

# Data Lineage mit ODI und OEMM

Holger Dressing

Oracle Deutschland BV & Co. KG

## Schlüsselworte

Datenintegration, DIS, Metadaten, Metadatenmanagement, Data Warehouse, DWH, Enterprise Metadata Management, OEMM, Oracle Data Integrator, ODI

## Einleitung

Informationen für Unternehmensentscheidungen werden aus den zugrunde liegenden Daten abgeleitet. Daher ist es immer wichtiger, die Daten und deren Semantik nachvollziehen zu können. Daten durchlaufen verschiedene unternehmenskritische Anwendung und Lösungen. Sie werden extrahiert, transformiert und in andere Anwendungen hineingeschrieben. Dabei können Strukturen verändert und Kennzahlen gebildet werden, dessen Definitionen und Berechnungsvorschriften für alle Benutzer wichtig sind. Diese Aussagen gelten für alle Unternehmensprozesse und darauf basierenden Daten, das Thema wird aber insbesondere immer wieder bei BI und Data Warehouse (DWH)-Lösungen diskutiert. Auch bei BCBS 239 für Banken ist OEMM ein Teil der Lösung von Oracle.

In der Vergangenheit hat Oracle die Metadatenverwaltung in die DWH- und BI- Lösungen integriert. Die Metadaten aus den ETL Tools wie dem ODI könnten in die Metadaten der BI Tools (OBI EE) integriert werden. Neu ist eine umfassende Metadatenverwaltung als eigenständiges Datenintegrationsprodukt namens Oracle Enterprise Metadata Management (OEMM), das alle Metadaten der Datenbanken, der Modellierungstools, der ETL-Tools und der BI-Tools zentral verwalten kann. Nicht nur die Metadaten von Oracle-Produkten werden unterstützt, sondern eine umfangreiche Liste unterschiedlichster Anbieter aus dem BI und DWH-Markt. Es läßt sich ein unternehmensweites Glossar aufbauen, mit dem die Daten von den ausgehenden Quellsystemen bis hin zu den Zielsystemen unterstützt werden und umgekehrt (Lineage/Impact-Analyse). Die Key Feature lauten wie folgt:

- Report to Source Lineage
- Impact Analyse
- Modell-Versionierung
- Nutzt verbreitet Metadaten-Standards
- Erstellt und verwaltet ein Unternehmens-Glossar
- Importiert BI Metadaten marktgängiger Anbieter
- Importiert ETL Metadaten marktgängiger Anbieter
- Importiert Datenbank-Metadaten marktgängiger Anbieter
- Big Data Unterstützung

## Reduziert die Risiken und erweitert das Vertrauen in die Daten

OEMM reduziert die Risiken, die immer entstehen, sobald es Änderungen in den Datenstrukturen gibt. Das beginnt bei den Quellen und setzt sich fort bis in die Reports des Berichtswesens. Immer wenn Geschäftslösungen und Entscheidungssysteme auf den gleichen Datensätzen beruhen, ist es wichtig, deren vor- und nachgelagerte Strukturen zu kennen. Die GUI von OEMM erleichtert es, solche abhängigen Komponenten zu identifizieren und die Risiken zu minimieren. In Kombination mit Unternehmensglossaren, die eine erweiterte Transparenz ermöglichen, wird das Vertrauen auf die Daten im gesamten Unternehmen vergrößert.

## Technische Grundlagen

OEMM 12.2.1.0.0 benötigt eine Oracle DB, mindestens 11.2.0.3. Man kann eine komprimierte Datei herunterladen mit einem JDK und einem vorkonfigurierten Tomcat 7. Verfügbar ist das unter

Windows oder Linux 64-Bit auf Serverseite und einen Browser Firefox, Google Chrome, Internet Explorer oder Safari auf der Client-Seite. Es muß nicht auf den gleichen Systemen wie das Data Warehouse laufen, ein kleines Hardwaresystem ist vollkommen ausreichend. Alle benötigten Komponenten bis auf die Datenbank sind Bestandteil der Installation. Anschließend muß nur noch der Start des Windows Services für OEMM sicher gestellt werden.

Die GUI entspricht der Logik, wie sie bei anderen DIS-Werkzeugen üblich ist. Es gibt Baumstruktur, die die logisch zusammengehörenden Metadaten zeigt und einen größeren Bereich, in dem die Details der jeweils aufgewählten Aktionen dargestellt werden. Die Baumstruktur beginnt mit dem logischen Namen des Repositories, in das die Modelle der verschiedenen Systeme aufnimmt.

### Abnehmen und Verknüpfen der Metadaten

Der erste Schritt ist die Übernahme der Metadaten. Metadaten können aus Business Intelligence Werkzeugen, Data Integration Werkzeugen, Datenmodellierungs-Werkzeugen, Datenbanken, Metadaten-Werkzeugen, Objekt- und XML-Modellierungswerkzeugen übernommen werden. Es werden nicht nur Oracle-Produkte, sondern auch solche von vielen anderen Herstellern übernommen. Die Liste ist lang und kann auf OTN<sup>1</sup> eingesehen werden.

Als erstes muß zu dem entsprechenden Werkzeug eine Verbindung aufgebaut werden und die Metadaten ausgelesen werden. Falls es eine Oracle Datenbankinstanz ist, sind die üblichen Connect-Daten einzugeben und anschließend das Schema in der Datenbank auszuwählen. Falls Metadaten aus ODI eingelesen werden sollen, sind die ODI Connect-Daten wie User/Passwort und ODI Repository einzugeben.

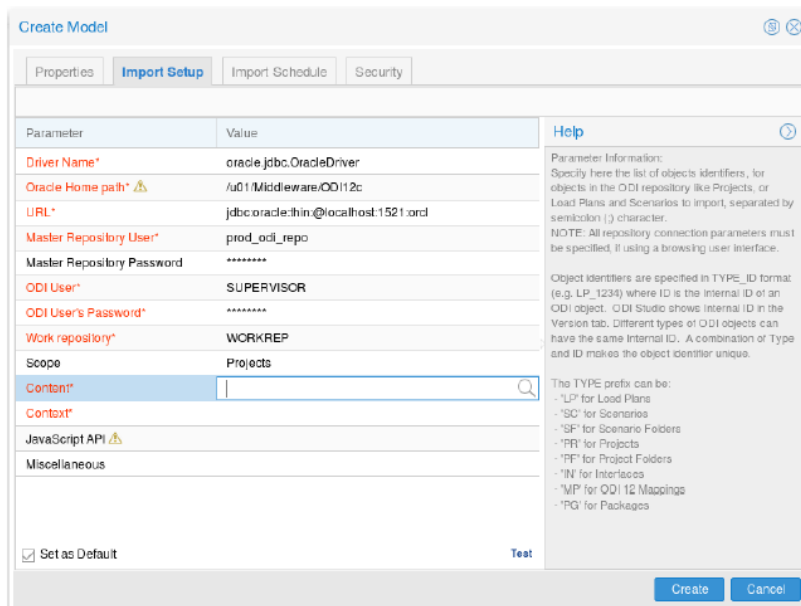


Abb. 1: Metadaten einlesen aus dem Repository des ODI

In den meisten Fällen wird man hier die Metadaten der Datenbanken der operativen Systeme und des Data Warehouses einlesen, ebenso wie die Daten der Modellierungswerkzeuge, der ETL Werkzeuge und das BI-Werkzeug. Anschließend müssen diese Daten konfiguriert werden, hier unterscheidet OEMM z. B. nach „OLTP Source“, „Load in Dimension“, „Multidimensionalen DW“ und „BI

<sup>1</sup> Siehe Link <http://www.oracle.com/technetwork/middleware/oemm/overview/index.html>, aufgerufen am 9.3.2015.

Reports“. OEMM versucht, die Attribute auf Basis von Namensgleichheiten innerhalb dieser Konfigurationen zuzuordnen, man kann hier auch manuelle Zuordnungen vornehmen. OEMM kann mehrere Versionen aller Metadaten verwalten. Im Prinzip ist dafür nur der Einlese-Vorgang zu wiederholen. OEMM wird dann die jeweiligen Metadaten mit einem Zeitstempel versehen. Es ist auch möglich, verschiedene Versionen miteinander zu vergleichen.

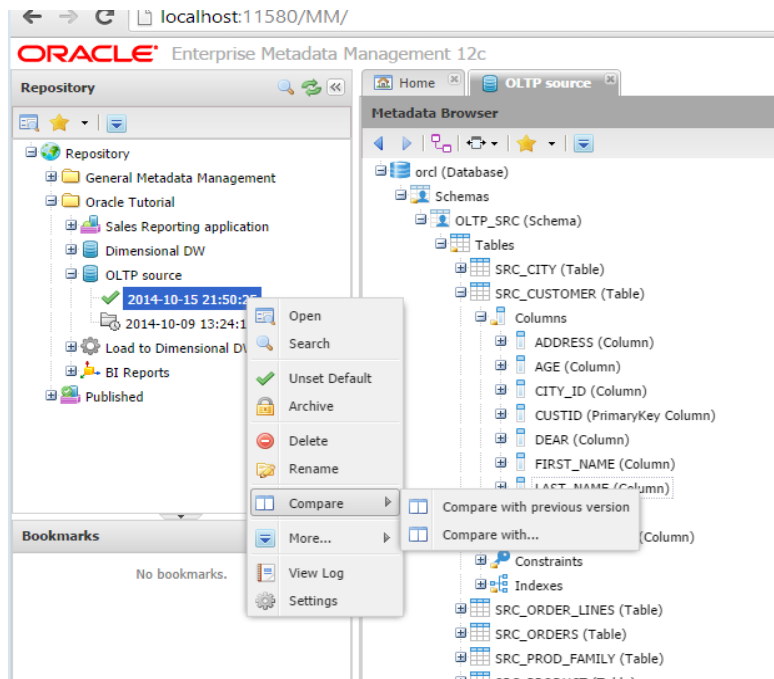


Abb. 2: Versionierung von Metadaten

### Durchsuchen der Metadaten mit Hilfe der Impact/Lineage-Analyse

Die Metadaten werden in den Strukturen der OEMM GUI angezeigt, in dem man die Daten nachverfolgen kann. Wenn man ein Mapping aus dem ODI in OEMM nachverfolgt, werden auch die spezifischen Strukturen von ODI dargestellt. Wäre es ein Mapping aus Informatica, würde man die entsprechenden Strukturen dieses Werkzeuges erkennen. In der Darstellung „Text“ kann der Inhalt aber auch quasi „neutral“ dargestellt werden.

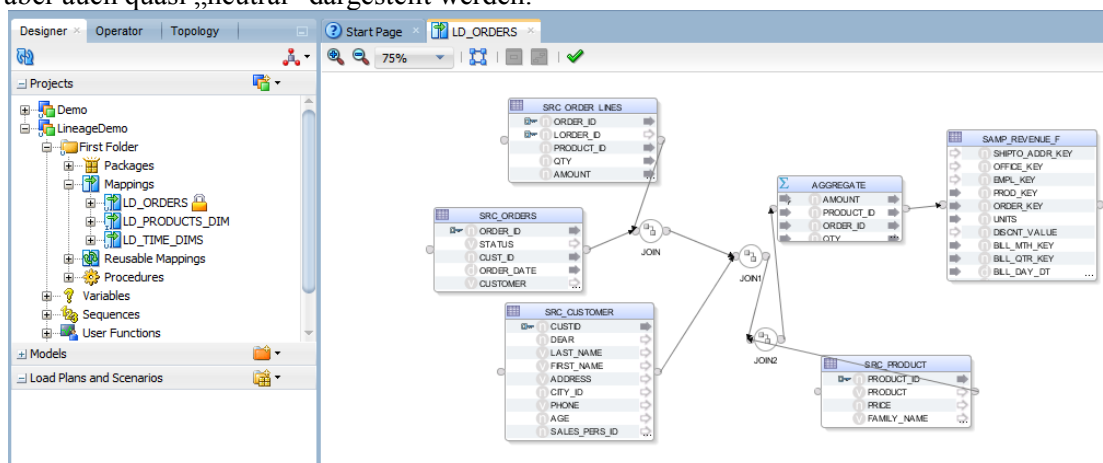


Abb. 3: Mapping in ODI

Wie im ODI und vielen anderen Werkzeugen gibt es ergänzende „Properties“-Fenster, um die Detailinformationen der Aggregationen und der Joins nachzuvollziehen.

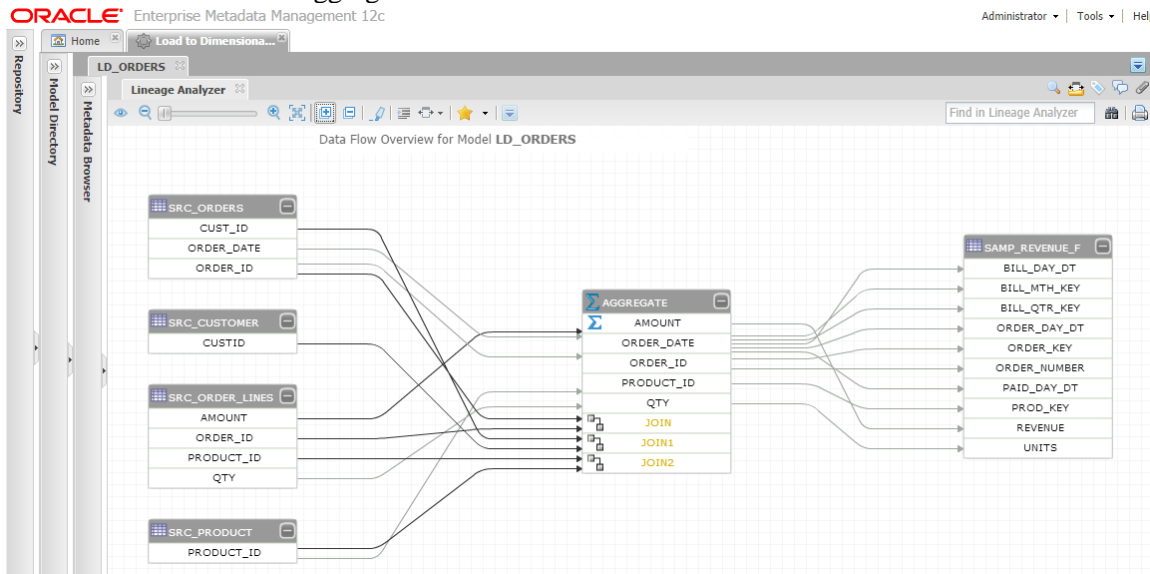


Abb. 4: Mapping aus dem ODI in OEMM

Darauf aufbauend kann man mit dieser Darstellung eine Lineage/Impact-Analyse erstellen. In Abb. 5 wird eine Lineage-Analyse gezeigt, vom BI Report über das ETL Tools bis zur Datenbank (ganz links dargestellt). Solche Analysen kann man einzeln für ein ausgewähltes Attribut oder auch über alle Attribute und Kennzahlen erstellen. Ebenso kann eine Impact-Analyse ausgehend von den Quellen erstellt werden, quasi in einer analogen Darstellung.

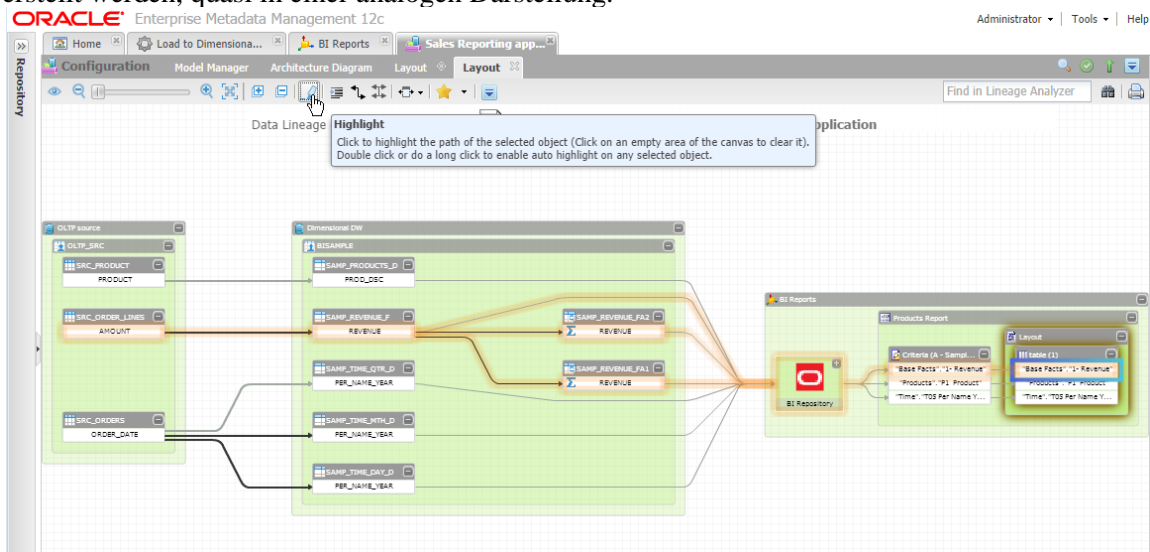


Abb. 4: Lineage Analysen auf Attributebene

Die gleiche Vorgehensweise ist auch mit Werkzeugen anderer Anbieter möglich. In der nachfolgenden Abb. 5 wird Erwin von CA für die Datenmodellierung eingesetzt. In einem solchen Fall kann es vorkommen, dass es Lücken in der Lineage-/Impact-Analyse gibt, die nicht automatisch gefüllt werden. Dafür gibt es ein Mapping-Feature, um solche Lücken manuell zu schließen oder anzupassen.

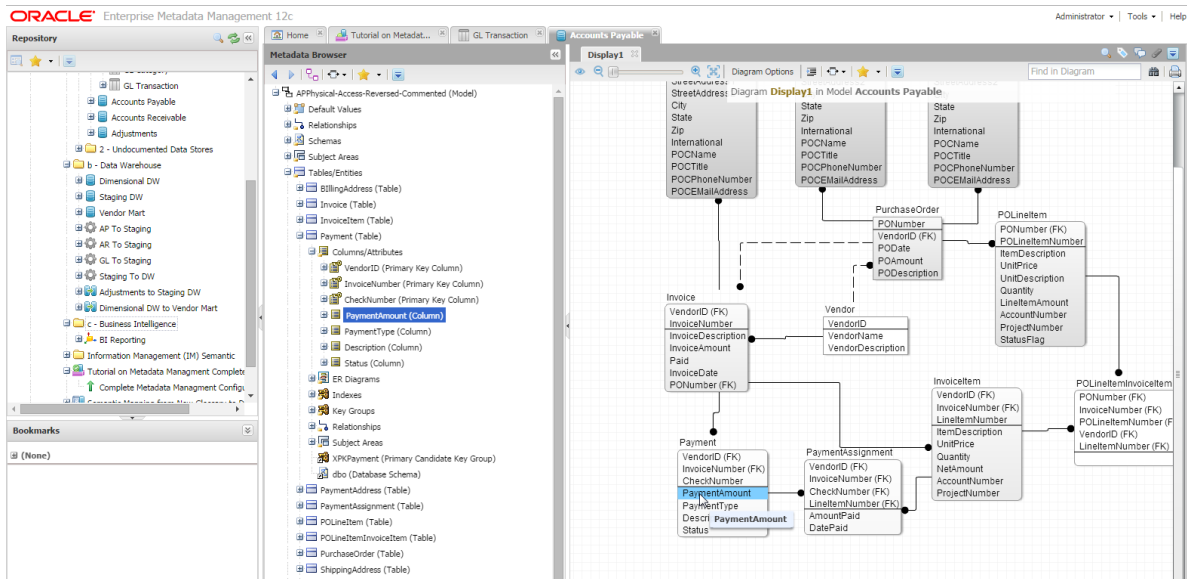


Abb. 5: Datenmodell aus Erwin in OEMM

### Business Glossar

OEMM bietet ein ISO 11179 basiertes Business Glossar für die Entwicklung und das Management aller Daten, um unternehmenweit Begriffe, Standards, Geschäftsregeln, etc. zu erfassen, zu definieren und zu pflegen. Zwischen dem Quell- und dem Zielmodell gibt es sogenannte semantische Mappings, die den Zusammenhang der Elemente zwischen dem Quell- und dem Zielmodell beschreiben. Ein Business Glossar ist in verschiedene Kategorien aufgeteilt, das Begriffe, Domänen oder andere Kategorien enthält. Begriffe und Domänen können bereichsübergreifend verlinked sein. Diese werden als semantische Mappings bezeichnet. Die Annahme ist, daß damit ein semantisches Modell über alle Metadaten und deren semantische Mappings erstellt wird.

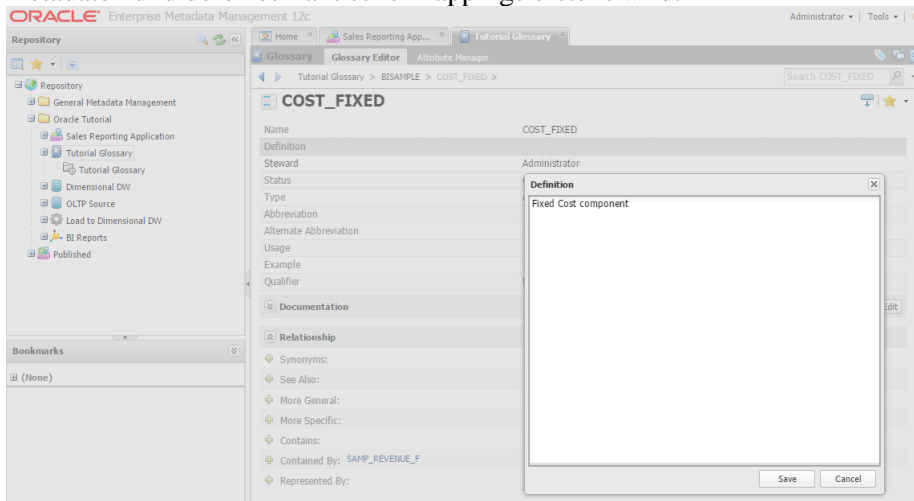


Abb. 6: Glossary Editor

Wenn ein Anwender nach einem Begriff in einem solchen Modell sucht, können damit alle Definitionen, Domänen, Geschäftsregeln und Kennzahlenberechnungen gefunden und verglichen werden. Wenn neue Metadaten nach OEMM geladen werden, versucht das System selbständig entsprechende Strukturen aufzubauen. Der Anwender kann eigene semantische Mappings erstellen und damit das Glossar ergänzen. Damit können Auswirkungs-/engl. *Impact* (Wo wird etwas benutzt?) und Abstammungs-/engl. *Lineage* (Was bedeutet diese Definition?) Analysen durchgeführt werden.

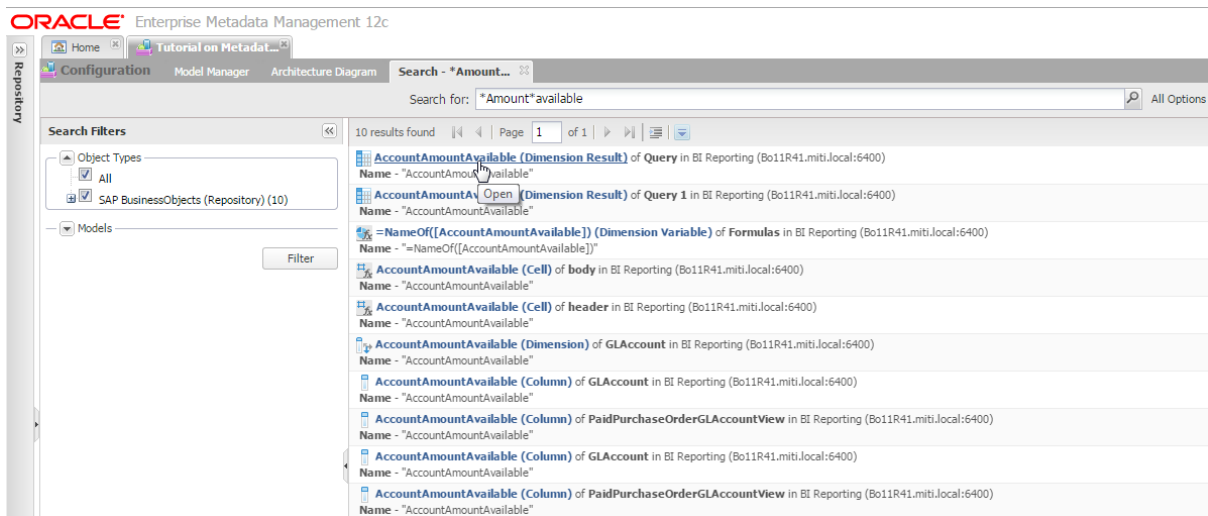


Abb. 7: Ansicht im Glossar

Damit ergibt sich die Möglichkeit, nicht nur über die Modelle in die Analyse einzusteigen, sondern auch das Glossar mit den Definitionen als Ausgangspunkt zu nehmen und aus dem Glossar heraus eine Lineage-/Impact-Analyse zu starten.

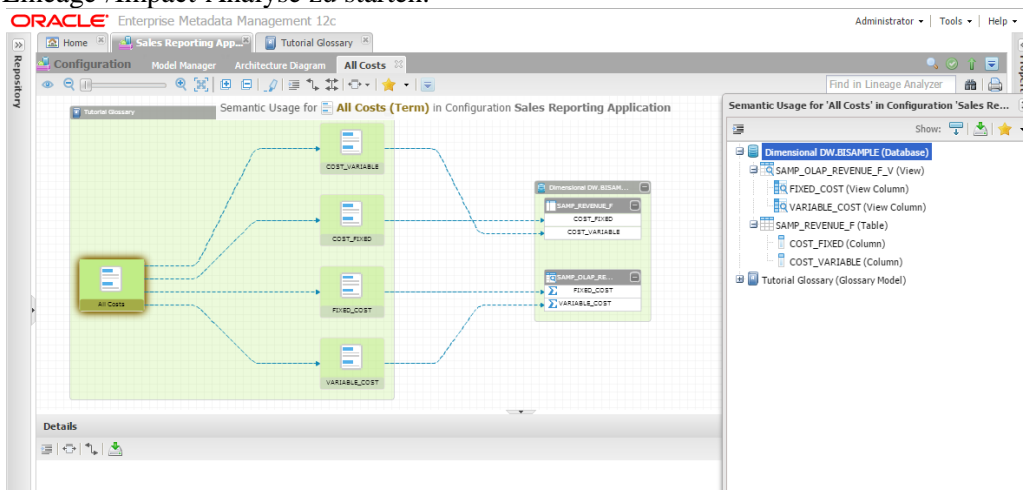


Abb. 9: Lineage Analyse zu den Definitionen im Glossar.

## Ausblick

Metadatenmanagement bei Oracle war auch vor OEMM nur durch die Integration verschiedener Oracle Produkte machbar. Der OEMM geht jetzt noch weiter: Er deckt viele marktgängige Produkte von der Datenbank über die Modellierungs- und ETL-Werkzeuge bis zum Reporting-Bereich ab. Wünschenswert wäre, die Metadaten nicht nur importieren zu können, sondern auch in OEMM zu editieren und anschließend in die verschiedenen Werkzeuge zurückspielen zu können.

## Kontaktadresse:

Dr. Holger Dresing  
Oracle Deutschland B.V. & Co. KG  
Kühnehöfe 5  
D-22761 Hamburg

Telefon: +49 (0) 89091-056  
Fax: +49 (0) 89091-056  
E-Mail: holger.dresing@oracle.com  
Internet: www.oracle.de