Metadata Management Tutorial

BI Documentation Reuse Using Meta Integration® Metadata Management (MIMM)

TABLE OF CONTENTS

TABLE OF FIGURES

1 Introduction

The need for more sophisticated and precise metadata management is a growing concern for most large organizations. Nearly all components that comprise modern information technology, from CASE tools, ETL engines, Warehouses, BI, EAI environments, as well as metadata repositories, contain, and often derive their processing from, metadata. The metadata for these environments can sometimes contain a wealth of documentation metadata (descriptions, domains, business rules, etc.). However, these nuggets of very useful information is generally distributed and duplicated throughout the information infrastructure, in BI design packages, embedded metadata, data models, etc.

This document provides tutorial guidance identifying many methodologies for identifying, collecting/reverse engineering, centralizing and normalizing this metadata documentation. The goal is generally one or more well-managed glossaries, Physical Data Models and/or conceptual models providing good documentation metadata to business users, engineers, data scientists and data analysts.

Disclaimer

Some of the features detailed in this document may not apply and/or be available for the particular Meta Integration® Metadata Management (MIMM) edition you may have.

1.1 How to use this document

It is certainly possible to skip through the tutorials, and thus simply glean an "management-level understanding" of Meta Integration® Metadata Management (MIMM) and its use within a metadata management environment. However, it is not recommended that one try to skip parts of the tutorials and then try to go through later parts. When following through the tutorial sections, it is very important to respect the order of the steps (and the order sections/labs within each section). The results of preceding tutorials are re-used and built upon in each successive lesson.

In addition, it is important to ensure complete understanding of the conceptual background provided in the sections leading up to and supporting the tutorial material. Thus, one should not simply jump into the tutorials with carefully reviewing the concepts presented in that section.

As this document include hand-on tutorials, a great deal of specificity is required. This detail includes specifying particular CASE, ETL, BI, etc., vendor's tools. While Meta Integration® Metadata Management (MIMM) environment itself is capable of working with over 100 different versions of third-party tools (see http://www.metaintegration.net/Products/MIMB/SupportedTools.html), it is necessary for the clarity conciseness of the tutorials to limit the cadre of tools that will be referred to. Please note that it is not necessary to have these tools on-hand to get the full benefit of the tutorials. Remember also, though you may intend to use Meta Integration® Metadata Management (MIMM) suite of tools with many of the supported third-party tools not specified in the tutorial, it is still quite valuable to learn the processes, methods and best practices presented here. Then one may reuse what one has learned and apply that knowledge and skill to the particular set of tools that are critical to one's own enterprise.

1.2 Conventions used in the tutorial

The following font conventions will be used throughout the tutorial.

- User Interface item – New
- Submenu item – New › Folder
- Terminology item – *model content* item
- Name or label reference – Accounts Payable

2 BI Design Documentation Reuse as Glossary

As pointed out, documentation metadata may reside anywhere in an organization. Many times, the most useful and accurate of this information is already being used as documentation for a subset of your architecture, including:

- External glossaries in CSV format
- Conceptual and logical models
- BI tools, especially in the design layer
- Data formats which carry their own metadata (e.g., XML)

Meta Integration® Metadata Management (MIMM) is built around a set of bridges that can reverse engineer this information and then forward engineer it into

- Glossaries and related Physical Data Models
- Conceptual and logical models
- or even other targets like
    - BI tools, including the design layer
    - DI/ETL tools
    - Data formats which carry their own metadata (e.g., XML)
    - Object models for implementation

We will use the example here, that is quite common, of a heavily documented BI tool design layer that we wish to bring into a master glossary. In addition, we will use the data lineage to reverse engineer the semantic mappings between the data warehouse and those glossary terms (derived from the OLAP model in the BI).

2.1 Overview

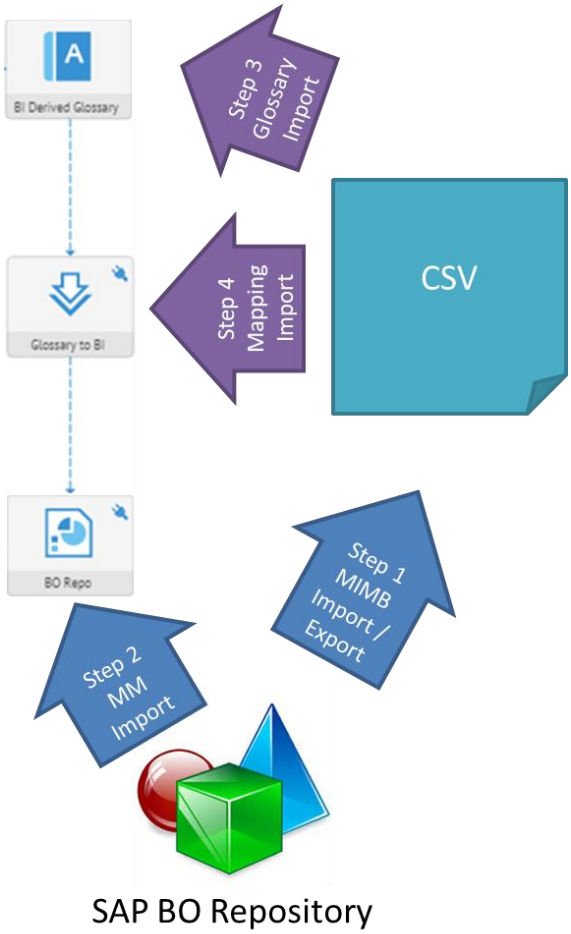


Figure 1 - Documentation Reuse Process Overview

In the above diagram, one may see the process we will step through. We will leverage both the documentation metadata (business names and descriptions) to identify and reverse engineer a glossary with related terminology. In addition, we will leverage the data flow architecture or lineage (stitching between the data warehouse and BI connection definition as well as the lineage in the BI tool itself to the documented objects on the design layer) to determine semantic mappings from the glossary to a Physical Data Model of the data warehouse.

2.2 Generate a Glossary (Terms and Mapping) CSV File (from BI)

The original documentation metadata is in the BI tool, so we will harvest directly from there. In this case, we will import the BI Reporting environment from SAP Business Objects as used in the Tutorial. We are harvesting two universes, or design models, from the tool. Those two reuse the same connection into the Dimensional DW (warehouse) database and there are six reports on one universe and one on the other:

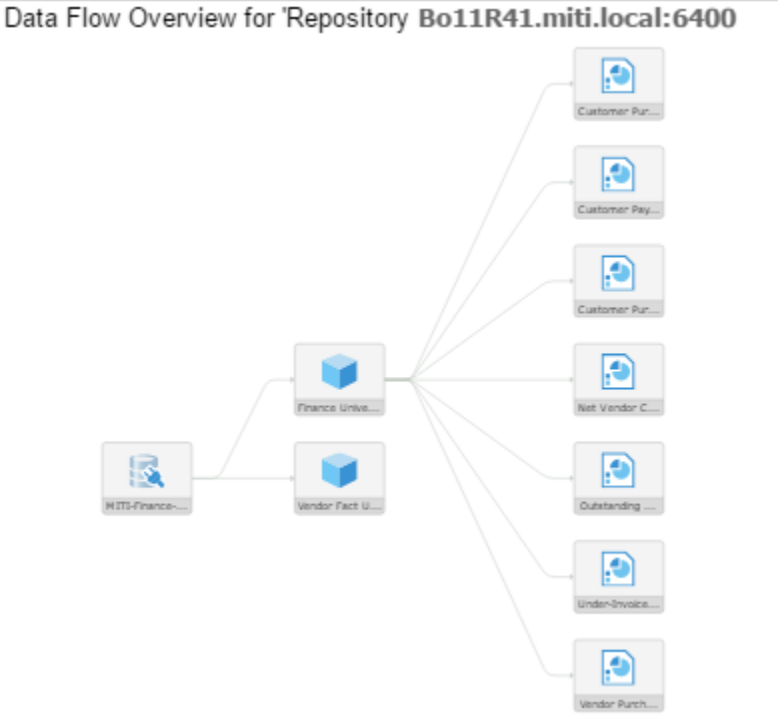


Figure 2 - Data flow overview of BI Reporting

Warning Disclaimer:

Although your BI metadata is already harvested in your Meta Integration® Metadata Management (MIMM), it currently cannot be directly exported (e.g. to a Glossary CSV file) because the original metadata (MIR XMI) is not in the Meta Integration® Metadata Management (MIMM) cache ($HOME/data/files), but rather in the Meta Integration ® Model Bridge (MIMB) cache for incremental harvesting ($HOME/data/MIMB).

Therefore, in order to generate the Glossary CSV file from your BI metadata**, you first need to obtain access to the standalone MIMB Web Application software and license, which is not part of your Meta Integration® Metadata Management (MIMM) Software**.

At this point you can import (again) your BI metadata and export it a Glossary CSV as explained below

2.2.1 Import from your BI Design

To import the original documentation metadata from the BI tool, we will harvest that directly into the MIMB Web Application.

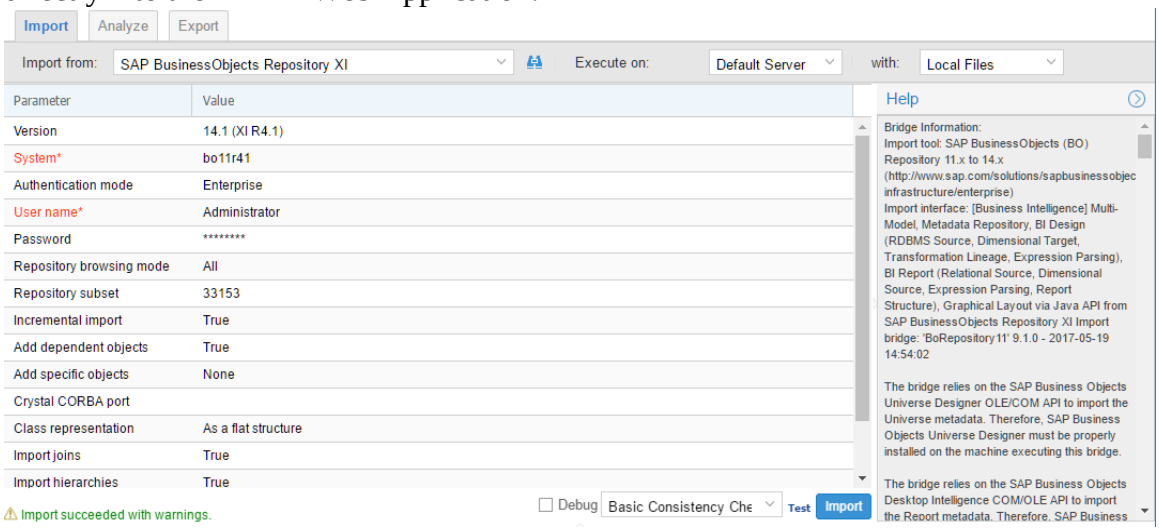


Figure 3 - Harvesting the BI tool Metadata

We are using MIMB as it will allow for a complete import of all the contents of the BI tool without profiling it, which would be required if we harvested directly into Meta Integration® Metadata Management (MIMM).

To import, merely populate the bridge parameters with the correct connection information to the SAP Business Object environment. As this Tutorial assumes one does not have the same environment handy, we will illustrate it here and then provide the .CSV file.

Click on the Import button.

When the import has finished and the log reports success, click on the Analyze tab.

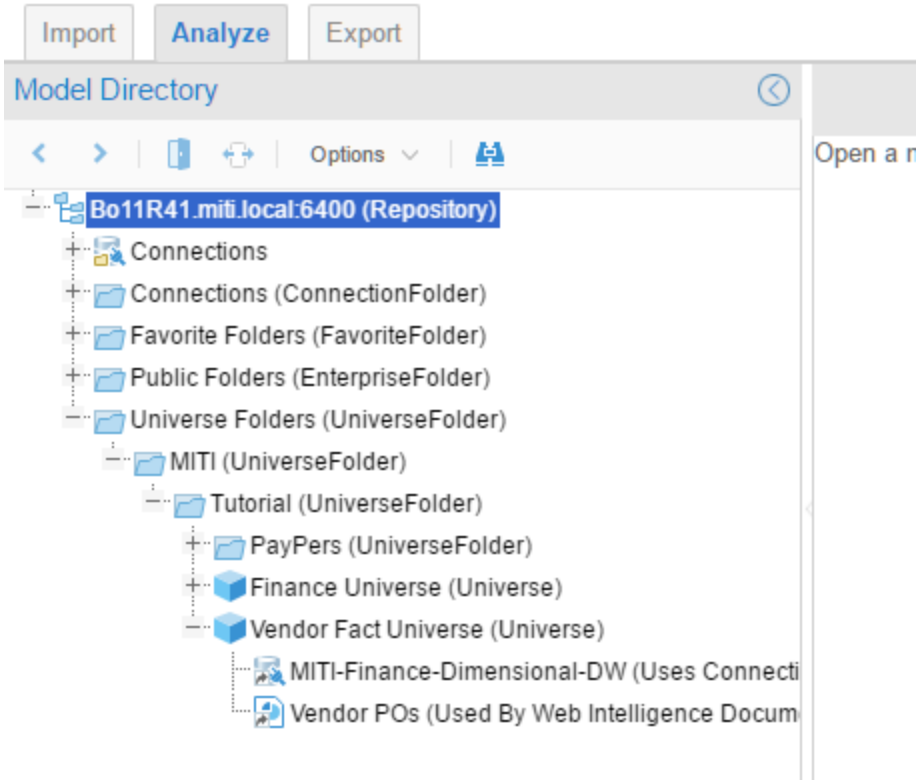


Figure 4 - Analyze tab

Note, two universes were imported, Finance Universe and Vendor Fact Universe. Double-click on the Vendor Fact Universe and expand some of the classes and features. Note, we have good definitions and logical names.

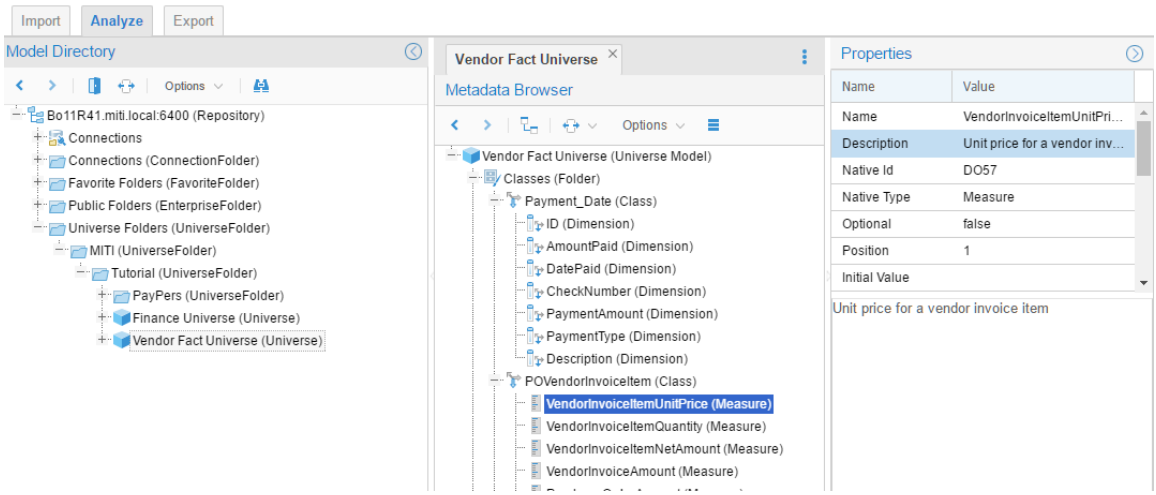


Figure 5 - Business names and descriptions

2.2.2 Export to a Glossary CSV File

We will now export to a glossary based .CSV format, which will contain both the terminology derived from the classes and features in the OLAP model as well as semantic mapping information back to the specific object in that OLAP model. The latter is so Meta Integration® Metadata Management (MIMM) may recreate the semantic mappings.

Click on the Export tab.

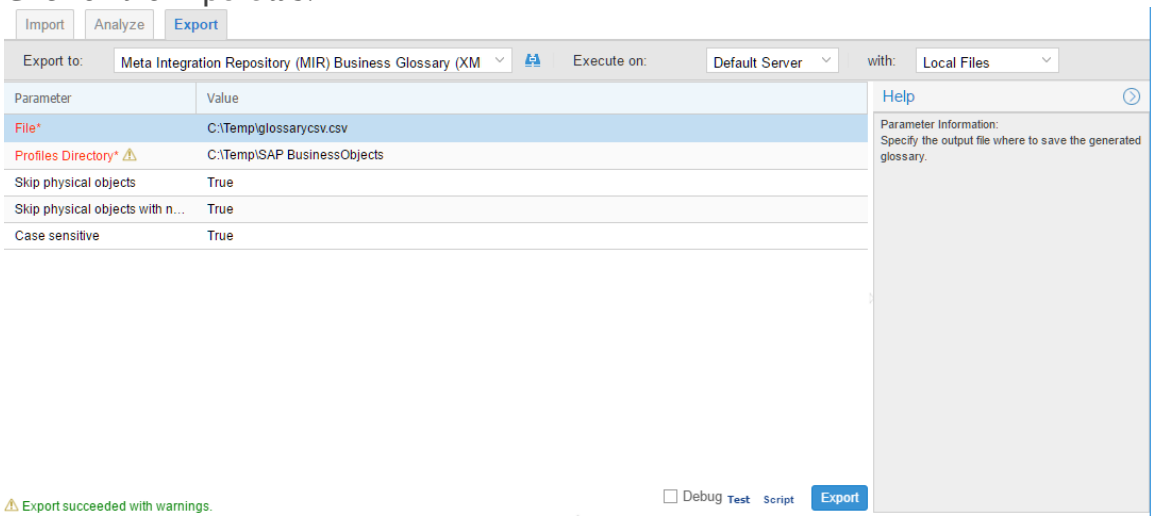


Figure 6 - Export tab

Then enter the appropriate bridge parameters, including the location to place the resulting CSV file.

Finally, click on the Export button.

When the export has finished and the log reports success, open the resulting .CSV file.

	A	B	C	D	E	F	G
1	Parent Category Path	Name	Description	Model	Object	Object Type	Id
2	Classes/Employee/	Birthdate	The person's birthdate	Universe Folders/MITI/Tutorial/Pay	Classes/Employee/Birthdate/	Dimension	i:B76D2B2598F88939BA37863
3	Classes/POVendorInvoiceI	BillingStreetAddress2	Second line of street addre	Universe Folders/MITI/Tutorial/Fin	Classes/POVendorInvoiceItem/BillingSt	Dimension	i:03E242FD9815EB91E58BAAE
4	Classes/POVendorInvoiceI	BillingStreetAddress2	Second line of street addre	Universe Folders/MITI/Tutorial/Ver	Classes/POVendorInvoiceItem/BillingSt	Dimension	i:40F1AD09E36BFD76BE5E279
5	Classes/POVendorInvoiceI	PurchaseOrderDescription		Universe Folders/MITI/Tutorial/Fin	Classes/POVendorInvoiceItem/Purchas	Dimension	i:03E242FD9815EB91E58BAAE
6	Classes/POVendorInvoiceI	PurchaseOrderDescription		Universe Folders/MITI/Tutorial/Ver	Classes/POVendorInvoiceItem/Purchas	Dimension	i:40F1AD09E36BFD76BE5E279
7	Classes/POVendorInvoiceI	POLineItemUnitDescriptio	Description of meaning of	Universe Folders/MITI/Tutorial/Fin	Classes/POVendorInvoiceItem/POLineIt	Dimension	i:03E242FD9815EB91E58BAAE
8	Classes/POVendorInvoiceI	POLineItemUnitDescriptio	Description of meaning of	Universe Folders/MITI/Tutorial/Ver	Classes/POVendorInvoiceItem/POLineIt	Dimension	i:40F1AD09E36BFD76BE5E279
9	Classes/EvaluationSupport	DocumentURL	UReference to the fully for	Universe Folders/MITI/Tutorial/Pay	Classes/EvaluationSupportingDocument	Dimension	i:B76D2B2598F88939BA37863
10	Classes/HistorySupporting	DocumentDescription	Document description	Universe Folders/MITI/Tutorial/Pay	Classes/HistorySupportingDocument/Dc	Dimension	i:B76D2B2598F88939BA37863
11	Classes/CustomerPOInvoic	ShippingPOCName	Name of Point of Contact	Universe Folders/MITI/Tutorial/Fin	Classes/CustomerPOInvoiceItem/Shippi	Dimension	i:03E242FD9815EB91E58BAAE
12	Classes/Payment/	Note	Special notes about this ad	Universe Folders/MITI/Tutorial/Pay	Classes/Payment/Note/	Dimension	i:B76D2B2598F88939BA37863
13	Classes/CustomerPOInvoic	CustomerPOLineItemStatu	Current status of this line i	Universe Folders/MITI/Tutorial/Fin	Classes/CustomerPOInvoiceItem/Custo	Dimension	i:03E242FD9815EB91E58BAAE
14	Classes/CustomerPaymen	CustomerPaymentType	Type of payment	Universe Folders/MITI/Tutorial/Fin	Classes/CustomerPaymentDate/Custom	Dimension	i:03E242FD9815EB91E58BAAE
15	Classes/CustomerPOInvoic	ShippingInternational	International shipping info	Universe Folders/MITI/Tutorial/Fin	Classes/CustomerPOInvoiceItem/Shippi	Dimension	i:03E242FD9815EB91E58BAAE
16	Classes/Vendor/	VendorID		Universe Folders/MITI/Tutorial/Fin	Classes/Vendor/VendorID/	Dimension	i:03E242FD9815EB91E58BAAE
17	Classes/Vendor/	VendorID		Universe Folders/MITI/Tutorial/Ver	Classes/Vendor/VendorID/	Dimension	i:40F1AD09E36BFD76BE5E279
18	Classes/Assignment/	Details	Details about the history it	Universe Folders/MITI/Tutorial/Pay	Classes/Assignment/Details/	Dimension	i:B76D2B2598F88939BA37863
19	Classes/Payment/	POPayDate	Date of first payment as sc	Universe Folders/MITI/Tutorial/Pay	Classes/Payment/POPayDate/	Dimension	i:B76D2B2598F88939BA37863
20	Classes/POVendorInvoiceI	PaymentAddressDescriptic	Description of location spe	Universe Folders/MITI/Tutorial/Ver	Classes/POVendorInvoiceItem/Paymen	Dimension	i:40F1AD09E36BFD76BE5E279
21	Classes/CustomerPOInvoic	InvoiceLineItemUnitDescri	Short description of what a	Universe Folders/MITI/Tutorial/Fin	Classes/CustomerPOInvoiceItem/Invoic	Dimension	i:03E242FD9815EB91E58BAAE
22	Classes/POVendorInvoiceI	ShippingState	State	Universe Folders/MITI/Tutorial/Fin	Classes/POVendorInvoiceItem/Shipping	Dimension	i:03E242FD9815EB91E58BAAE
23	Classes/POVendorInvoiceI	ShippingState	State	Universe Folders/MITI/Tutorial/Ver	Classes/POVendorInvoiceItem/Shipping	Dimension	i:40F1AD09E36BFD76BE5E279
24	Classes/POVendorInvoiceI	POLineItemQuantity	Number of units for a line	Universe Folders/MITI/Tutorial/Fin	Classes/POVendorInvoiceItem/POLineIt	Measure	i:03E242FD9815EB91E58BAAE

Figure 7 - CSV file

This is the result required in order to reverse engineer into Meta Integration® Metadata Management (MIMM), creating a glossary and semantic mappings required.

2.3 Populate the Glossary (from CSV)

As mentioned earlier, we will not only use the information contained within the BI design environment, but also the lineage constructed by Meta Integration® Metadata Management (MIMM) to populated and semantically map the ultimate Physical Data Model representation of the data warehouse from the glossary. Thus, we need a current configuration with the warehouse and the BI environment stitched.

We have already built up a subset of the architecture that includes those two as well as a staging warehouse (Staging DW) and the data integration in between that and the Dimensional DW.

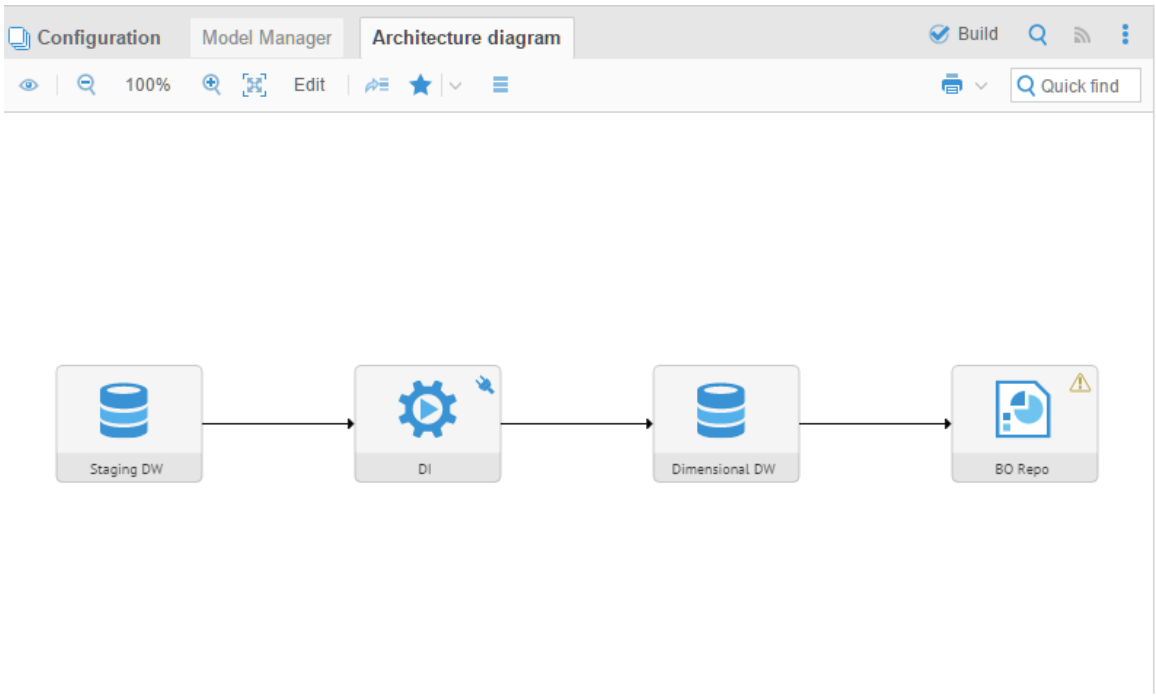
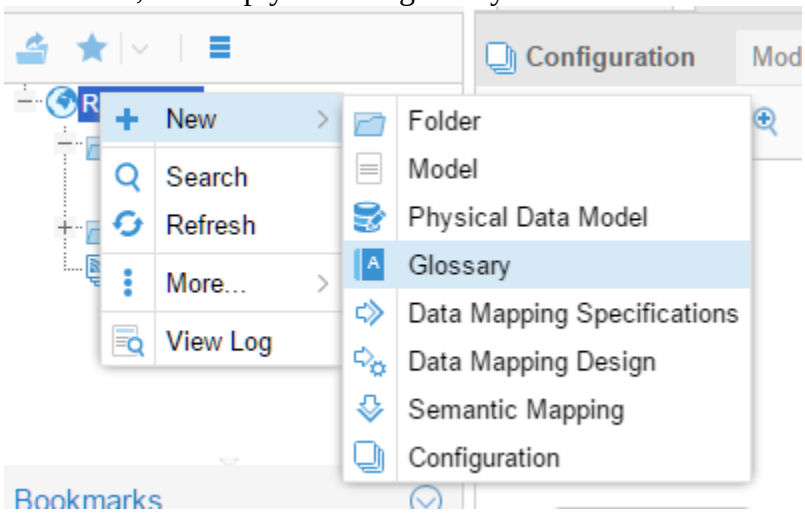


Figure 8 - Architecture subset

From here, we simply create a glossary:



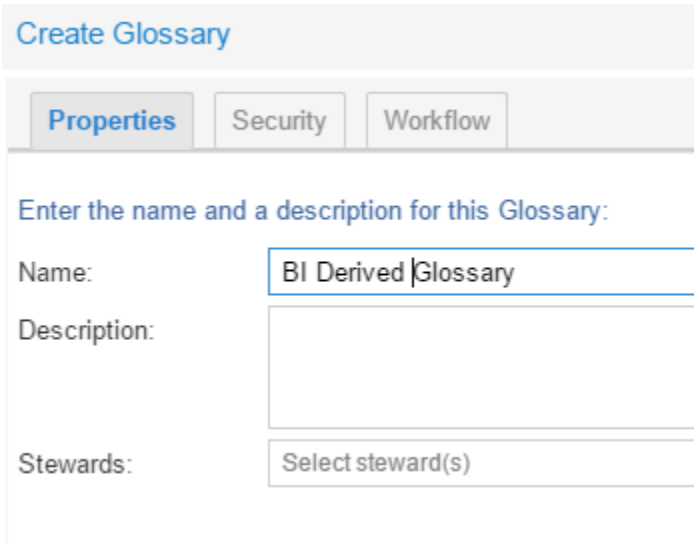


Figure 9 - Create glossary

Then, include it in the configuration:

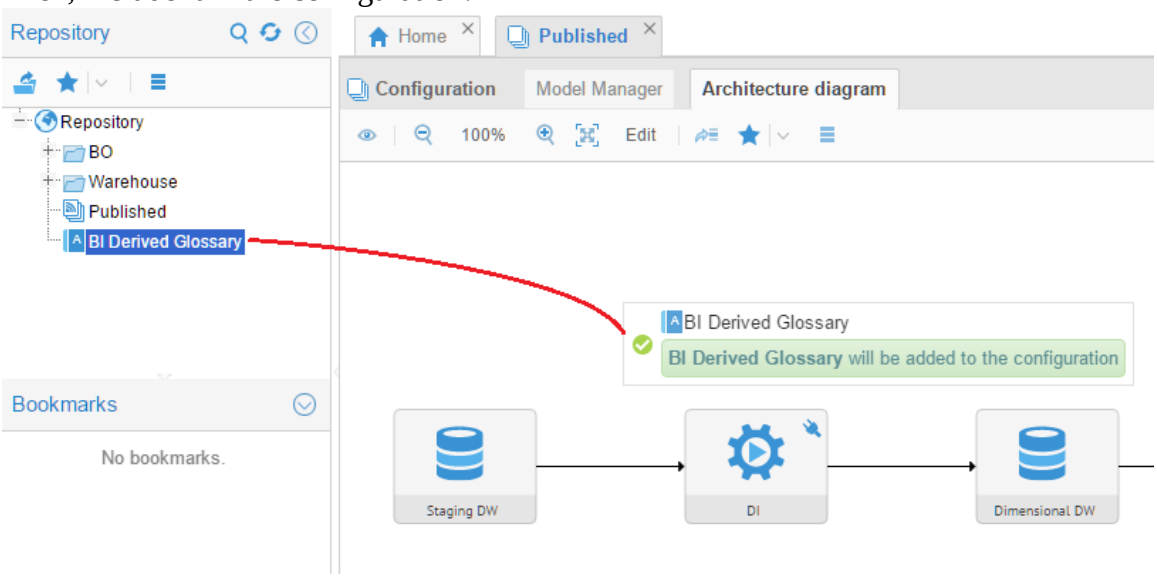


Figure 10 - Drag the glossary into the configuration

Build the configuration and open it in the Explorer UI.

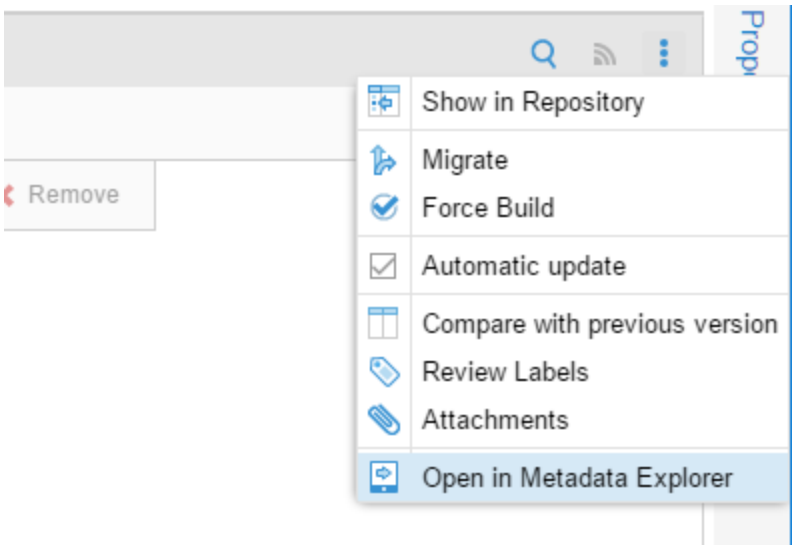


Figure 11 - Explorer UI

Browse to the glossary:

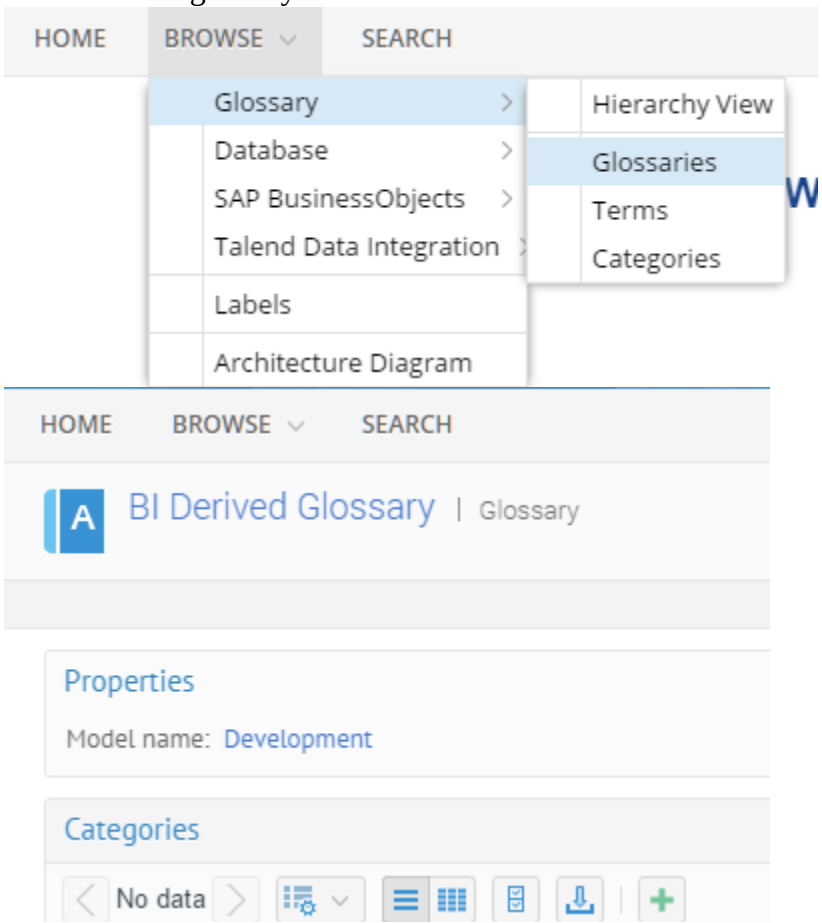


Figure 12 - Browse to the glossary

And import the .CSV file created in MIMB.

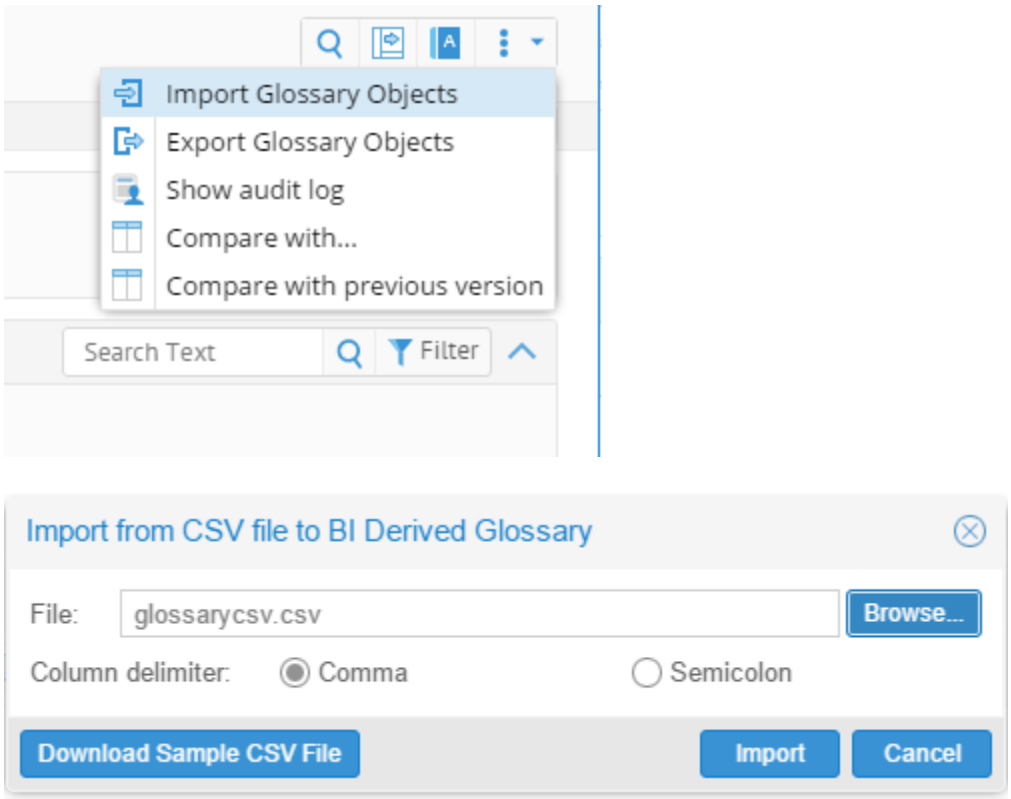


Figure 13 - Import the .CSV file

Once the log reports success you have populated the glossary.

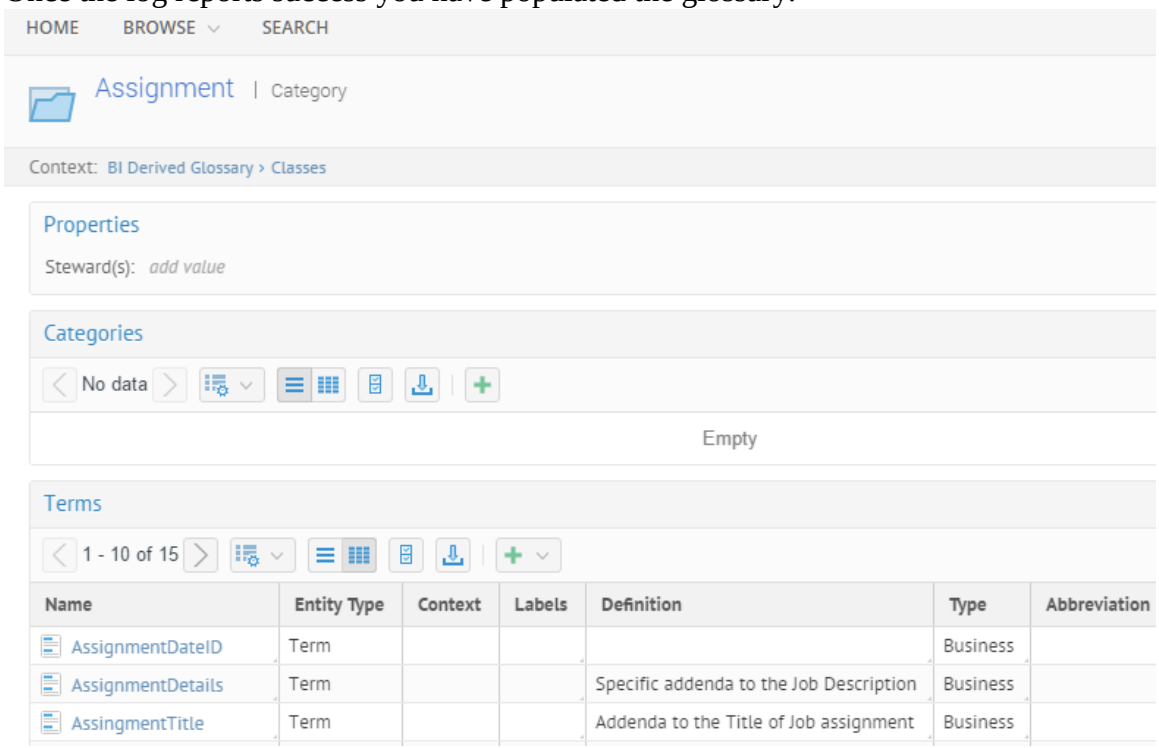


Figure 14 - Resultant glossary

2.4 Populate the Semantic Mapping (from CSV)

Now, as pointed out earlier, the .CSV file also contains information to reconstruct the semantic mappings from the glossary to the BI environment.

So, create a new semantic mapping:

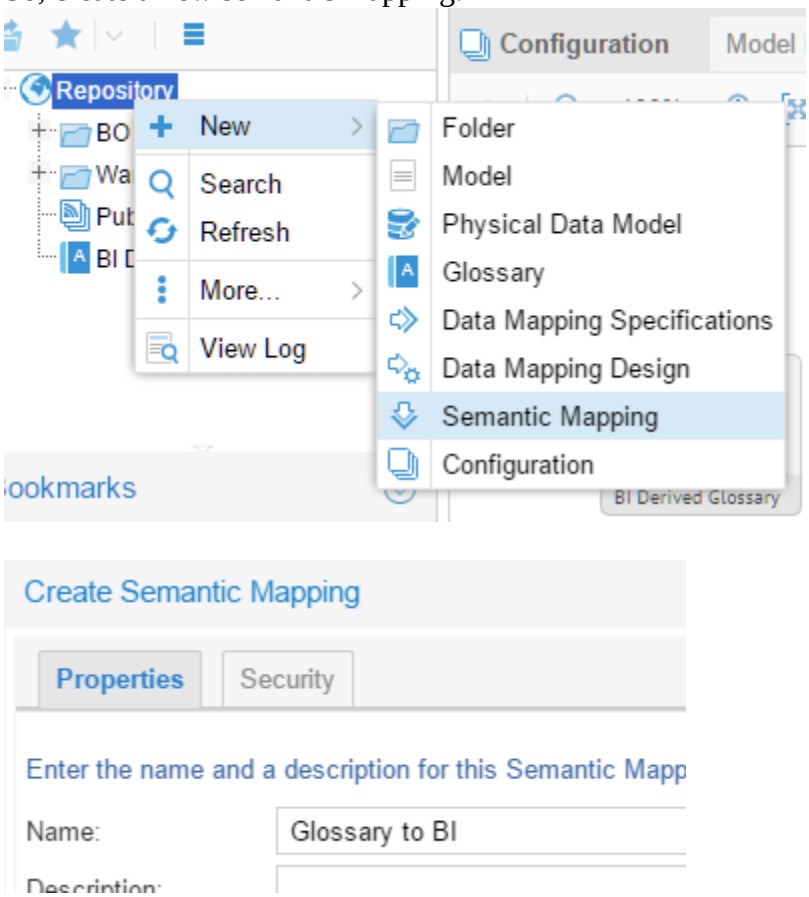


Figure 15 - Create new semantic mapping

Drag the BI Derived Glossary into the left of the mapping:

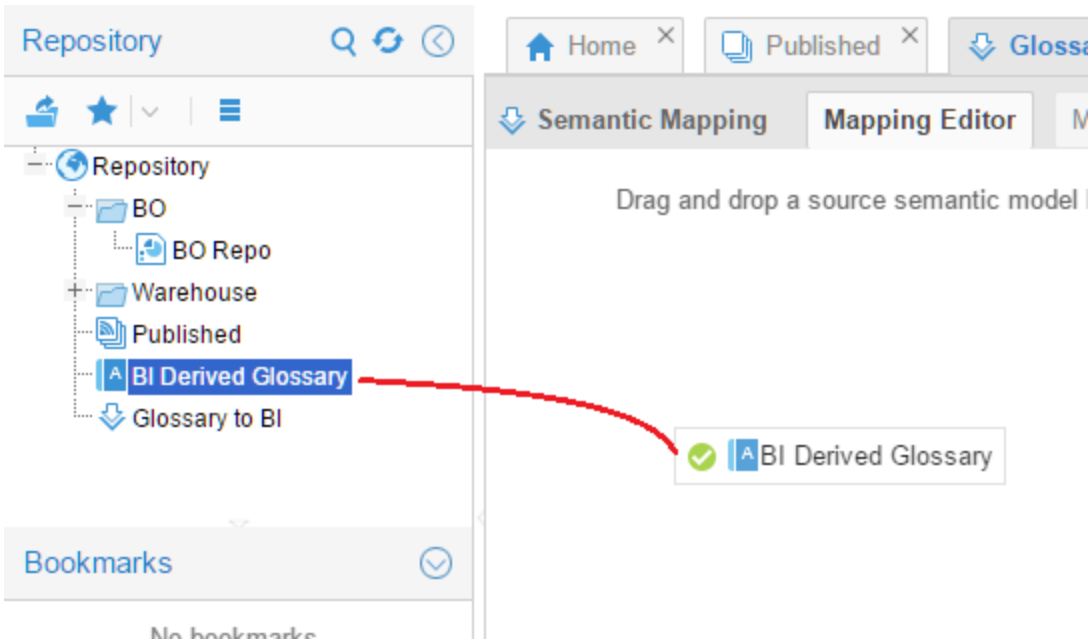


Figure 16 - Drag and drop

Drag the BO Repo model into the left of the mapping:

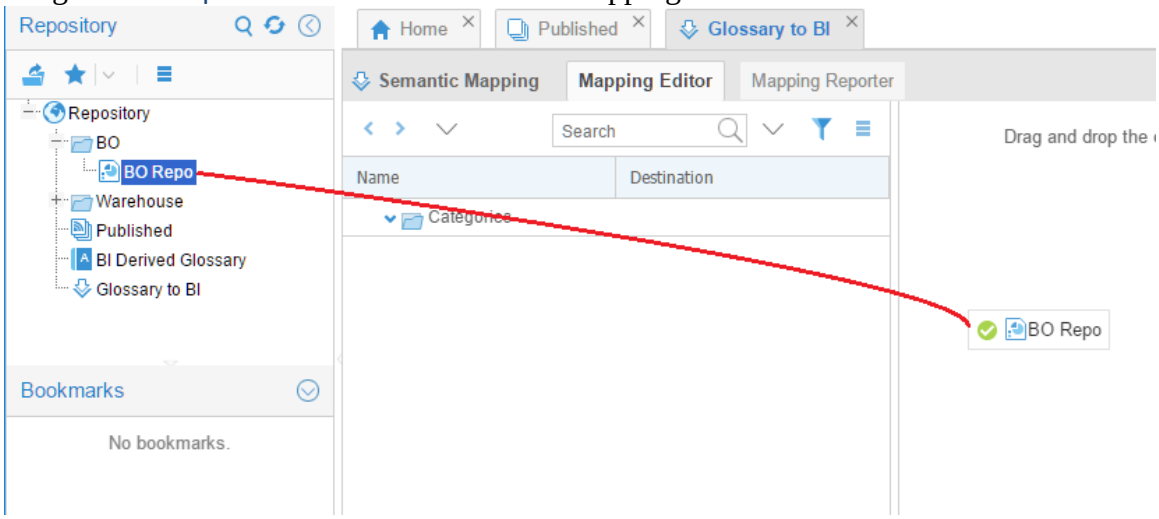
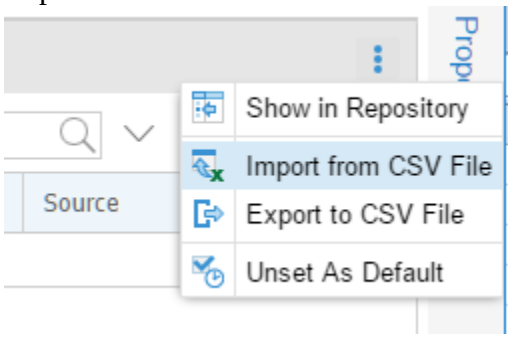


Figure 17 - Drag and drop

Import from the same .CSV file:



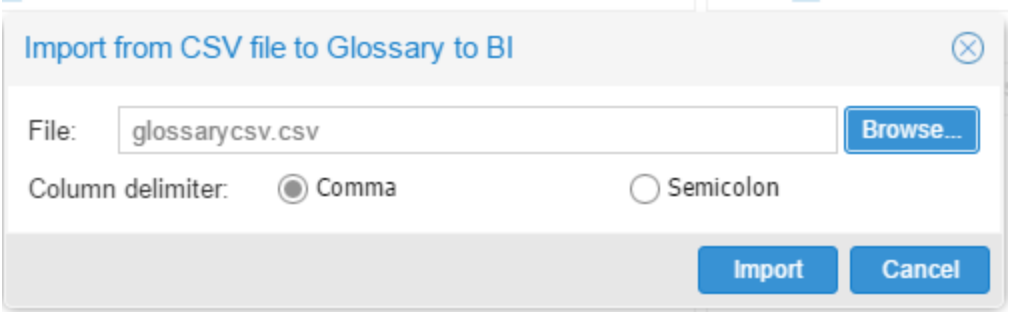


Figure 18 - Import semantic mappings from CSV

And the resulting mappings are there:

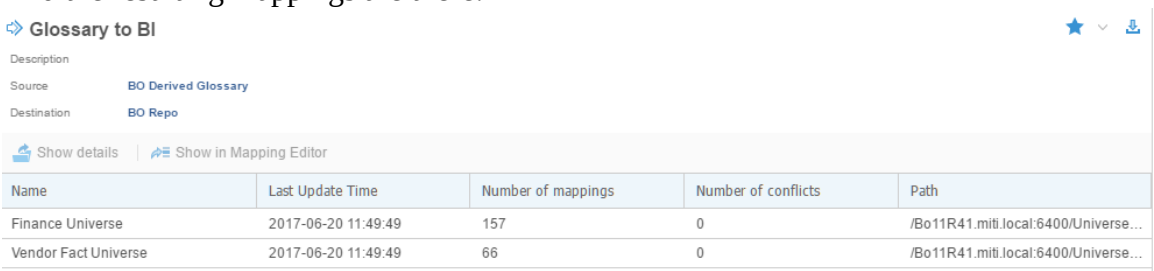


Figure 19 - Resulting semantic mappings

Close the semantic mapping and add it to the configuration:

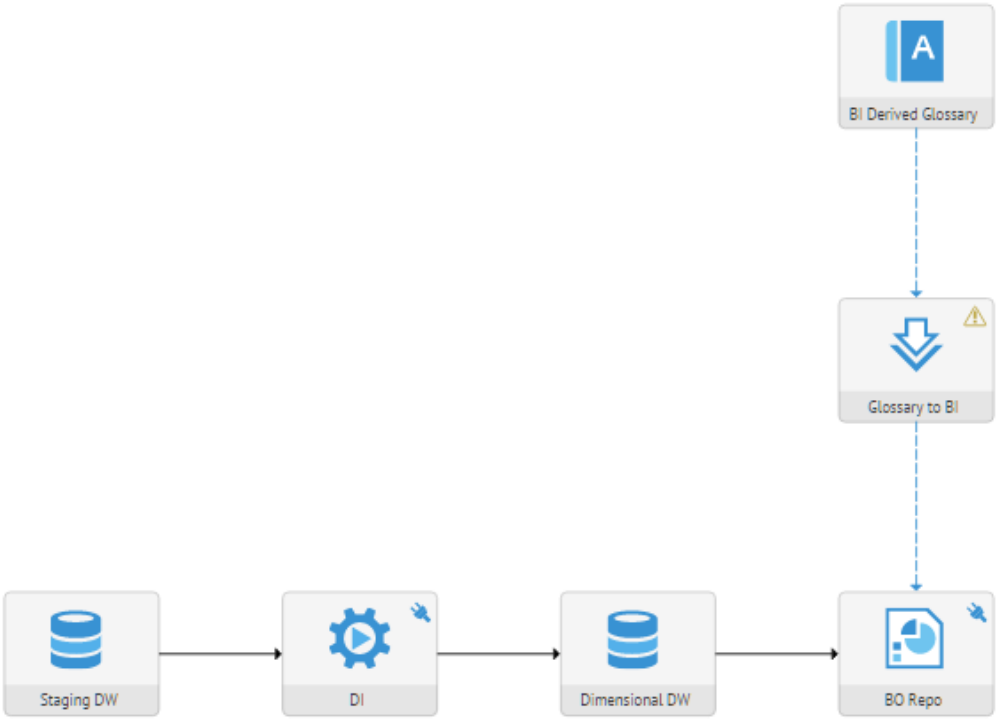


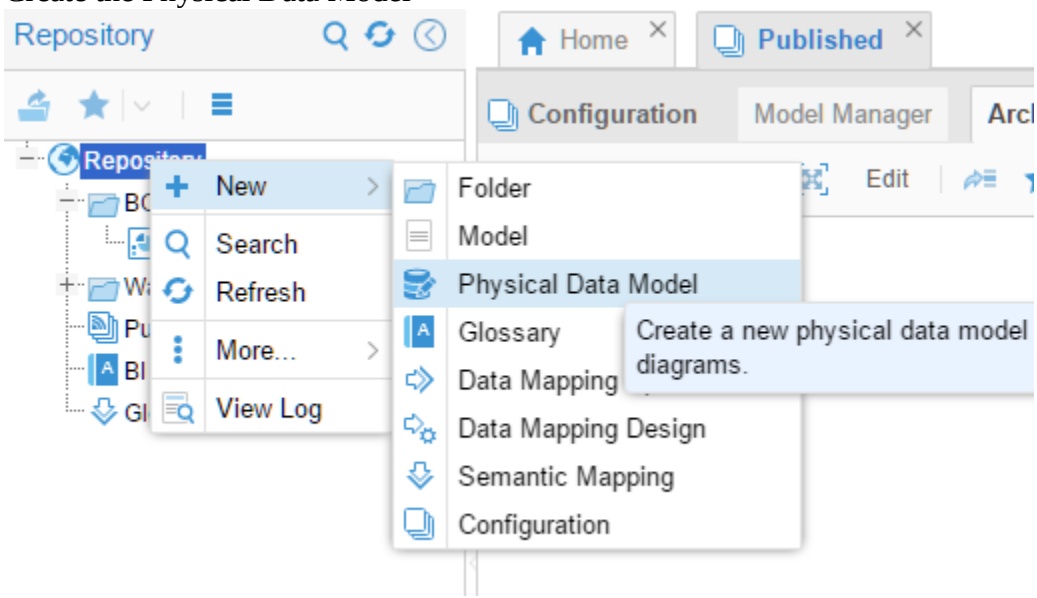
Figure 20 - Semantic mapping in the configuration

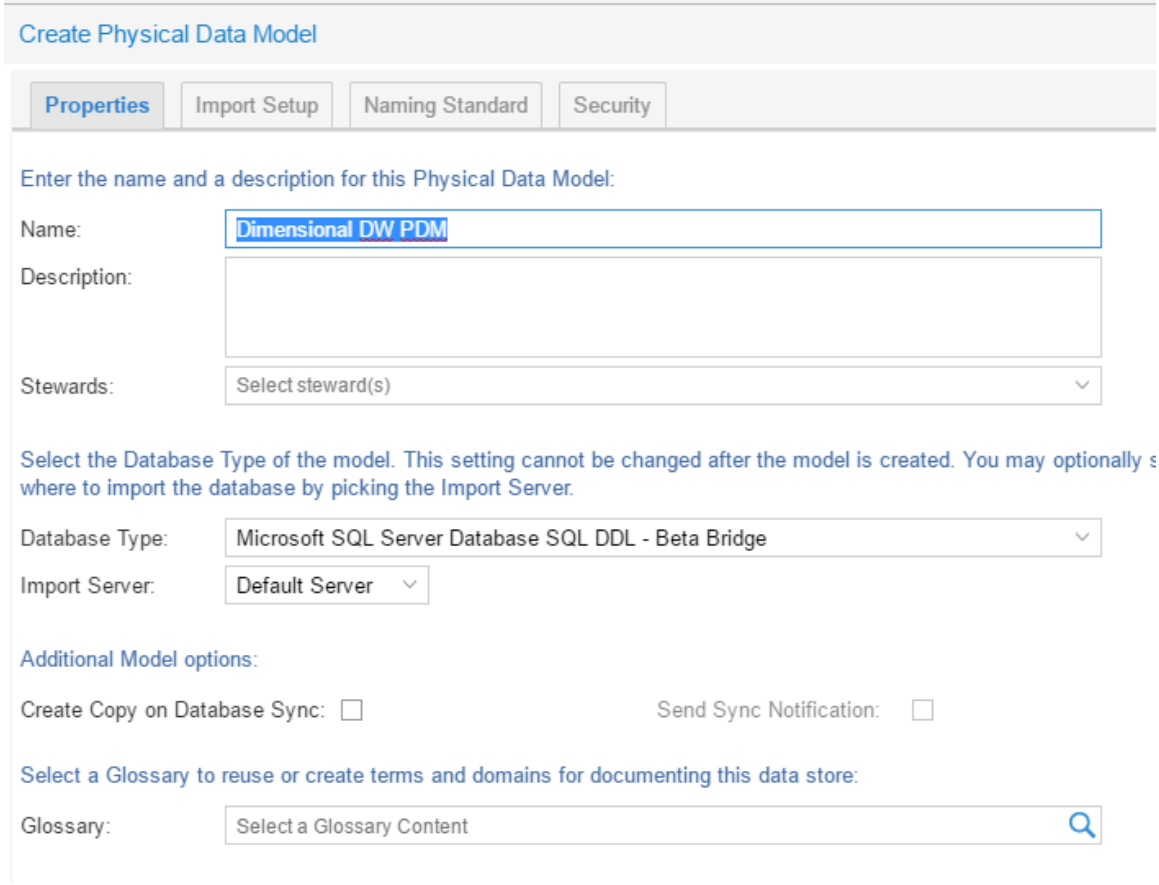
3 BI Design Documentation Migration to the Data Warehouse

3.1 Convert the Data Warehouse into a Physical Data Model

Now, we will create the [phsycialDataModel] for the Dimensional DW and include that in the configuration, if it is not already done.

Create the Physical Data Model





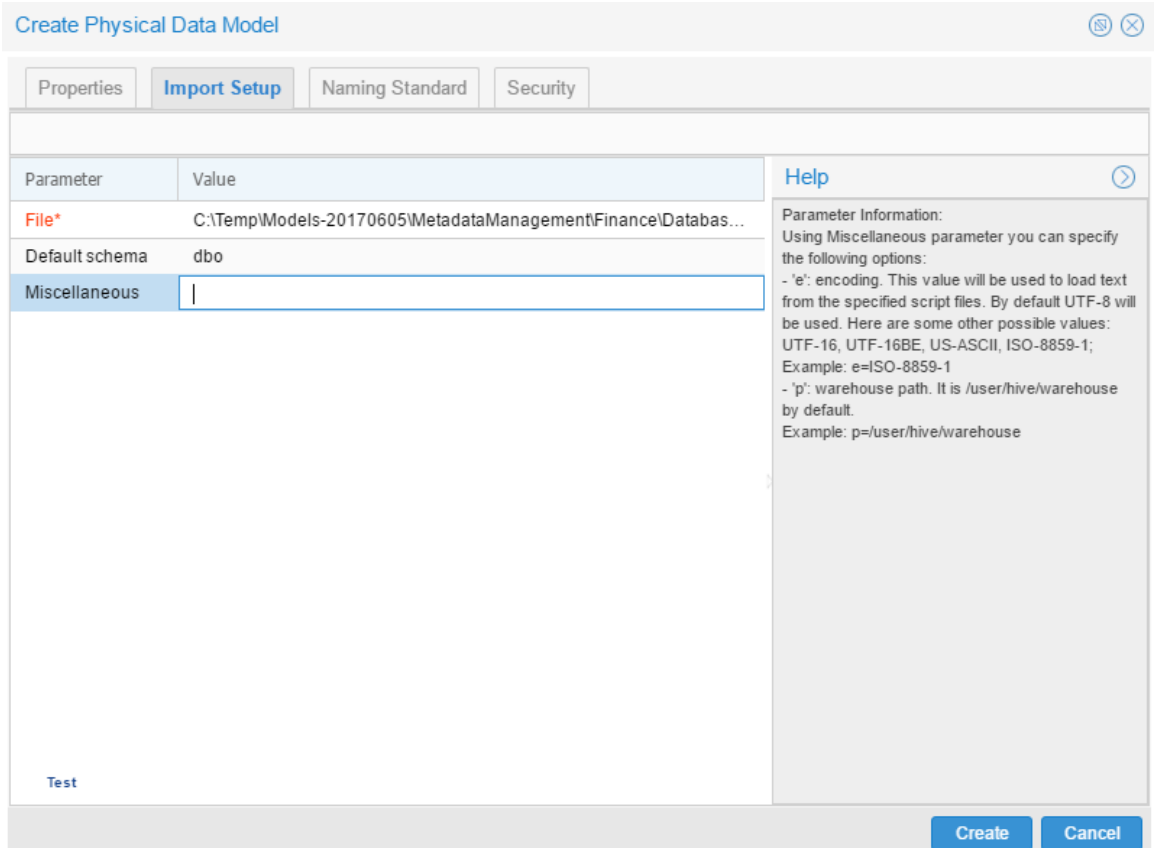


Figure 21 - Create the Physical Data Model

Click on the Create button.

Then click on the Yes button to import the underlying physical schema and sync with it.

Now, remove the existing warehouse model in the configuration.

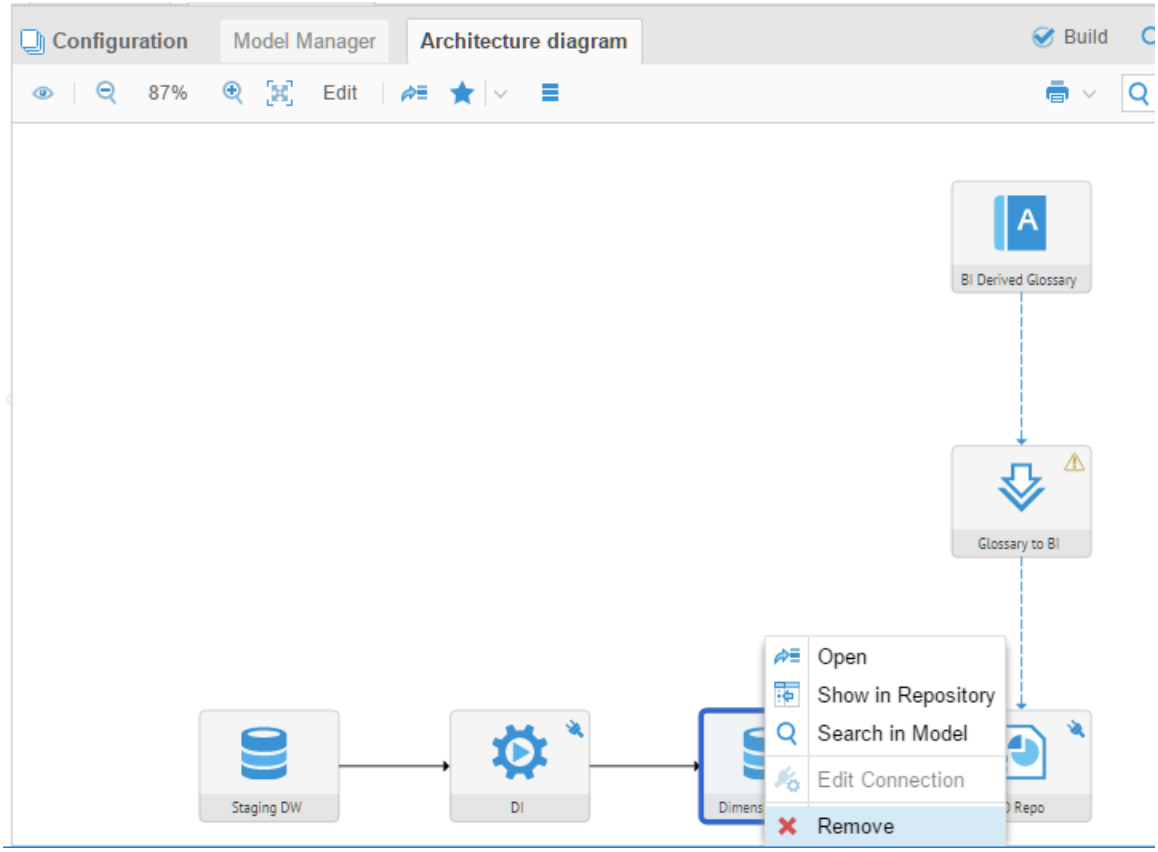


Figure 22 - Remove

And replace it with the new Physical Data Model:

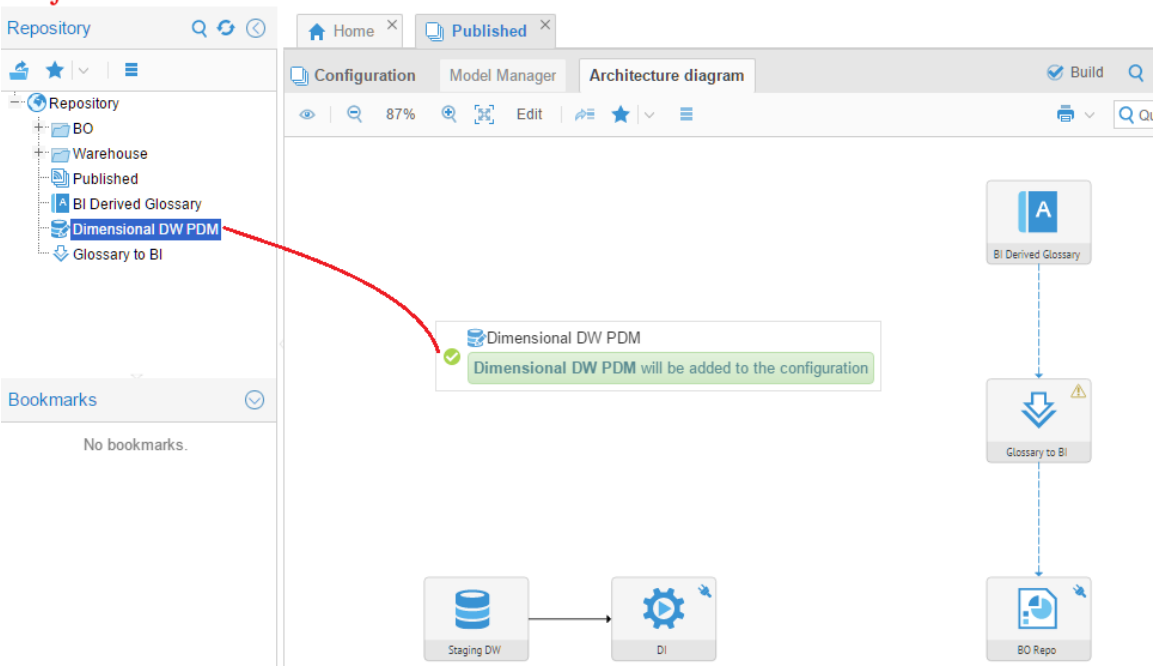
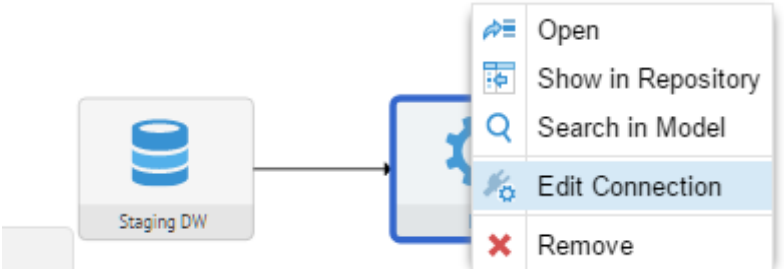
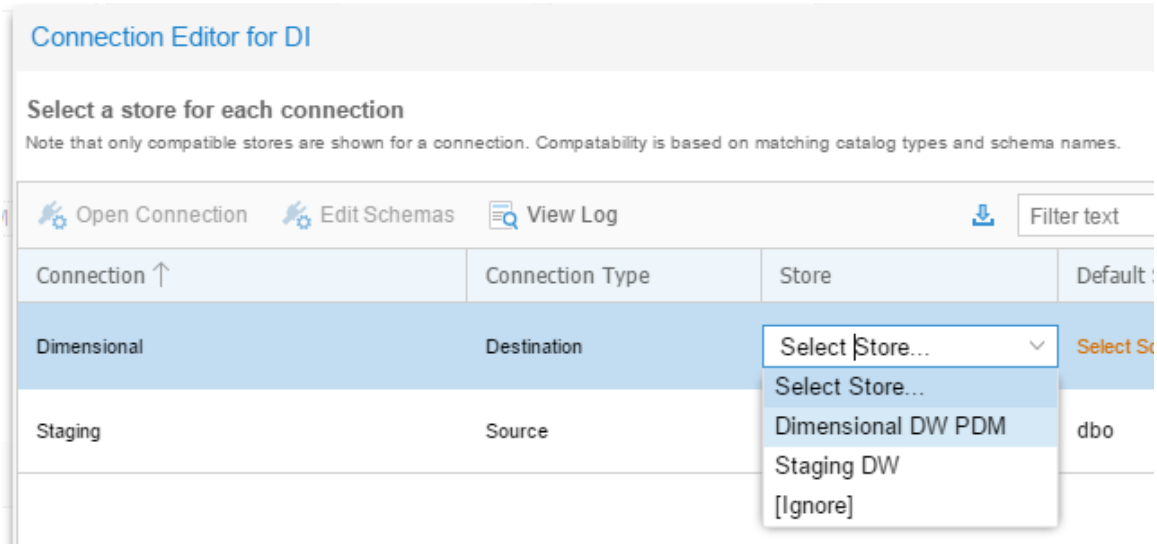
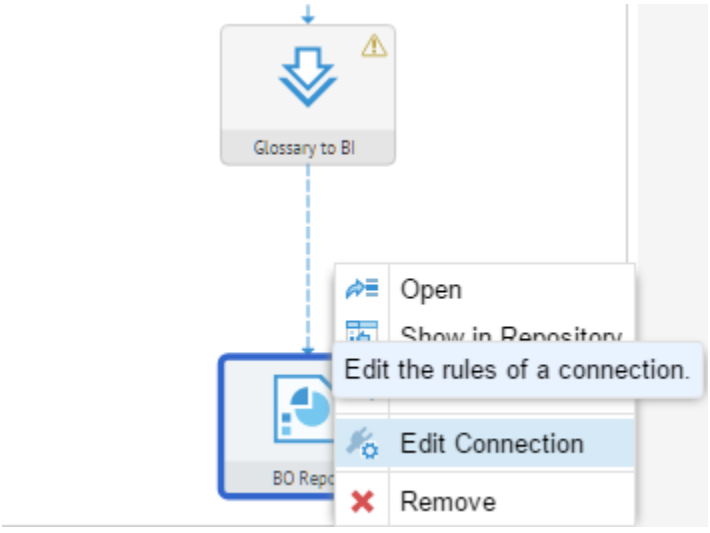


Figure 23 - Replace

And stitch it into the architecture:







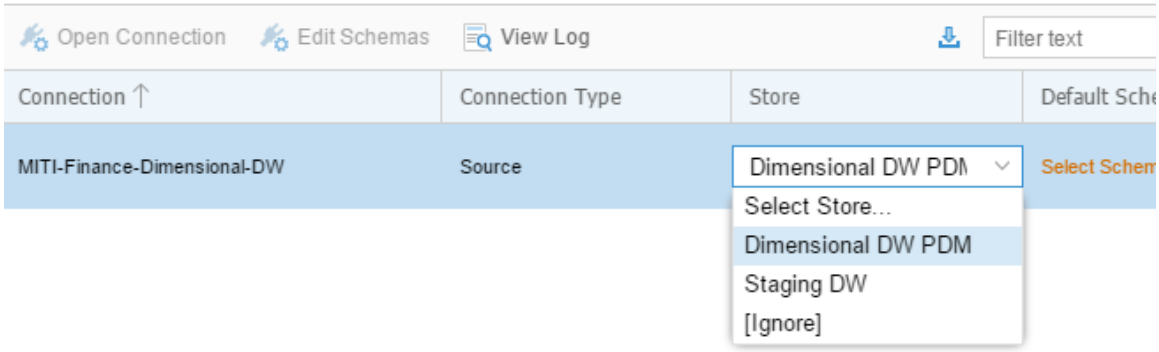


Figure 24 - Stitch

Layout the diagram and the result is:

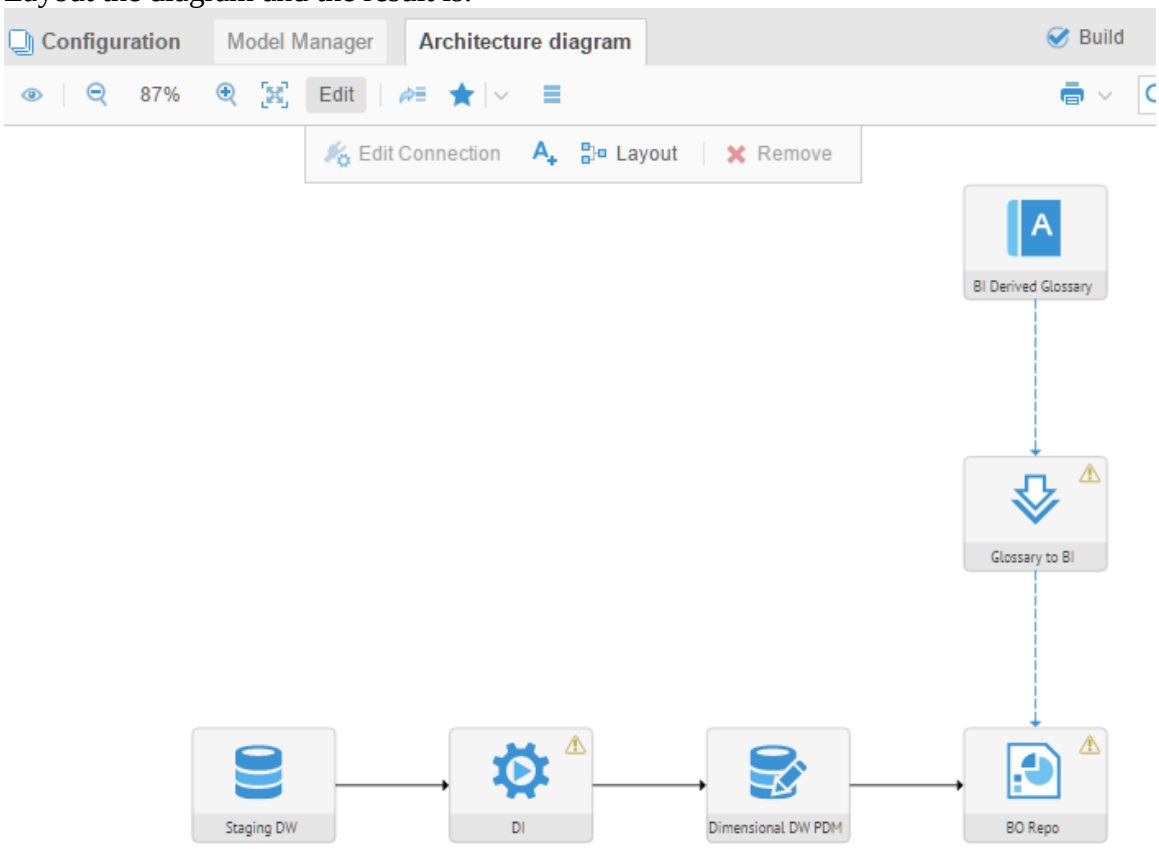


Figure 25 - Result

3.2 Migrate the Semantic Mapping to the Physical Data Model

It is now time to migrate the semantic mappings (from the glossary to the BI environment) into a semantic mapping from the glossary to the Dimensional DW.

First, create the semantic mapping:

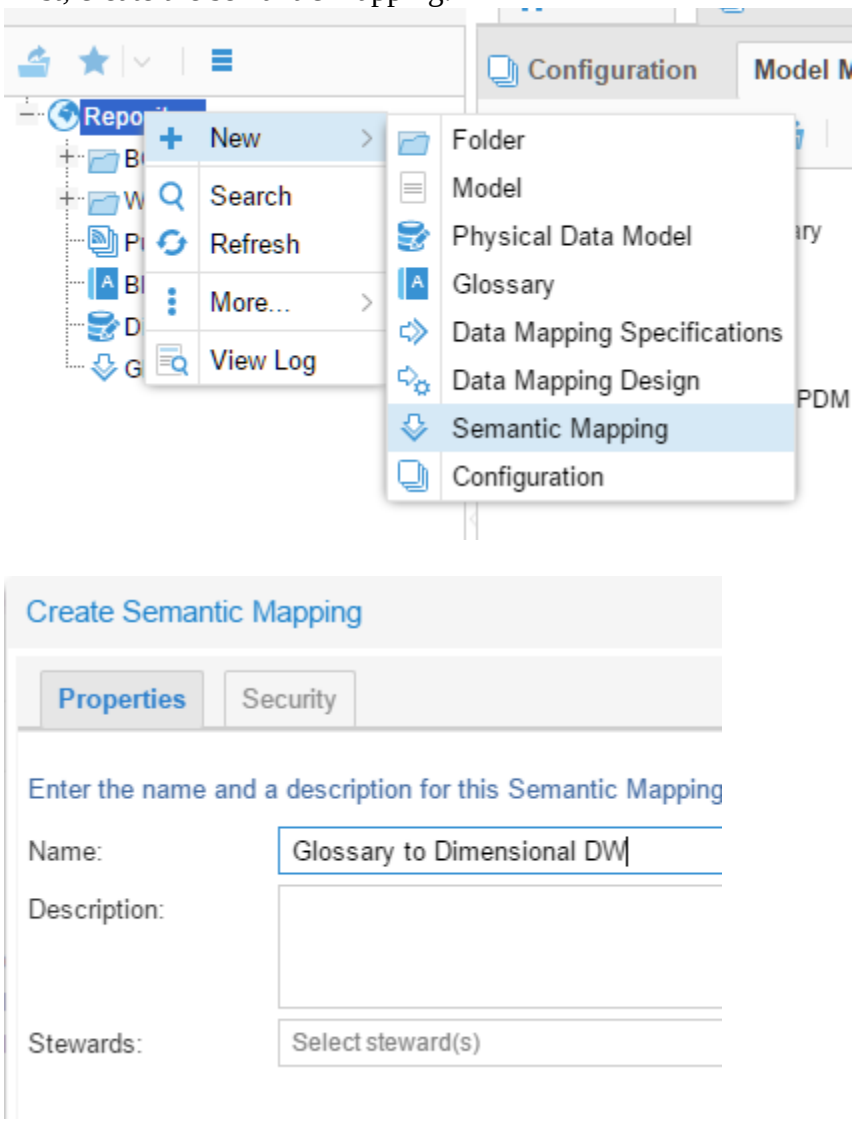


Figure 26 - New semantic mapping

Drag the BI Derived Glossary into the left of the mapping:

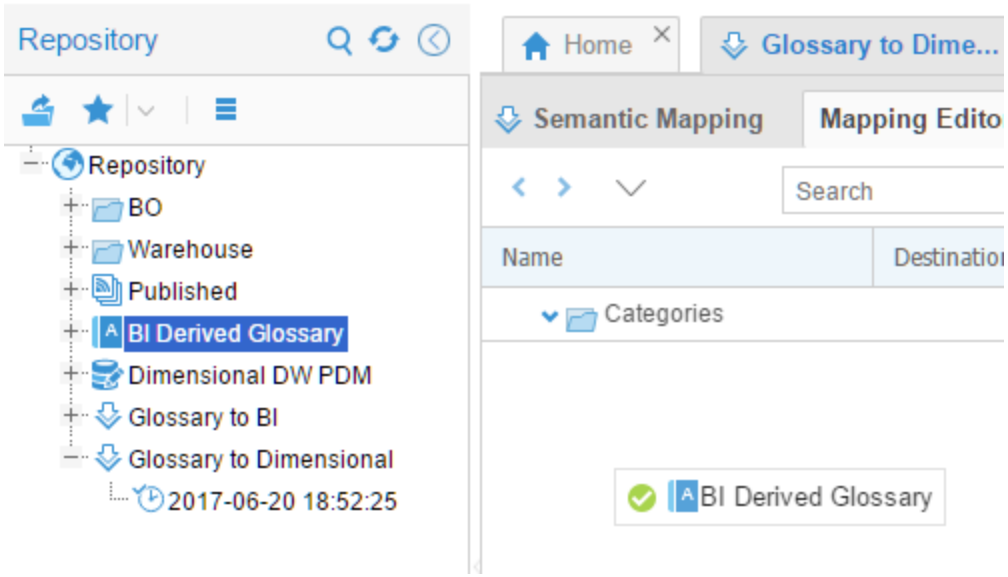


Figure 27 - Drag and drop

Drag the Dimensional DW PDM model into the left of the mapping:

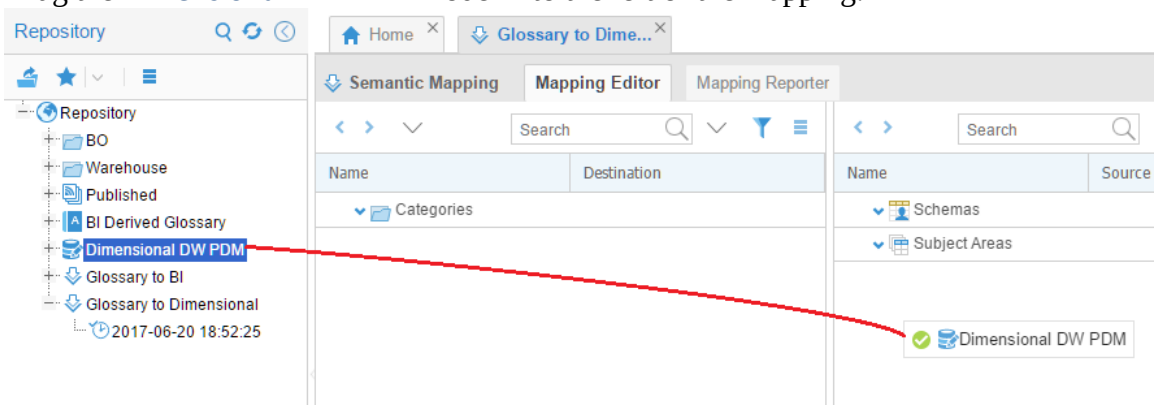


Figure 28 - Drag and drop

Close the semantic mapping and add it to the configuration:

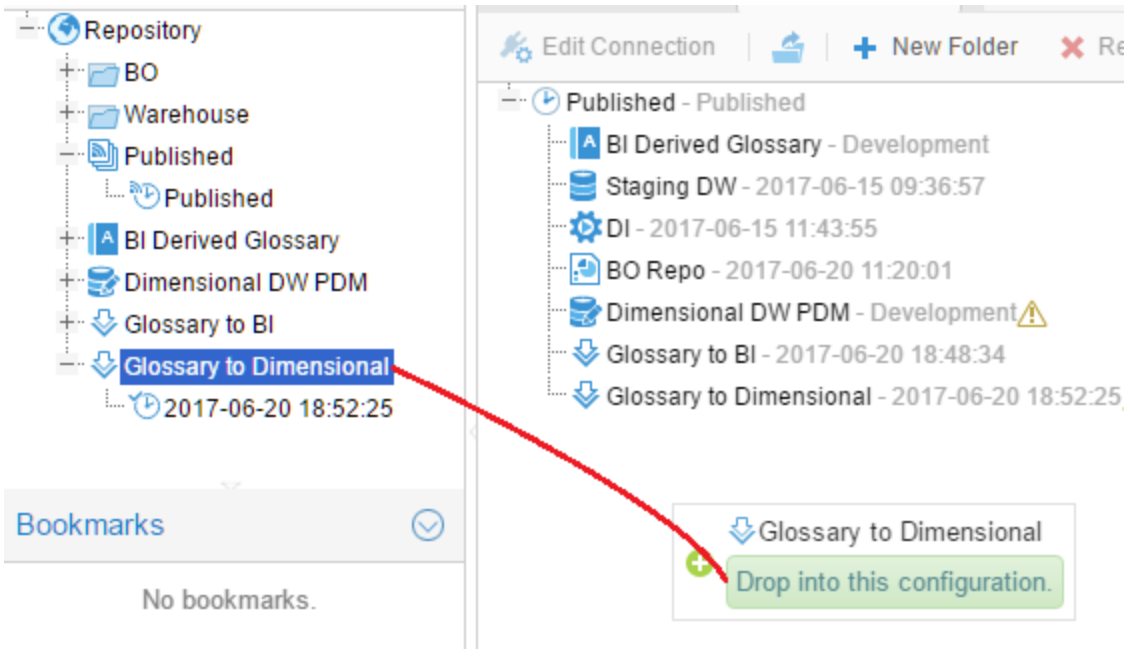


Figure 29 - Drag and drop

Now, right-click on the mapping and select Auto Map:

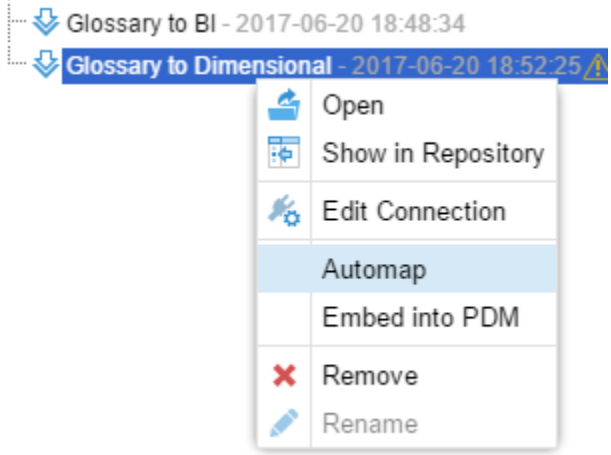


Figure 30 - Auto Map

And we have a semantic mapping:

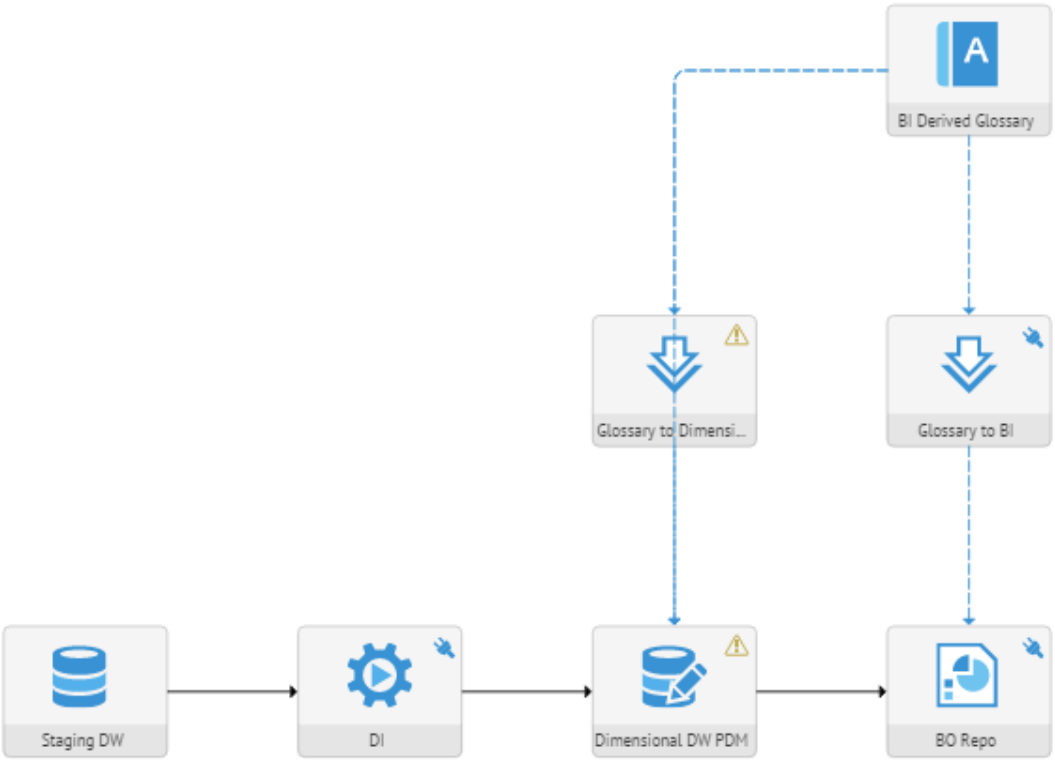


Figure 31 - Semantic mapping

3.3 Embed the Semantic Mapping into the Physical Data Model

Now, in order to use the glossary as the associated one for the Dimensional DW PDM model, one may embed the mapping, removing it as a standalone mapping and embedding it in the documentation process for the Dimensional DW PDM.

Go to the Model Manager tab, right-click on the mapping that you just auto mapped and select Embed into PDM:

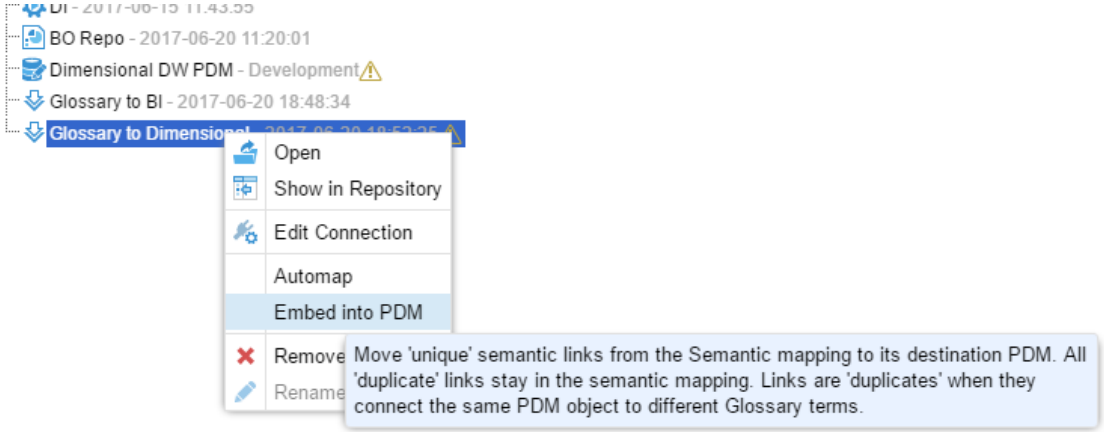


Figure 32 - Embed the mapping

Now, to clean up, remove the two semantic mappings:

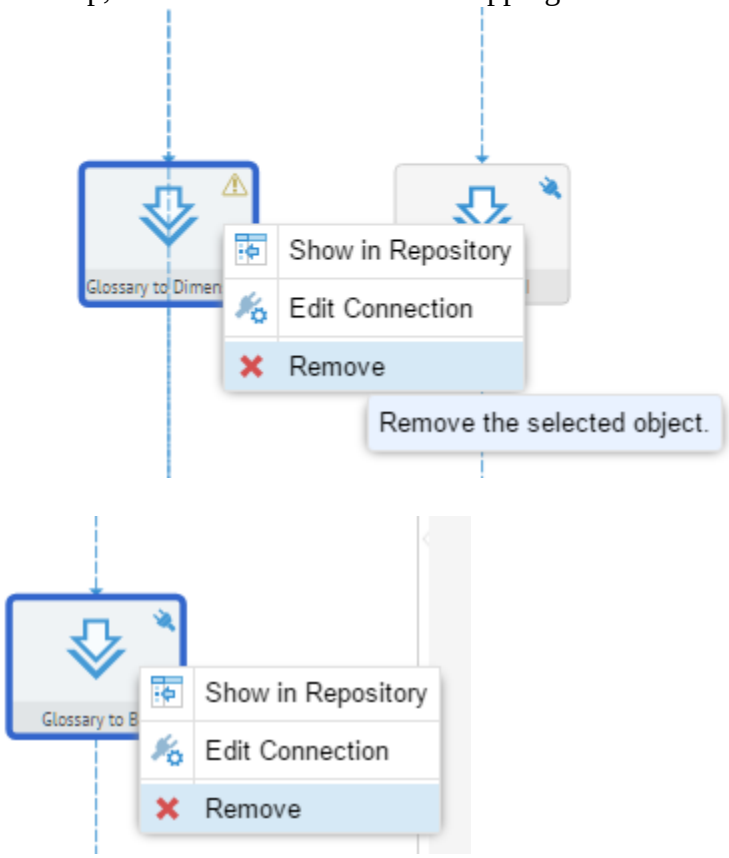


Figure 33 - Clean up the configuration.

Build the configuration.

Now, one may continue to document the Dimensional DW PDM.

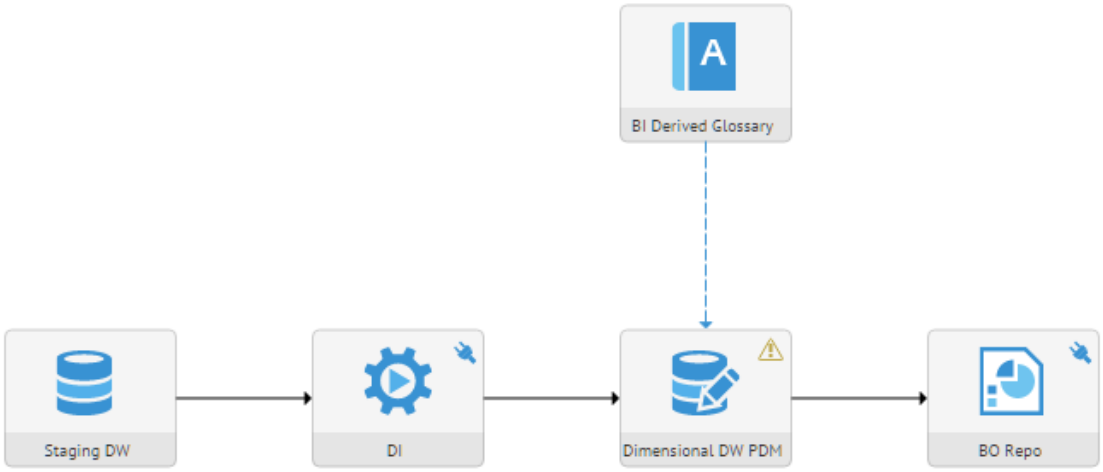


Figure 34 - New configuration

4 BI Report Documentation Reuse as Glossary

TBD