

Data-Warehouse-Metadaten-Management: Aufgaben, Tool-Unterstützung, Projekt-Strategie

Dr. Jörg Westermayer und Andreas Weigl, SHS VIVEON AG

Metadaten-Management im Data Warehouse wird nach wie vor meist stiefmütterlich behandelt. Dieser Artikel zeigt die Notwendigkeit eines zentralen Metadaten-Managements, beschreibt Ansätze für eine Projekt-Strategie und stellt die Möglichkeiten des Oracle Warehouse Builders und der Oracle Business Intelligence Enterprise Edition zum Austausch und zur Konsolidierung von Metadaten dar.

Ohne Data-Warehouse-Daten ist eine ganzheitliche Steuerung der Geschäfts- und Entscheidungsprozesse in einem Unternehmen kaum mehr denkbar. Analog müssen DWH-Metadaten diese Funktion für das DWH selbst, seinen Betrieb, die Entwicklung und seine Nutzung übernehmen. Entscheidungen über die Auswirkung von Weiterentwicklungen, über den Ausbau der Infrastruktur bis hin zur Gestaltung des BI-Dienstleistungsangebots sind ohne Metadaten nicht fundiert zu treffen. Für die Anwender sind Informationen über die Definition von Kennzahlen, die Aktualität und Qualität von Daten unverzichtbar.

Doch des Schusters Kinder haben oft die schlechtesten Schuhe. Nach einer Studie aus dem Jahr 2004 (Florian Melchert, Metadaten-Management im Data Warehousing, empirische Studie, Universität St. Gallen) billigen zwar 97 Prozent aller Teilnehmer dem Thema Metadaten-Management hohe Relevanz zu, aber weniger als die Hälfte betreibt überhaupt Metadaten-Management. Und dort, wo solch eine Initiative besteht, überwiegen Lösungen, die lediglich einzelne Aspekte des Themas tangieren. Die Beratungspraxis zeigt, dass sich das Bild seit 2004 kaum verändert hat.

Um das Thema zu fassen, werden im ersten Abschnitt dieses Artikels die fachlichen, logischen, physikalischen und operativen Aspekte des Metadaten-Managements im DWH skizziert. Im zweiten Teil wird den Ursachen für den stiefmütterlichen Umgang mit Metadaten in der Praxis nachgegangen. Im dritten Abschnitt sind Lösungsan-

sätze für ein Metadaten-Management mit Oracle Warehouse Builder und Business Intelligence Enterprise Edition vorgestellt.

Metadaten im DWH und ihr Nutzen

Bereits anhand eines beliebigen Reports kann man vielfältige Kategorien von Metadaten anschaulich darstellen. Für die korrekte Interpretation und Auswertung der im Report dargestellten Zahlen werden Definitionen von Kennzahlen und der zugrunde liegenden Geschäftsregeln ebenso benötigt wie Informationen über die Aktualität der Daten oder über den Ansprechpartner zum Report. Diese Metadaten lassen sich auf einer fachlichen und einer organisatorischen Ebene abbilden (siehe Abbildung 1).

Ein Entwickler, der den Report wartet, findet Unterstützung durch

Informationen über den Ersteller des Reports, die Ladeläufe der Report-Basisdaten sowie die dem Report zugrunde liegenden Datenmodelle und fachlichen Metadaten. Diese Metadaten-Kategorien lassen sich auf einer technischen und einer organisatorischen Ebene wiederfinden (siehe Abbildung 1).

Metadaten-Management trägt wesentlich zu Effizienz in Entwicklung und Betrieb bei, es reduziert Einarbeitungs- und Wartungsaufwände und ermöglicht Abhängigkeits- und Impactanalysen bei Weiterentwicklungen. Die Kategorien der Metadaten im DWH zeigen, dass es notwendig ist, mit einem sehr weiten Metadaten-Begriff und klaren Kategorien zu arbeiten, um die Anforderungen aller Nutzer zu erfassen. Die Diskussionen über die Zielsetzung und Reichweite eines Metadaten-Managements im DWH

	Struktursicht	Betriebs- und Prozesssicht
Fachliche Ebene	Themen Kennzahlen Geschäftsregeln Dimensionen und Fakten	Qualitätskennzahlen Verfügbarkeit und Aktualität Reportnutzung
Technische Ebene	Data Warehouse Architektur Datenmodelle und BI Meta Layer Objekt-Abhängigkeiten Datenbewirtschaftungsketten (ETL) Ladeprozesse (Schedules)	Datenbewirtschaftungsprozesse Ressourcenverwendung Änderung und Entwicklung Datenqualitätsmonitoring
Organisatorische Ebene	Themenverantwort / Ansprechpartner Fachabteilungen DWH Abteilungen Berechtigungen	Requestmanagement Bauftragungen Projekte Budgets und Istaufwand

Abbildung 1: Metadaten-Kategorien

und ein gemeinsames Verständnis im Unternehmen sind nur durch eine frühe und klare Strukturierung einer Metadaten-Initiative zu führen.

Stand und Herausforderungen in der Praxis

Die große Diskrepanz zwischen der Bedeutung, die dem Metadaten-Management zugemessen wird und dem Stand der Realisierung ist sicherlich begründet durch die große Anzahl der generierenden und haltenden Systeme, die nicht oder nur unzureichend in ihren Metadaten integriert sind. Die Liste dieser Systeme ist lang und meist steht hinter jedem System ein anderer Hersteller: Business-Intelligence-Applikationen, ETL-Tools, CASE-Tools, Data Dictionaries, Versionierungs-Tools, Scheduling-Tools, Trouble-Ticket-Tools, Request-Management-Tools und so weiter.

Darüber hinaus werden viele Metadaten in der Konzeption von DWH-Projekten – aber auch während des Betriebs – in vielfältigen Office-Dokumenten erfasst, die dann verstreut abgelegt werden und für eine Nutzung jenseits des unmittelbaren Entstehungszweckes verloren sind.

Ein weiterer Punkt sind Metadaten, die gar nicht erst erzeugt werden. Das Spektrum der Versäumnisse ist hier weit und reicht von unzureichenden Konzepten über fehlende Kommentierungen in Programm-Code/-Tools bis zu lückenhaftem Logging der Ladeprozesse.

Metadaten-Initiativen stehen daher vor folgenden Herausforderungen:

- Die konzeptionelle und fachliche Integration der Metadaten aus den DWH-Systemen muss über ein Ziel-Metadatenmodell verfügen, das den Bezug zwischen den verschiedenen Metadaten-Kategorien herstellt und in der Lage ist, Struktur und Betriebsdaten aufzunehmen.
- Die Architektur des Metadaten-Management-Systems erfordert für die technische Realisierung der Integration Schnittstellen für die Extraktion, den Import und idealerweise den Austausch der Metadaten zwischen den Systemen und ein oder

mehrere Repositorien für die Speicherung. Hierbei ist insbesondere die Verarbeitung unstrukturierter Metadaten in Business-Dokumenten eine nicht zu unterschätzende Aufgabe.

- Zusätzlich ist zu klären, in welcher Form die Metadaten den Anwendern zur Verfügung gestellt werden, sei es durch DWH-Portale oder durch die Veröffentlichung und Verlinkung der Metadaten in BI-Tools und -Reports.

Vor allen Überlegungen zur Projektstrategie, Technologie und Architektur des Metadaten-Managements sollte die Einsicht stehen, dass jeder Ansatz ohne eine stetige Unterstützung durch einen definierten Prozess über den gesamten Lebenszyklus des DWH zum Scheitern verurteilt ist.

Lösungsstrategien für ein Metadaten-Management

Die geschilderte Herausforderung der Integration der Metadaten in ein system- und herstellerunabhängiges Meta-Modell will das „Common Warehouse Metamodel (CWM)“ meistern. Dies wird durch die Object Management Group seit 1999 mit führenden BI- und DWH-Herstellern definiert. Ziel ist eine gemeinsame Sprache und Struktur für den plattform- und systemunabhängigen Austausch von Metadaten zwischen den Systemen der Hersteller.

Das Ziel, solche übergreifenden Metadaten-Systeme zu schaffen, ist nach Einschätzung der Autoren in der Praxis verfehlt worden. Das Meta-Modell hat eine zu hohe Komplexität und verlangt erhebliche technische Erweiterungen bei der Abbildung von Metadaten. Im praktischen Einsatz findet man das CWM vornehmlich als Format für den Export von Metadaten aus den Tool-Repositorien.

Metadaten-Systeme können meist drei Architekturvarianten zugeordnet werden:

- Die zentrale Architektur ist das Zielbild, das hinter dem CWM steht. Alle Tools nutzen ein zentrales Metadaten-Repository, mit dem sie über Schnittstellen Metadaten in beide

Richtungen austauschen, gemeinsam nutzen und konsistent halten. Die Realisierung dieser Architektur-Variante scheitert meist daran, dass die Applikations-eigenen Metadaten-Repositorien zentrale Funktionalitäten der Applikation steuern. Diese hohe Abhängigkeit zwischen Metadaten und Applikation steht der herstellerunabhängigen Standardisierung und Zentralisierung in einem gemeinsamen Repository im Wege.

- Die dezentrale Architektur ist das Standard-Szenario für Best-of-Breed- und die meisten Hersteller-Lösungen. Jedes Tool besitzt ein lokales Repository und verwaltet dort die benötigten Metadaten. Zwischen den Tools erfolgt ein Austausch von Metadaten. Dieser erfolgt meist nur in eine Richtung (zum Beispiel vom ETL-Tool in das Reporting-Tool) und wird häufig nur einmal im Entwicklungsprozess zur manuellen initialen Versorgung der nachgelagerten Tools eingesetzt. Auch wenn die Tools die Metadaten in den Strukturen des CWM exportieren können, werden nicht immer alle relevanten Metadaten von den Export-Schnittstellen erfasst. Der Aufbau eines konsistenten Metadaten-Stands zwischen den Tools ist mit der dezentralen Architektur nur mit hohem manuellem Aufwand und unter der Bedienung einer Vielzahl von Schnittstellen möglich.
- Eine hybride Architektur berücksichtigt die Probleme und Begrenzungen des zentralen und des dezentralen Architektur-Ansatzes und versucht, die zentrale und konsistente Bereitstellung von Metadaten mit den Anforderungen der Tools in Einklang zu bringen. Die DWH- / BI-Tools exportieren ihre Metadaten in ein zentrales Repository, aus dem sie wiederum Metadaten der anderen Komponenten beziehen. Die Anzahl der Schnittstellen ist gering. Die Lösung verfügt über ein einheitliches Metadaten-Modell mit einem konsistenten Datenbestand im Repository, das für Analysen und Reports zur Verfügung steht. Die größte Hürde für die Implementie-

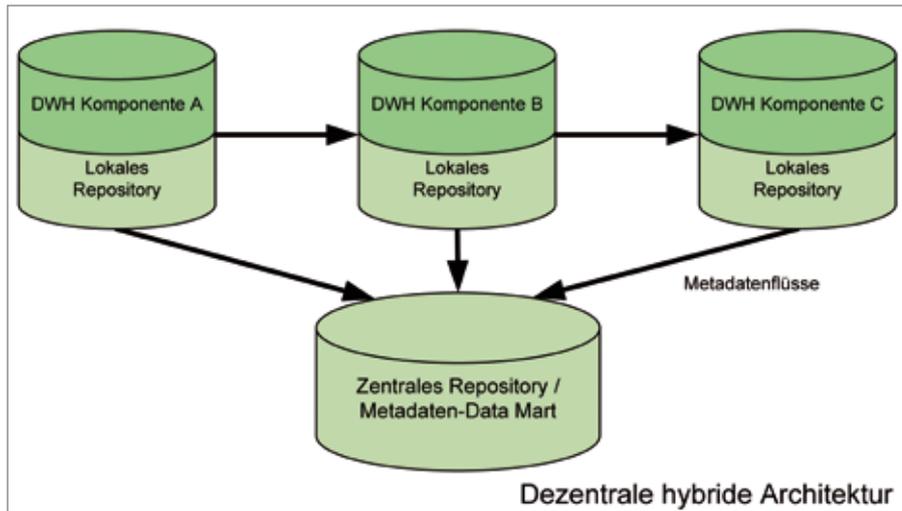


Abbildung 2: Extraktion und Sammlung der Metadaten in einem zentralen Repository

Die wesentliche Eigenschaft dieser Architektur-Variante stellt der wechselseitige Austausch von Metadaten zwischen dem zentralen Repository und den Tools dar.

Die Autoren verfolgen häufig einen pragmatischen unidirektionalen Ansatz, der die von den Herstellern unterstützten Schnittstellen zwischen den Tools als Quelle einsetzt und ein DWH-weites Metadaten-Management über ein oder mehrere Metadaten-Repositories beziehungsweise Data-Marts realisiert. Diese Metadaten-Data-Marts integrieren Metadaten-Extrakte aus den Repositories und stellen diese für Analysen und Reports zur Verfügung.

Der Aufbau des zentralen Metadaten-Management-Systems (siehe Abbildung 2) wird durch DWH-Lösungsansätze und ein DWH-Projektvorgehen realisiert. Die Tools werden als operative Quellsysteme betrachtet, die einen Data Mart mit Daten beliefern. Der Einsatz von „DWH-Bordmitteln“ und eines bewährten Projektvorgehens gewährleistet einen sicheren, gegebenenfalls stufenweisen Aufbau eines zentralen Metadaten-Managements. Der stufenweise Aufbau empfiehlt sich auch, um der Komplexität des Problems gerecht zu werden und frühzeitig den erforderlichen hohen Prozess-Reifegrad in der Organisation ausprägen zu können.

Die Nutzung der Tool-spezifischen Repositories und ihrer Schnittstellen

wird in diesem Projekt-Ansatz flexibel und nach Abwägung von Reifegrad und Nutzen integriert. Chancen und Grenzen eines Metadaten-Managements mit den Werkzeugen eines Herstellers zeigt der folgende Abschnitt.

Metadaten-Management mit OWB 11g R2 und BI EE 10.1.3

Die beiden Repositories von OWB und BI EE decken inhaltlich einen beachtlichen Teil der relevanten Metadaten-kategorien des DWH ab (siehe Abbildung 3). In dieser weiterhin dezentralen Architektur besitzen sowohl der OWB als auch BI EE ihre lokalen Repositories und verwalten dort ihre Metadaten. Ein

beachtlicher Teil der Metadaten wie Dimensionen und Attribut-Bezeichnungen wird dabei in beiden Systemen gehalten und verwendet, ein weiterer Teil der Metadaten wird zwar aus dem OWB in BI EE übernommen, dort jedoch weiter angereichert und bearbeitet. Andere Metadaten wie die Definition von Variablen, Zugriffssteuerungen und Lokalisierungen wie Mehrsprachigkeit werden dagegen nur im OBIEE Repository gepflegt.

Ein konsistenter Metadaten-Stand kann nur erreicht werden, wenn alle Objekte nur in einem Repository konfiguriert sind. Objekte, die in beiden Repositories verwaltet werden, sind daher durch Ableitung zu erzeugen, um die Repositories konsistent zu halten und den Aufwand manueller Nacherfassungen und Korrekturen zu verringern. Auch ein mehrfacher Austausch der Metadaten im Laufe des DWH-Lebenszyklus' muss hierbei konsistent durchführbar sein (siehe Abbildung 4).

Der OWB 11g R2 unterstützt jetzt die UDML-API des Oracle BI-Servers. Somit ist es möglich, ein komplettes und funktionsfähiges BI-EE-Repository aus den Metadaten des OWB abzuleiten, das alle gemeinsam verwalteten Metadaten-Objekte enthält. Dieses kann dann im Oracle BI-Administration-Tool weiterbearbeitet und um Objekte, die man nur im BI-EE-Repository verwaltet, ergänzt werden.

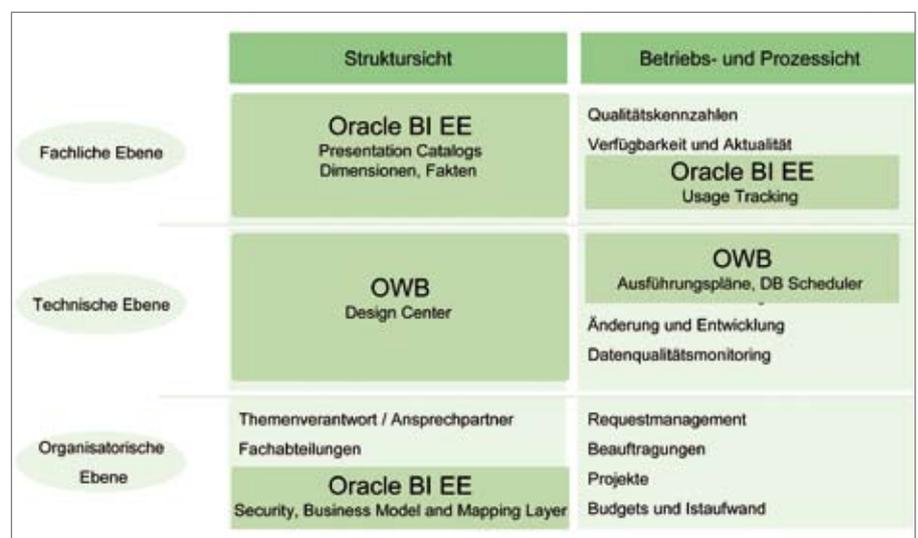


Abbildung 3: Metadaten in einer OWB- und BI-EE-Architektur

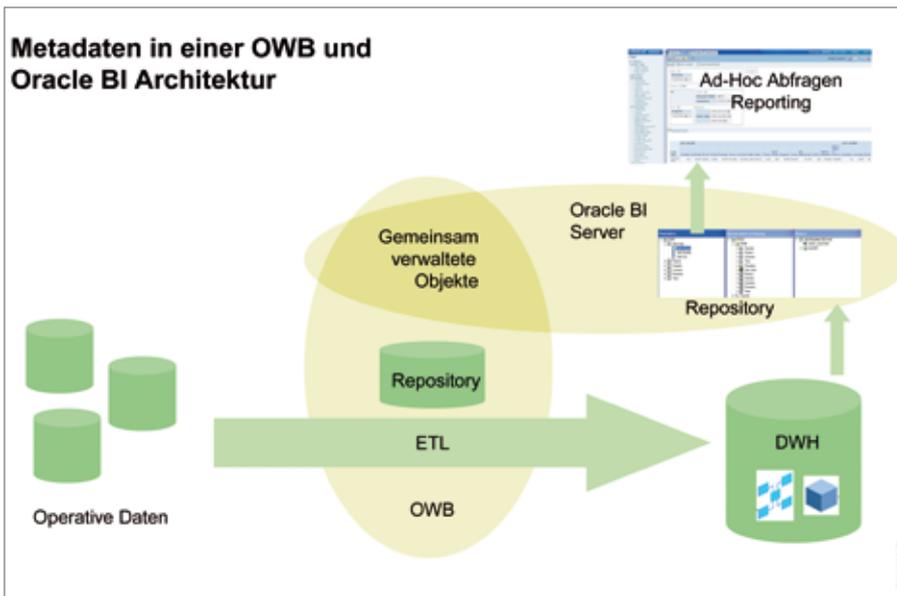


Abbildung 4: Architektur der OWB- und BI-Repositories

Die DWH-Strukturen und damit die verbundenen Metadaten ändern sich jedoch durch Weiterentwicklung und Wartung fortlaufend. Sowohl das OWB als auch das BI-EE-Repository müssen gepflegt werden. Da letztgenanntes in der ersten Version nun sowohl durch Ableitung aus dem OWB-Repository als auch durch manuelle Bearbeitung entstanden ist, muss eine wiederholte Übergabe der gemeinsam verwalteten Metadaten zwischen OWB und BI sicherstellen, dass Änderungen konsistent integriert werden können. Dazu lässt sich die Merge-Funktionalität des BI-Administration-Tools nutzen. Der Merge besteht aus der Ermittlung der Änderungen in den aus OWB abgeleiteten Repositories und der Übernahme dieser Änderungen in das Master-Repository (siehe Abbildung 5).

Die beschriebene Ableitung bildet einen DWH-Lifecycle ab, der Betriebs- und Wartungsprozesse des DWH unterstützt. Durch Weiterentwicklung operativer Systeme oder geänderter Reporting-Anforderungen werden Modifikationen der Metadaten veranlasst. Diese Änderungen werden zunächst im OWB-Repository vorgenommen und anschließend in ein BI-Repository überführt, das die gemeinsam verwalteten Objekte enthält. Die Korrekturen an den gemeinsam verwalteten Metadaten werden im Merge-Prozess in das

Master-BI-EE-Repository eingeführt. Anschließend lassen sich Erweiterungen und Ergänzungen mithilfe des BI-Administration-Tools pflegen und auf den BI-Server bereitstellen.

Oracle ist mit dem Merge der beiden Repositories einem integrierten und konsistenten Metadaten-Management ein Schritt nähergekommen, auch wenn man die Metadaten nur in einer Richtung austauschen kann. Allerdings bedarf das generierte BI-EE-Repository in der Regel manueller Nachbesserung, was ebenso wie der manuelle Merge notwendigerweise fehleranfällig ist.

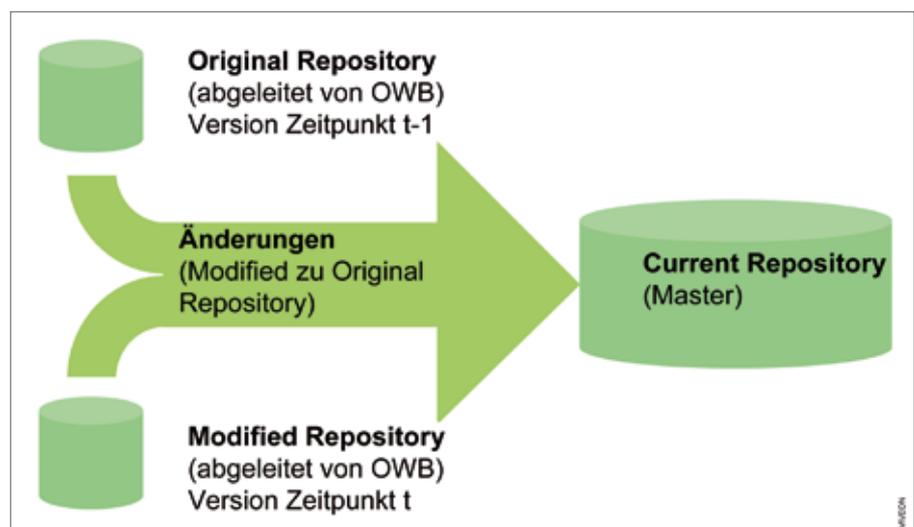


Abbildung 5: Merge der Repositories

Die konsistente Ableitung der Repositories über den gesamten DWH-Lebenszyklus hinweg belohnt jedoch mit dem entscheidenden Nutzen, jederzeit mithilfe des OWB-Repositories eine Abhängigkeitsanalyse der DWH-Objekte bis auf die Ebene des Presentation Layers der BI EE durchführen zu können.

Fazit

Metadaten-Management ist ein unverzichtbarer Bestandteil eines DWH, das als eigenständiges Projekt und mit eigenem Prozess zum Erfolg geführt werden muss. Zwar hat sich die technologische Unterstützung durch die Software-Hersteller verbessert, aber auch in DWH-Landschaften mit Applikationen nur eines Anbieters ist die konsequente Unterstützung eines zentralen Repositoriums noch nicht zufriedenstellend gelöst. Als Alternative integrieren Metadaten-Data-Marts die Metadaten von DWH-Sub-Systemen und bedienen zentrale Metadaten-Reporting- und -Analyseanforderungen. Hierbei sind natürlich die in den BI-Applikationen bereits vorhandenen Methoden zum Austausch von Metadaten zu nutzen. Metadaten-Management lässt sich mit den Strategien, Methoden und Technologien des Data Warehousing erfolgreich umsetzen.

Kontakt:

Dr. Jörg Westermayer
joerg.westermayer@shs-viveon.com