingangsdaten einer Risikokennzahl bis hin zu den operativen Ursprüngen zurückverfolgen zu können, ergibt sich eine Vielzahl weiterer Möglichkeiten. Werden beispielsweise in einem Operativsystem Änderungen an einer Datentabelle, etwa die Vereinheitlichung von bisher in Cent gespeicherten Beträgen in EUR, vorgenommen, so können per Knopfdruck die bestehenden Abhängigkeiten angezeigt werden. Ohne eine solche Analysemöglichkeit entsteht oftmals ein beträchtlicher Aufwand zur Identifikation sämtlicher historisch gewachsener Abhängig-

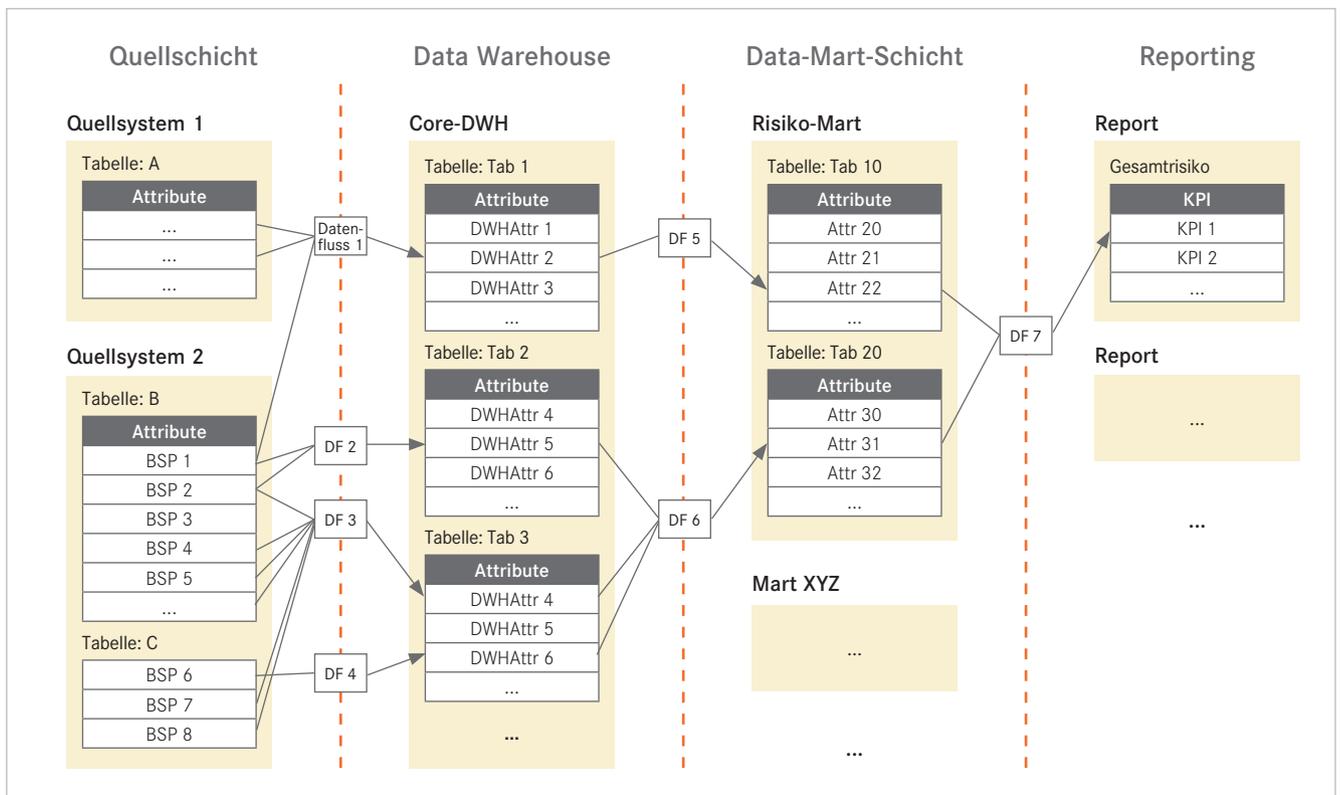


Abbildung 1: Vollständige Data-Lineage für eine beispielhafte Risikokennzahl

keiten. Weiterhin besteht die Gefahr, dass einzelne auf das Feld zugreifende Transformationen nicht angepasst werden, wodurch Fehler und Datenqualitätsprobleme entstehen.

Lesen Sie mehr zu diesem Thema in unserer aktuellen Studie **banking insight. BCBS 239 – Überregulierung oder Impuls für ein richtungsweisendes Risikomanagement?**. Erfahren Sie, welchen Umsetzungsstandard die von uns befragten Institute nach eigener Einschätzung haben, wo sie noch vom Zielbild abweichen und wie sie dem gegensteuern können.

Inhalt der Studie (Auszug)

- > **Data Governance:** Noch hapert es bei vielen Banken an der Grundvoraussetzung für BCBS 239: der Data Governance. Sie sorgt für eine Datenstrategie, die vom Vorstand aus alle Ebenen durchzieht. Problematisch: Der Vorstand selbst verantwortet die Steuerung der Daten in über der Hälfte der Fälle gar nicht, in über einem Viertel kennt er sie nicht im Detail.
- > **Datenqualitätsmanagement – Von einem Randthema zu hoher Relevanz:** BCBS 239 geht gleich an mehreren Stellen auf die Datenqualität ein – und rückt das Thema damit stärker ins Rampenlicht des Risikomanagements. So essenziell es ist, so groß sind hier auch die Defizite.
- > **Risikoberichterstattung:** Im Bankgeschäft lauern viele Wagnisse. Regelmäßige Risikoberichte sollen einen Überblick über die Lage geben. Viele Kreditinstitute kommen dabei aber den Zeitvorgaben der Aufsicht nicht hinterher und berichten unvollständig.
- > **Vier Fragen an drei Experten:** Wie Dirk Jäger vom Bankenverband, Ralf Goebel vom Sparkassen- und Giroverband und Andreas Leonhard von der BayernLB zu BCBS 239 stehen, wo sich die Fachleute einig sind und an welchen Stellen sich ihre Antworten unterscheiden.

msgGillardon-Metamaster – Vorgehen

1. Schritt: Datenkataloge erfassen: Zu sämtlichen Architekturschichten – im einfachsten Aufbau Quellsysteme, Datawarehouse und Berichte – sind alle Tabellen und Attribute vollständig zu erfassen.

2. Schritt: Fachliche Beschreibungen und Parameterlisten hinterlegen: Die Basis für jedes Metadatenmanagement, anders als häufig angenommen jedoch noch nicht das Zielbild, ist die Ergänzung fachlicher Beschreibungen und Ausprägungsmöglichkeiten zu Tabellenfeldern.

3. Schritt: Datenflussgruppen auf Tabellenebene anlegen: technische und fachliche Beschreibung von Verbindungsregeln zwischen Quelltabellen, Filtern und Aggregationsvorschriften für Datensätze.

4. Schritt: Datenflüsse auf Attributebene anlegen: Erfassung der geschäftlichen Transformationslogik auf granularster Ebene als Basis für die Data Lineage.

5. Schritt: Optionale Kopplung an die technische Realisierung: Kopplung zwischen Datenflussbeschreibungen (Metadaten) und tatsächlicher Realisation durch zum Beispiel ETL-Jobs (Extract – Transform – Load).

Autor



Markus Nicklas

Partner Business Analyse & Prozessmanagement,
msgGillardon AG

> +49 (0) 6196 / 7750 - 5453

> markus.nicklas@msg-gillardon.de