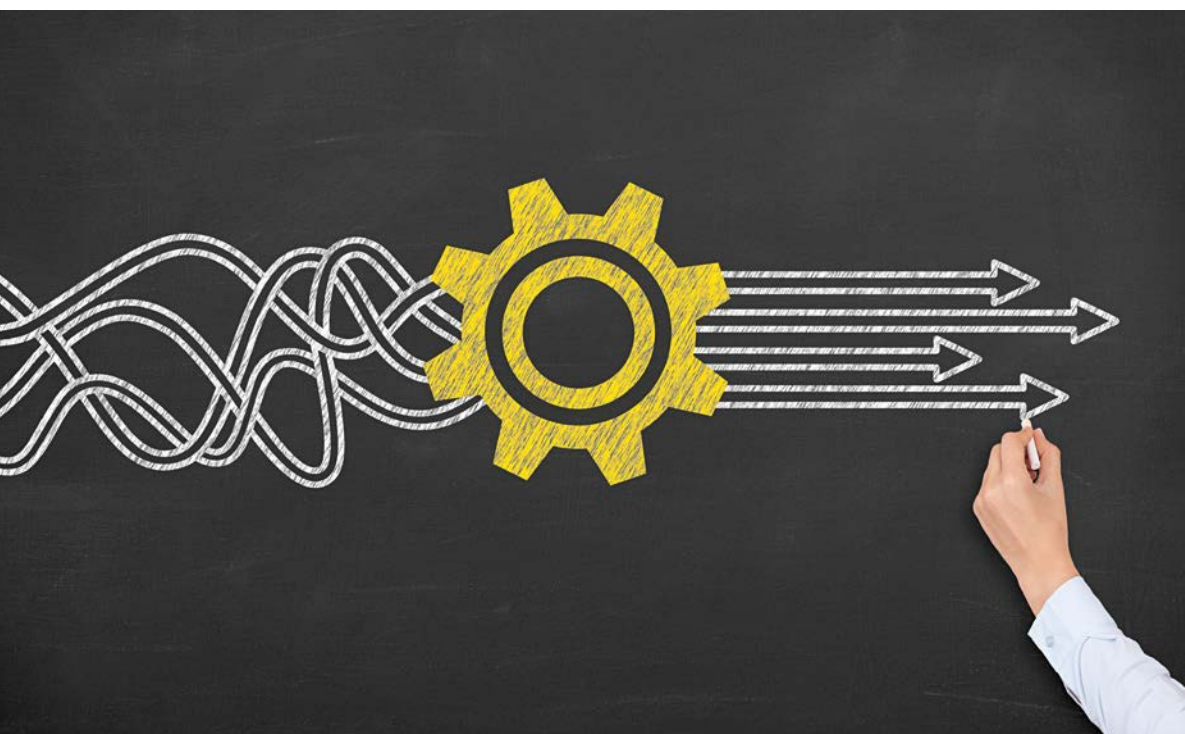


# Sprechen Sie Meta?

Metadaten gelten als die DNA aller Prozesse und damit eines Unternehmens. Damit haben sie eine große Bedeutung für die Geschwindigkeit und Qualität von Business Intelligence und Datawarehouse Projekten. Aber was sind Metadaten und wie nutzt man sie?



Business-, technische und operative Metadaten liefern wichtige Informationen aus einem Unternehmen.

**M**etadaten werden als die DNA des Datawarehouse bezeichnet<sup>1</sup>. Ich würde sogar noch deutlich weiter gehen: Metadaten sind die DNA aller Arbeitsprozesse und somit Ihres Unternehmens! Jeder sammelt Metadaten. Meist jedoch unbewusst und in unstrukturierter Form – aber kaum einer nutzt sie.

## Was sind eigentlich Metadaten?

Metadaten sind im Grunde Informationen über Merkmale anderer Daten, die es in drei

Arten zu unterscheiden gilt: Business Metadaten, technische Metadaten und operative Metadaten.

### Business Metadaten

Business Metadaten beschreiben einen Prozess – sowohl in der Sprache als auch aus Sicht der Fachabteilungen. Es geht also zum Beispiel um einen Beschaffungsprozess, ungeachtet der Frage, in welchem System sich dazu Daten finden. Neben der Beschreibung von Vorgehensweisen und Anforderungen sind hier auch Informationen wie

z.B. über den „Prozess Owner“ (Wer ist für ein Thema zuständig?) oder Fälligkeiten anzusiedeln.

### Technische Metadaten

Die technischen Metadaten hingegen liefern Informationen über technische Abläufe und darüber, welche Datenbanken genutzt werden. Sie beschreiben Datentypen, Tabellenaufbau und Spalteninhalte im Datawarehouse genauso, wie zum Beispiel Reihenfolgen und Zeitpunkte im Ablauf von Aktivitäten und Prozessen.

### Operative Metadaten

Operative Metadaten sind häufig ein (sehr nützliches) Abfallprodukt. Sie werden automatisch generiert und erfassen z.B. die Laufzeit von ETL-Prozessen, verarbeiteten Datenmengen und Zugriffszahlen.

Alle drei Typen haben eines gemeinsam: Sie beschreiben Daten (eine Ressource, eine Entität, ein Objekt)<sup>2</sup> und helfen somit, die Struktur und Art der in diesen Daten enthaltenen Informationen zu beschreiben und zu dokumentieren. Um diese Metadaten sinnvoll nutzbar zu machen, ist es unabdingbar, dass sie in strukturierter Form, z.B. in einer eigenen Metadatenbank, vorliegen. Unstrukturierte Daten mögen zwar für die Katalogisierung und Sammlung von Prozessinformationen ausreichend sein, eine Stan-

dardisierung oder gar Automatisierung wäre so jedoch kaum möglich.

### Welchen Mehrwert bieten Metadaten?

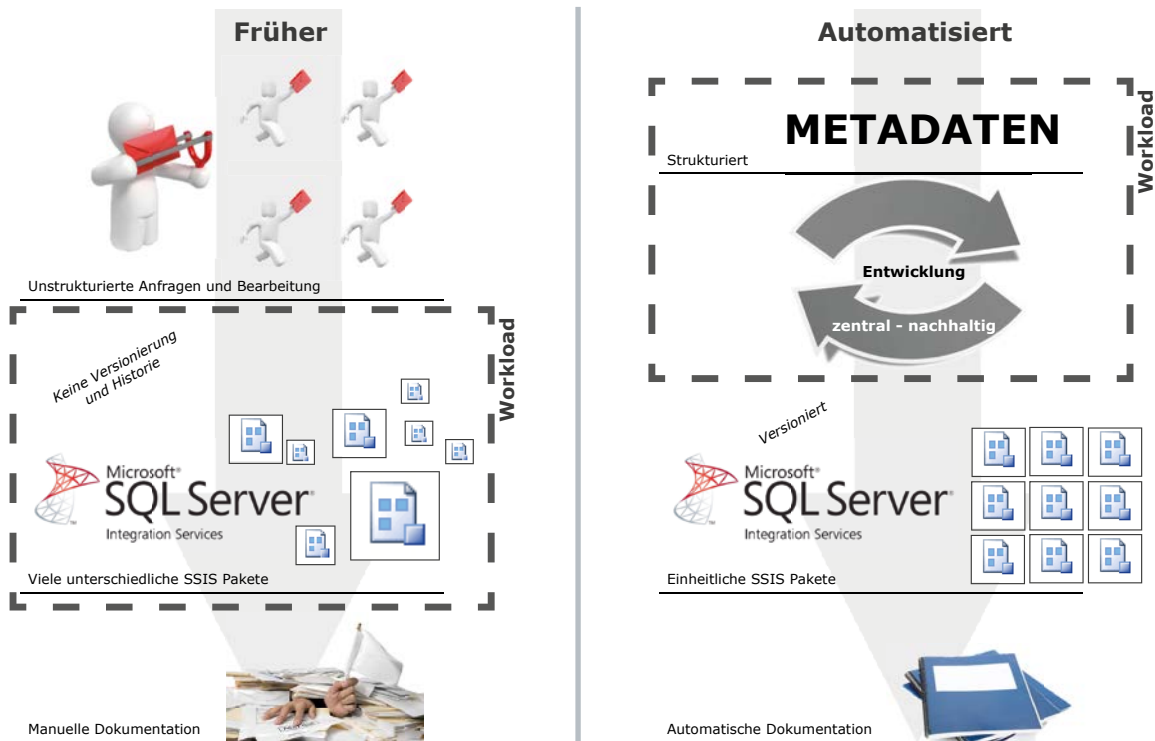
In erster Linie zwingt ein klares Metadatenmodell alle Beteiligten, sich an definierte Prozesse und Regeln bei der Übergabe zwischen Fachabteilungen und IT zu halten. Das Ergebnis ist eine klare, interne Übergabeschnittstelle von Business zu IT, da Anforderungen in strukturierter Form übergeben werden müssen, was Schattensystemen (Systeme und Strukturen, die nicht Teil der unternehmensweiten IT-Struktur und Strategie sind) und personellen Abhängigkeiten klar vorbeugt.

Durch Versionierung und Speicherung der Metadaten ergibt sich des Weiteren eine aktuelle und rückwirkend abrufbare Basis für jede Form einer aus den einheitlich aufgebauten Metadaten reproduzierbaren und zeitlich abgrenzbaren Dokumentation. Aufwändige, eventuell zeitlich verschobene oder inhaltlich abweichende Dokumentationsaktivitäten entfallen damit und die üblichen Anforderungen bezüglich Compliance („Wann wurde etwas geändert“-Szenarien und -Fragen) sind deutlich leichter zu erfüllen.

Eine solche greifbare Dokumentation stärkt das Verständnis des Business für die Herkunft von Daten. Die Überleitung von Business Metadaten zu technischen Metadaten schafft eine klare Terminologie, um

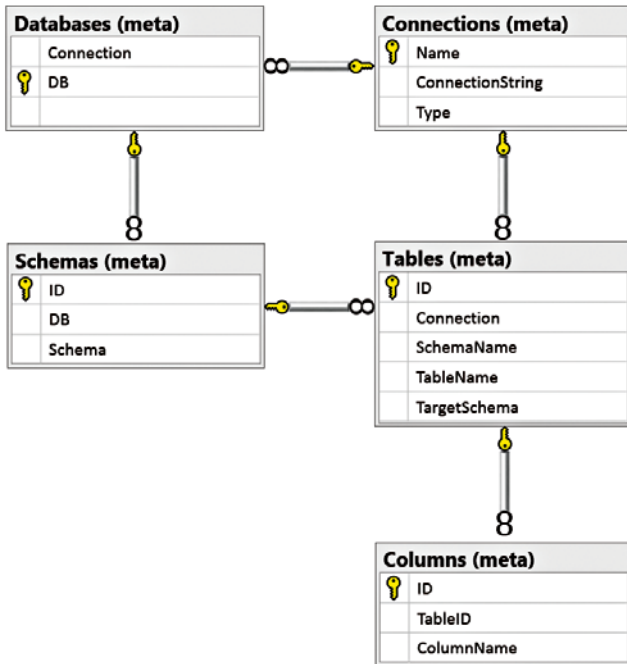
## Abb. 1: Entwicklungsprozess mit und ohne Metadaten

Klassischer, manueller Entwicklungsprozess vs. Metadaten getriebener, automatisierter Prozess



## Abb. 2: Beispiel eines Metadatenmodells

Auszug eines beispielhaften Metadatenmodells zur Beschreibung der Quelldaten eines Datawarehouse



zum Beispiel Begriffe wie „Umsatz“ oder „Verkaufsmengen“ eindeutig zu definieren – ein häufiges Problem in klassischen BI-Umgebungen.

### Einheitliche Metadatenmodelle als Optimierungsmotor

Darüber hinaus erlaubt ein klares, einheitlich strukturiertes Metadatenmodell auch die einmalige Programmierung von Logiken, welche automatisiert anhand der Metadaten parametrisiert werden können. Damit muss nicht mehr symptombezogen, reaktiv an einzelnen Problemfeldern gearbeitet werden. Allgemeingültige Standards können geschaffen werden, Qualität und Performance des gesamten Projekts verbessern sich damit automatisch.

Wie sieht ein perfektes Metadatenmodell aus? Diese Frage ist genauso einfach zu beantworten, wie die Frage nach dem perfekten Geschäftsmodell: Es ist so individuell wie Ihr Unternehmen und somit gilt nicht „one size fits all“.

### Elemente, die in allen Modellen zu finden sind

Es gibt jedoch durchaus Elemente, die in nahezu allen Modellen zu finden sind. Im Speziellen trifft dies auf technische Metadaten zu. Beschreiben Sie einen ETL-Prozess, so wären dies mit Sicherheit Informationen wie zum Beispiel:

- » Quellsysteme
- » Quelltabellen und -spalten nebst ihren Datentypen

- » Erforderliche Einzelpakete/-prozesse
- » Der Ablauf bzw. die Orchestration Ihrer ETL-Pakete

Auch hier gilt, wie so oft, weniger ist mehr. Starten Sie mit einem schlanken Modell und lassen Sie es mit Ihren Anforderungen wachsen. Ein zu komplexes Modell wird schnell unbeherrschbar und ist zum Scheitern verurteilt!

### Ist das technisch überhaupt zu bewältigen?

Definitiv: Ja! Wie bei vielen Themen, die auf den ersten Blick schwer beherrschbar erscheinen, lässt sich die Komplexität reduzieren, indem man die wesentlichen Bereiche trennt. Zur Pflege eines Metadatenmodells gibt es ganz unterschiedliche Varianten, die sich letztendlich an den vorhandenen Möglichkeiten und der Komplexität der Anforderungen orientieren.

Starten Sie auch hier schlank! Es empfiehlt sich meist nicht, direkt mit einem eigens entwickelten Frontend oder gar einem Stammdatenframework, wie den Master Data Services, zu starten. Ein in Excel verwaltetes Modell kann am Anfang deutlich pragmatischer sein. Erstellen Sie Ihr Modell zunächst an einem Flipchart oder ähnlichem und bewerten Sie so, ob ein entsprechend mächtiges Werkzeug notwendig ist.

Achten Sie jedoch darauf, dass das Modell versionierbar ist. Entweder in Form von Backups, einem Quellcodeverwaltungssystem oder durch Temporal Tables. Letzteres, wenn Sie das Modell in einer SQL Datenbank speichern. Sie antworten damit auf die interne Frage, wann etwas geändert wurde als auch auf eventuelle interne oder externe Compliance Themen!

### Automatisierung von Prozessen (wie z.B. ETL Strecken)

Sobald Ihre Prozesse einmal in den Metadaten beschrieben sind, lässt sich auch der eigentliche ETL-, Datawarehouse- und Data-



### Benjamin Weissman

und sein Team der Solisyon GmbH beschäftigen sich seit 2000 mit

Berichtswesen, Unternehmensplanung und operativem Controlling für namhafte Kunden weltweit – stets mit der richtigen Balance zwischen individuellen Prozessen und erprobter Best Practice. Er ist regelmäßiger Sprecher auf nationalen und internationalen Konferenzen.  
b.weissman@solisyon.de

mart-Prozess – das betrifft sowohl die Generierung als auch die Orchestrierung – hervorragend automatisieren. An dieser Stelle hat man die Wahl zwischen fertigen Lösungen und eher flexiblen Framework-Sprachen, wie z.B. Biml – der Business Intelligence Markup Language<sup>1</sup>. Der Weg über eine eigene Lösung in Biml hat den Nachteil, dass viele Teile einer bestehenden Lösung erst entwickelt werden müssen. Dieser Nachteil wird aber deutlich durch die gewonnene Flexibilität wettgemacht. Mit Biml lässt sich Ihr individuelles Modell 1:1 und ohne Schnittstellenverlust einbinden und abbilden<sup>2</sup>.

### Smarte Programmierungen

Durch einfache, smarte Programmierungen wird es ermöglicht, einzelne Schritte (wie das Befüllen einer Tabelle per „TRUNCATE AND LOAD“) nur einmal ausgestalten zu müssen. Somit können Änderungen an solchen Schritten (wie z.B. ein nachträgliches Protokollieren) jederzeit ergänzt werden. Eine zentrale Änderung wirkt sich automatisch auf alle Tabellen aus – egal ob es sich

um eine oder um 250 handelt. Ihr Datawarehouse generiert sich aus Ihren Metadaten ohne die Notwendigkeit ständiger manueller Programmierung.

### Dokumentation

Da unser Metadatenmodell bereits alle wichtigen Informationen enthält, können wir die Dokumentation daraus auch 1:1 erstellen und ableiten. Die äußere Form und das technische Vorgehen hängen hier etwas von der gewählten Technologie zur Pflege ab. Haben Sie sich für die direkte Pflege in einer Datenbank wie MS SQL oder MS Access entschieden, lassen sich daraus die entsprechenden Dokumente ableiten. Entweder direkt oder über zugehörige Tools und Komponenten, wie den MS Reporting Services. Das macht das häufig ungeliebte Thema Dokumentation wieder einmal – ohne Abstriche bei

Inhalt oder Qualität - zu einem Abfallprodukt. Da nicht manuell nachdokumentiert werden muss, ist die Dokumentation stets up to date.

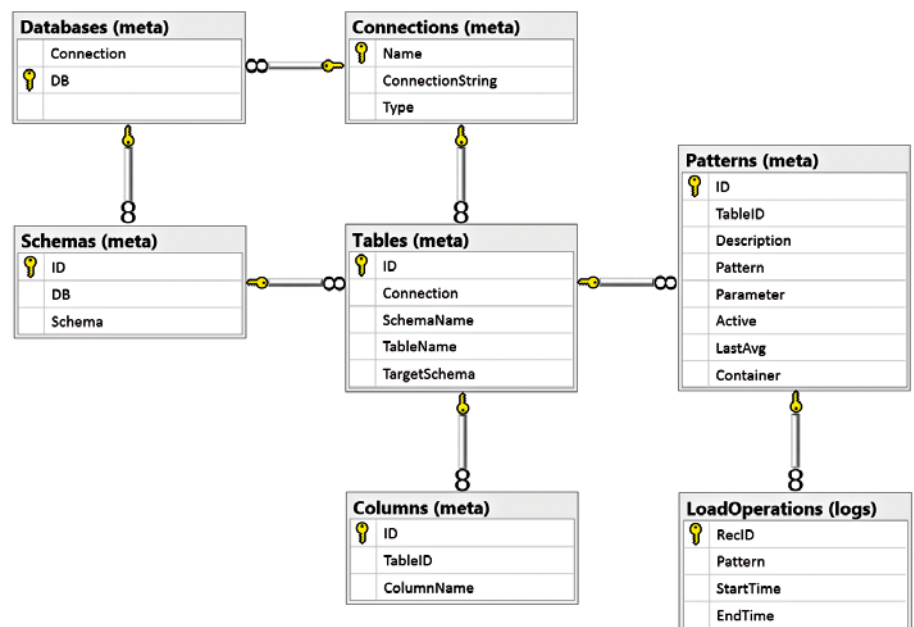
Apropos Abfallprodukt: Wir könnten nun zum Beispiel unser Metadaten-Modell um ein paar operative Informationen, wie z. B. Laufzeiten, ergänzen.

### Möglichkeiten zur Verbesserung

Mittels Technologien, wie Machine Learning, können wir proaktiv herausfinden, an welchen Stellen es Möglichkeiten zur Verbesserung von Performance und Stabilität gibt, anstatt im Notfall reagieren zu müssen. Auf diese Weise generiert unser System eigene Metadaten, lernt aus ihnen und passt sich den Gegebenheiten an. Somit sollten Metadaten kein Selbstzweck sein – können sich aber zum Selbstläufer entwickeln! ■

## Abb. 3: Erweiterung eines Metadatenmodells

Mögliche Erweiterung des Datenmodells zur Verwaltung und Überwachung verschiedener Lademöglichkeiten (Patterns) je Tabelle



1) <http://www.biml-blog.de/>

2) <https://www.solisyon.de/abgeleitetes-metadaten-modell-sql-server-biml-bimlexpress/>