# Metadata Management Tutorial

Data Modeling
Using erwin Metadata Management
(EMM)

#### **TABLE OF CONTENTS**

1	Intro	oduction	4
	1.1	How to use this document	5
	1.2	Conventions used in the tutorial	6
2	Phys	sical Data Modeler	Error! Bookmark not defined.
	2.1	Create a physical data model	8
	2.2	Bootstrapping from existing documentation	10
	2.3	Business Glossary Integrated Documentation	15
	2.3.1	Create the physical data model	
		Linked term generation and classification	
		Name/Abbreviation standards	
	2.3.4		22
	2.4	Subject Areas and Diagrams	29
	2.4.1	Creating a Diagram	29
		Editing a Diagram	30

#### **TABLE OF FIGURES**

Figure 1 -	Settings for the Staging DW model	8
Figure 2 -	New model from erwin DM	11
Figure 3 -	Dimensional DW model from erwin DM	12
Figure 4 -	Drag documented model into physical data model to bootstrap	13
Figure 5 -	Fully documented model	
Figure 6 -	Sync with Database	14
Figure 7 -	Reference model specifications	15
Figure 8 -	Naming standards settings for Web Contact model	17
Figure 9 -	Dragging the Web Contact model into the Finance configuration	19
Figure 10 -	Open in Metadata Explorer	
Figure 11 -	Entering the name Activity	20
Figure 12 -	Entering the description	20
Figure 13 -	Glossary integration with physical data model	21
Figure 14 -	Link to the new term	
Figure 15 -	Get standardized name	23
Figure 16 -	Open in Metadata Manager	23
Figure 17 -	Apply naming standard in Manager Ui	24
Figure 18 -	Apply Naming Standard in Explorer UI	
Figure 19 -	Results of apply naming standard	25
Figure 20 -	Display grid	25
Figure 21 -	Changing table name in a physical data model with a naming standard defined	26
Figure 22 -	Changing column name in a physical data model with a naming standard defined	27
Figure 23 -	Renamed term in naming standard glossary	27
Figure 24 -	Naming standards updates applied	28
Figure 25 -	Create new diagram	29
Figure 26 -	Tables added to Contact Diagram	30
Figure 27 -	Display preferences dialog	31
Figure 28 -	Relationship from Identity to User	31
Figure 29 -	Edit Relationship dialog	32
Figure 30 -	Relationship join condition	32
Figure 31 -	New shape	
Figure 32 -	Edited annotation	33

#### 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 is distributed and duplicated, often times active, and generally represented in a variety of methodologies, depending upon the underlying technology they represent.

erwin Metadata Management (EMM), provide strikingly expanded set of capabilities in many facets of metadata management, including:

- Business Glossary
- Data Governance
- Metadata comparison, integration, and mapping
- Version and configuration management
- Data life cycle related metadata management
- Lineage and impact analysis
- Enterprise architecture development, management and deployment.

This document is the culmination of many years of experience in supporting the enterprise metadata management and integration requirements of numerous clients. It presents in detail and with supporting tutorials this vision, metadata management process and methods, best practices, as well as, many strategic scenarios that leverage erwin Metadata Management (EMM). These examples are comprehensive and directed at realworld examples tied to business-oriented goals and return on investment. In all, anyone who completes the exercises in this document should find it straight-forward to implement and deploy an effective and comprehensive erwin Metadata Management (EMM) based metadata management environment.

#### Disclaimer

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

#### 1.1 How to use this document

It is certainly possible to skip through the tutorials, and thus simply glean a "management-level understanding" of erwin Metadata Management (EMM) 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 the erwin Metadata Management (EMM) environment itself is capable of working with over 100 different versions of third-party, 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 erwin Metadata Management (EMM) 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 Physical Data Modeler

This is a continuation of the exercises in the Metadata Management part of the Metadata Management Tutorial. Thus, you should first complete that tutorial before proceeding. If you wish to start directly from this point, without completing the first tutorial, delete the Tutorials folder in the repository and execute the script entitled Completed to produce the complete tutorial to that point.

A physical data model may be based upon a harvested database. In this case, the physical data model starts out with the schemas, tables and columns as harvested into it and then may be extended to include relationships, diagrams, subject areas, logical attribution (business name and description, e.g.) of a complete data model. This physical data model may then have a life of its own, or it may be *synchronized* with changes (due to harvesting new versions) to the underlying harvested model.

Thus, one may point to a data store (like a database, big data cluster via Hive, etc.) and then produce a full data model based upon that underlying data structure. One may then set that model to be synchronized with any new versions of that underlying data structure as they are harvested.

The erwin Metadata Management (EMM) Physical Data Model allows customers to document existing data stores, like databases, big data sources, imported as models in erwin Metadata Management (EMM), and publish the resulting documented data stores to the enterprise. When documenting a model, one may:

- Add business names and descriptions to objects, like tables and columns
- Document relationships including
  - o Cardinality
  - o Join conditions
  - Verb phrases
  - to better understand interrelated objects
  - Organize and view objects graphically using diagrams and subject areas
  - Annotate diagrams.

As data stores change over time, the [dataDocumenter] protects the investment in documenting those data stores, as the Physical Data Modeler migrates existing documentation to new data store versions automatically as they are re-imported.

In addition, the Physical Data Modeler is integrated with the business glossary, allowing users to manage terminology and domain definitions across the enterprise. One can reuse these definitions not only between different enterprise applications but between versions of the same application. A business glossary can be used to define and maintain data type domains and their business rules, like column possible values. Physical Data Modeler allows to link tables and columns to terms and columns to business glossary data types.

#### 2.1 Create a physical data model

We have already seen how to create a Physical Data Model in the Metadata Management Tutorial, specifically the chapters on harvesting and on initial business glossary exposure.

However, as a reminder, we will do so again for the Accounts Payable database. In this case, there will be two model, one that is harvested (Accounts Payable) and one that is a physical data model (Accounts Payable PDM) synchronized with the changes to the harvested model.

Right-click on the 2 − Data Warehouse folder select New → Physical Data Model, and confirm the following details:

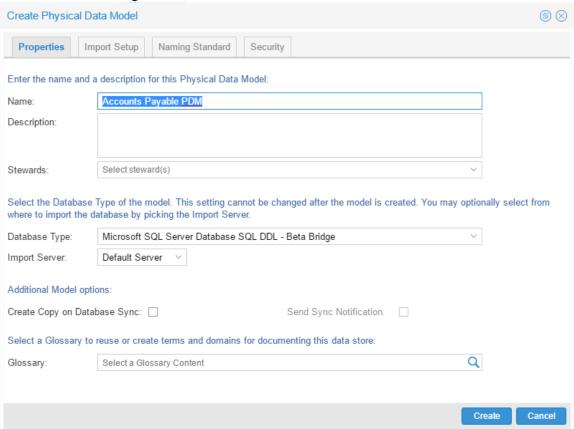
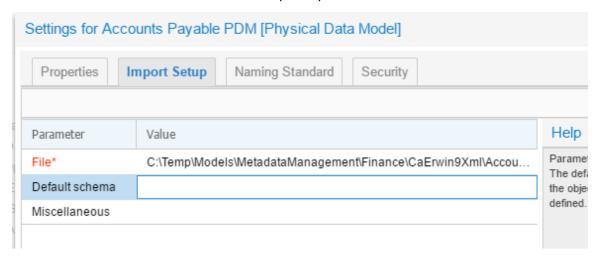


Figure 1 - Settings for the Staging DW model



Be sure to enter this in the File location (or browse for it):

sql

Click the Create button.

Then click the No button so as not to import.

#### 2.2 Bootstrapping from existing documentation

If documentation of a data store - even partial - is already available in a DM, ETL or BI tools one can use erwin Metadata Management (EMM) to bootstrap a physical data model imported from the metadata source. One merely needs to import the documentation into MM and drag & drop it into the physical data model. erwin Metadata Management (EMM) copies relevant content, like business names, descriptions, joins, and diagrams to the model.

If the target (documentable physical) model has PK/FK metadata erwin Metadata Management (EMM) assumes it matches the source model documentation and does not try to bring it to the target model. Otherwise, erwin Metadata Management (EMM) will bring in all available foreign keys as logical relationships. The process overwrites existing documentation and is designed for initial documentation population only.

For this, we already have an erwin DM model ready, just as before,

Create Model				
Properties	Import Setup	Security		
Enter the name a	nd a description f	for this Model:		
Name:	Accounts P	Accounts Payable Model		
Description:			_	
Stewards:	Select stewa	ard(s)	=	
Select an import I changed after the	_	odel. Optionally, you may specify the	S	
Server:	Default Ser	ver ∨		
Bridge:	erwin 9.x D	ata Modeler (File)		
Additional Model	options:			
Lightweight Version: Send				

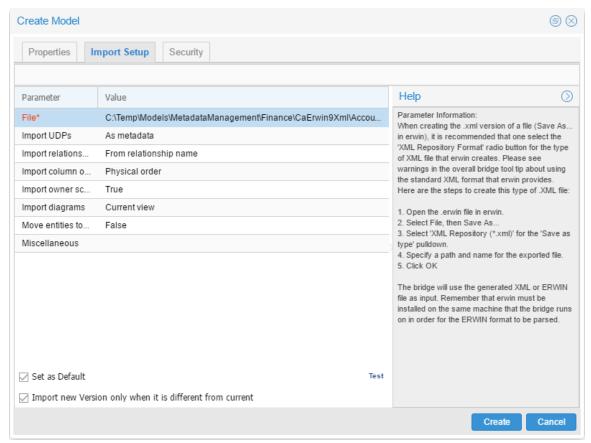


Figure 2 - New model from erwin DM

#### Be sure to:

- Not specify the Lightweight version, as these cannot be used to populate a glossary
- Enter this in the File location (or browse for it):
   C:\Temp\Models\MetadataManagement\Finance\CaErwin9Xml\AccountsPayable.xml

Import the model. The result is below (using the Show in Diagram option after selecting a table):

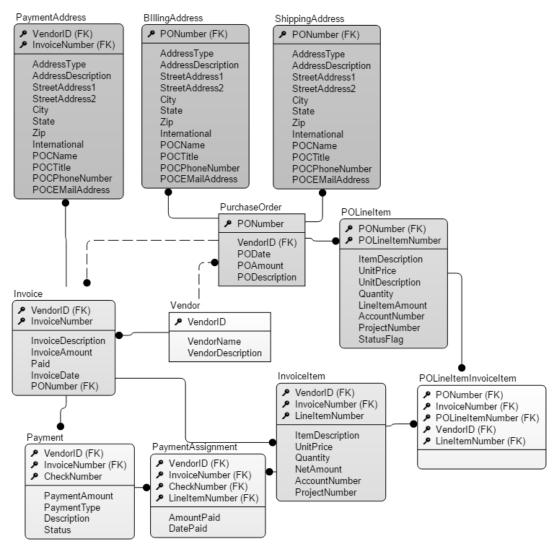


Figure 3 -Dimensional DW model from erwin DM

Open the Accounts Payable PDM physical data model we created in the last step, drag the Accounts Payable model (in the 1 – Operational Data Stores folder) into this Accounts Payable PDM physical data model:

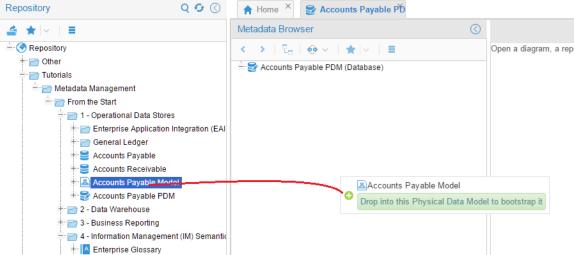


Figure 4 - Drag documented model into physical data model to bootstrap

After the log is finished, open the new version of the AP physical data model:

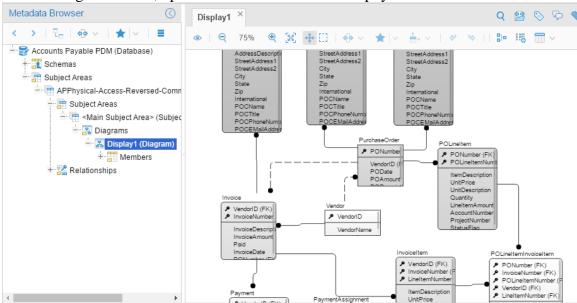


Figure 5 - Fully documented model

Now, if one wishes to sync with the underlying database, then right-click on the PDM and select Sync with Database

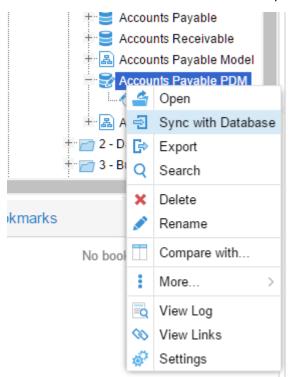


Figure 6 -Sync with Database

Then click the Import button.

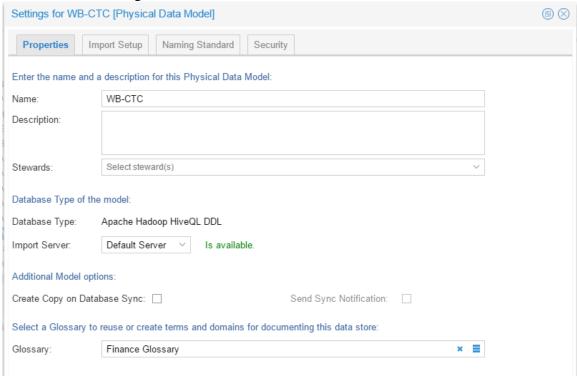
#### 2.3 Business Glossary Integrated Documentation

As one is documenting a new model, it would be very productive to integrate or even synchronize that effort with the building and maintaining of the business glossary. erwin Metadata Management (EMM) facilitates this in several ways:

- Linked term generation and classification
- Naming standards automatic documentation

#### 2.3.1 Create the physical data model

Right-click on the 2 – Data Warehouse folder select New → Physical Data Model, and confirm the following details:



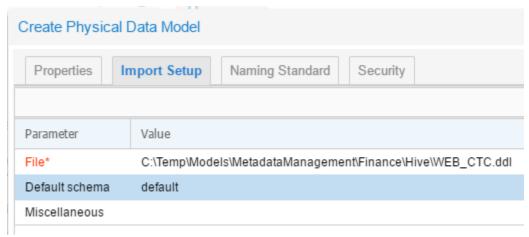


Figure 7 - Reference model specifications

DDL file is at C:\Temp\Models\MetadataManagement\Finance\Hive\WEB CTC.ddl

#### 2.3.2 Linked term generation and classification

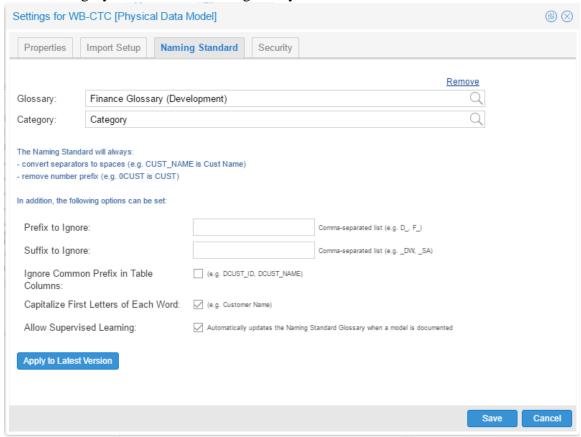
Note in the Properties tab we specified the Finance business glossary as the business glossary. Assigning a business glossary here means that we can use existing business glossary terms to document tables and columns in this model.

Also note we have not specified Create copy on Database Sync.

#### 2.3.3 Name/Abbreviation standards

In this scenario, we will also develop naming standards and use them to document a newly imported documented model.

In the Naming Standard tab, specify the same Finance business glossary and create and select a category inside that business glossary named "Contact":



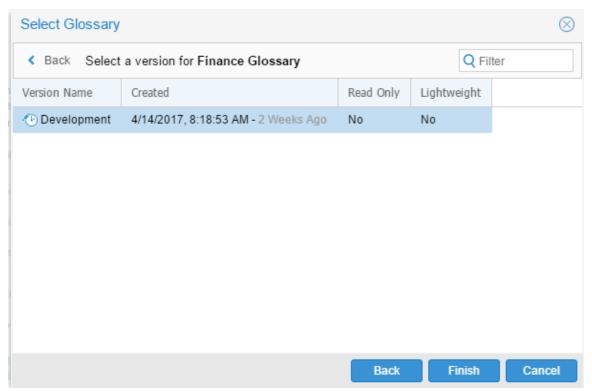


Figure 8 - Naming standards settings for Web Contact model

Assigning a business glossary to the Naming Standard tab means that the Abbreviation property of terms can be used to associate with physical names, and thus define recommended or standard business naming as well as associated physical names.

click the Save button.

Abbreviation/Naming standard (naming standard) can help one to derive business names from physical ones in a consistent manner. For example, a physical name <code>empl\_first\_name</code> can be interpreted as a business name <code>Employee First Name</code> by tokenizing the physical name over the "\_" separator, assuming "empl" is an abbreviation of "employee" and capitalizing the first letters of each word. This process is *automated* by erwin Metadata Management (EMM) based upon any defined *naming standard*.

A naming standard is a list of abbreviations and words they represent, defined as terms in a glossary. erwin Metadata Management (EMM) allows one to define a naming standard once and reuse it for documenting different models and their new versions. erwin Metadata Management (EMM) can generate an initial naming standard from a physical model. It is a list of all unique abbreviation tokens produced from table and column physical names. The application tokenizes names by separating characters (e.g. "\_") and capitalized letters (e.g. EmployeeName). The application cleanses the list from obvious noise (e.g. numbers).

There are a fair number of options available on this dialog. Some databases use standard prefixes or suffixes that are not clearly separated from the rest of the name (e.g. OCUST or CUSTD). For example, SAP prefixes physical names with a character from 0 to Z to

differentiate between system and custom objects. Some data warehouses suffix table names to differentiate between facts and dimensions. In case names have prefixes or suffixes that are not clearly separated from the rest of the name, erwin Metadata Management (EMM) allows one to specify a list of prefixes and suffixes to ignore (e.g. 0CUST -> CUST).

NOTE: if multiple tokens are join into one word without separators (e.g. CUSTNAME) the application does not help to separate tokens. Even if CUST abbreviation is defined the application will not apply it to CUSTNAME.

Abbreviation	Word	
usr	User	
nm	Name	

The generated business glossary has a term for each unique abbreviation (physical name token) in the model. Each term is named after its abbreviation and has the abbreviation attribute filled out. The term's description is set to a number of times the abbreviation is used in the model. One can sort terms by description (usage frequency) and use the sorted list to assign words to the most popular abbreviations only. One can decide to delete all abbreviations that are used only once.

The same naming standard can be used for documenting multiple models. The more models are involved the greater the return on the investment in creating a naming standard. When there are many models to document one can generate all their naming standards in the same business glossary category, erwin Metadata Management (EMM) keeps track and totals abbreviation usage frequency number across models. It can come in handy when an abbreviation is used multiple times in multiple models but zero or one time in each model. It can help to find abbreviations popular across models but not in any particular model, the more popular abbreviations the user can can find and document the better.

One can apply a naming standard to a freshly imported physical model to generate business names for its objects.

A business glossary category can reference terms defined in a different category. It promotes terms reuse. One can maintain multiple naming standards in a business glossary by staging them in different categories. In this case, one can have all abbreviation terms common to multiple naming standards reside in one category for ease of maintenance.

You will see four tables in this model. It is an extraction from a big data source with web log along with user and contact information. The four tables should be documented as so:

Physical Name	(Business) Name Description	
act	Activity	Activity on the web site
cntct	Contact	Basic contact information

		associated with an activity
idnty	Identity Identifying information	
	•	associated with a User
usr	User	Registered user

Editing a physical data model is often the task of non-technical business users, e.g. a subject matter expert. Thus, it is performed in the Metadata Explorer UI.

First, add this physical data model to the Finance configuration and validate so that it is available in that UI.

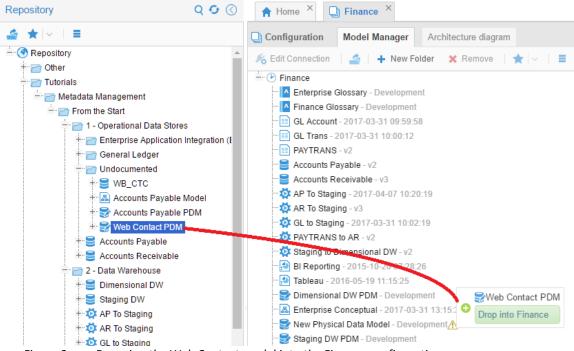


Figure 9 - Dragging the Web Contact model into the Finance configuration

Build the configuration.

Now, right-click on the act table in the Web Contact model and select Open in Metadata Explorer.

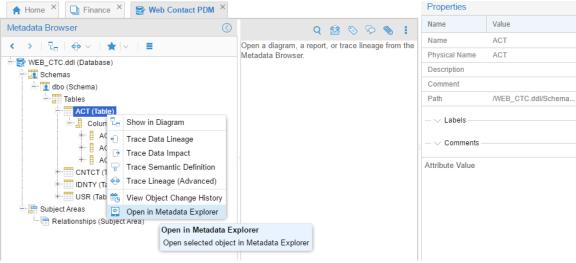


Figure 10 - Open in Metadata Explorer

To edit a table name, simply click on the name and you may edit in place. Enter the name Activity.

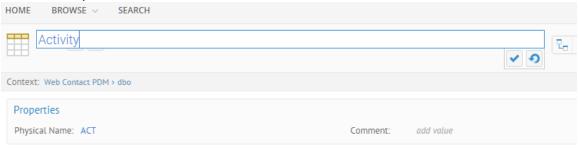


Figure 11 - Entering the name Activity

Similarly, one may do so for the description by simply clicking in the empty spaced below the name where the description would be displayed.



Figure 12 - Entering the description

However, we are going to do more. Note the [10] (Link to Term) icon next to where you can edit the name:

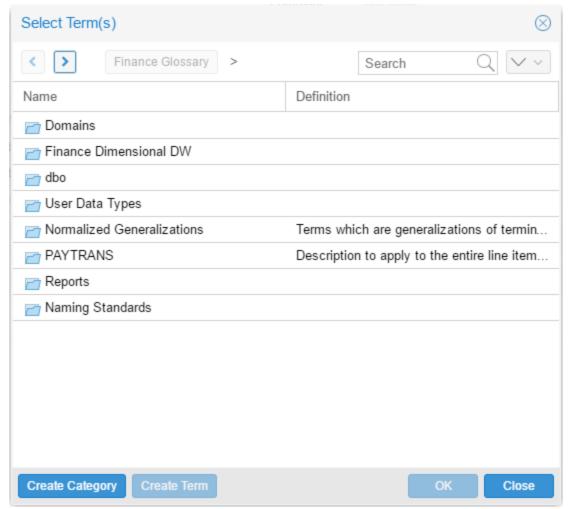


Figure 13 - Glossary integration with physical data model

Clicking on that icon takes you to the integrated business glossary we assigned to this physical data model. Using this method, one can ensure that terms are reused within this new physical data model and also that new terms may be created.

Now, we are ready to document the business names and descriptions from the above table. However, for each one we will either re-use a term in the business glossary or create a new one.

Navigate to the dbo category in the Finance business glossary and click the Create button.

Be sure that term is selected and click the OK button.

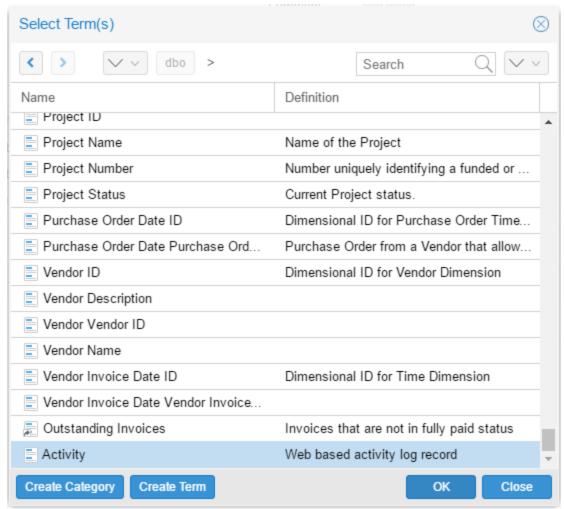


Figure 14 - Link to the new term

Note, the Name property has the text (Term) next to it indicating that this business name is based on a term in the glossary. When using terms as names within a physical data model, you are actually reusing the term name and definition within the model. This is not a real-time link, and thus updating the business glossary will not automatically update the table name and description. Instead, one may

- Diconnect from the term and then reconnect.
- Reharvest the model

and the name and description will be appropriately updated.

#### 2.3.4 Naming standard supervised learning

Now, let's take advantage of the supervised learning that has already taken place. Recall, we renamed act into Activity. Thus, in the Finance glossary in the Contact category, we have a new term Activity with abbreviation act. We can then use this information in naming the rest of the columns in the Activity table.

Click on the act\_ctxt column and click on the Get Standardized Name icon:



Figure 15 - Get standardized name

Note, the first part of the name is changed to Activity, and the name now has spaces rather than underscores, according to our rules laid out earlier.

We can also do this in bulk in the Metadata Manager UI. Click on the Open in Metadata Manager icon:



Figure 16 - Open in Metadata Manager

And then click on the More Options menu and select Apply Naming Standard:

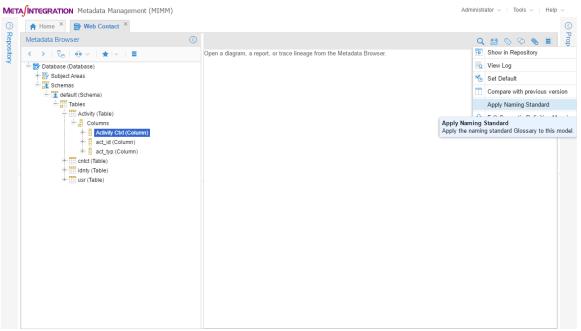


Figure 17 - Apply naming standard in Manager Ui

You may also do this from the Metadata Explorer UI by opening the model as a whole and clicking on the Apply Naming Standard icon:

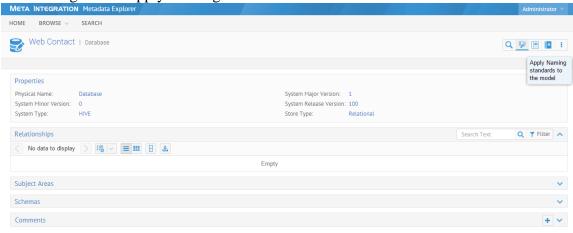


Figure 18 - Apply Naming Standard in Explorer UI

Either way, now when you view those columns, you will see the naming standard has been applied in bulk:

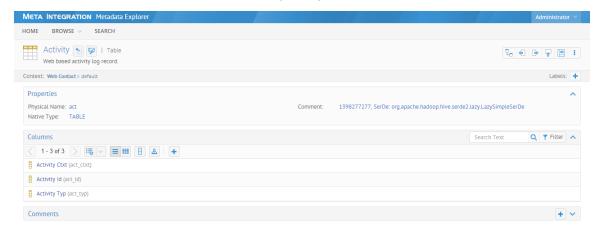


Figure 19 - Results of apply naming standard

We can actually provide names for the remainder of these column names using the grid view. Click on the Display grid icon:

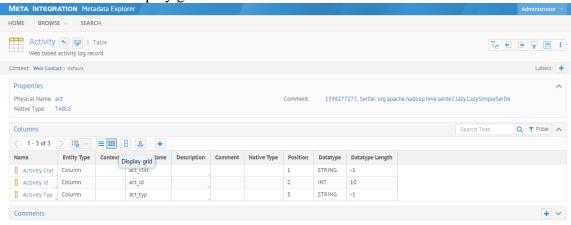


Figure 20 - Display grid

And double-click on each name and update according to the table:

Physical Name	(Business) Name	Description
act	Activity	Activity on the web site
act_ctxt	Activity Context	Context within the activity was captured
act_id	Activity Identifier	Unique identifier for activity on the web site
act_typ	Activity Type	Type of activity captured on the web site

Remember, this means that id is now an abbreviation for Identifier, etc.

It is now time to edit the standard to define words for abbreviations (usr=User) and then apply the standard to the model to generate business names (User Name).

In our example, click on the idnty table. Now, change the name to Identity and the description to "Identifying information to be associated with a User".

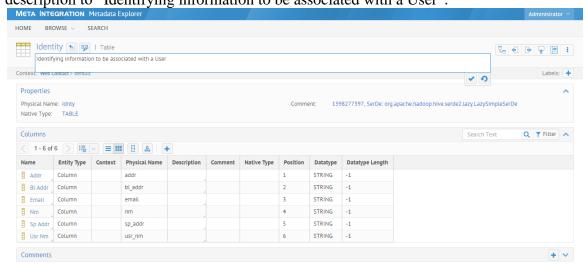


Figure 21 - Changing table name in a physical data model with a naming standard defined

Now, select the usr\_nm column. Rename this column "User Name" and the description to "User Name associated with this identifying information".



Figure 22 - Changing column name in a physical data model with a naming standard defined

Now, go to the Finance business glossary and open the Contact category. You will then see that the terms, like User and Name, are there with the appropriate abbreviations:

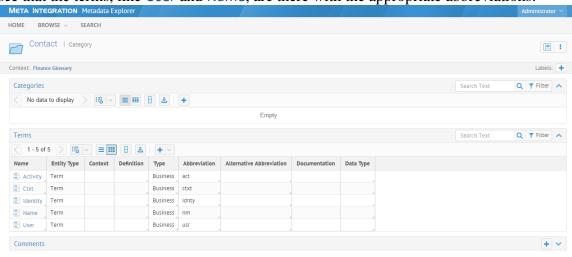


Figure 23 - Renamed term in naming standard glossary

Now, return to the Web Contact physical data model and click on Apply Naming Standard: After applying the updated naming standard, note that all the tables and columns with nm or usr in their name now have Name and User as business names:

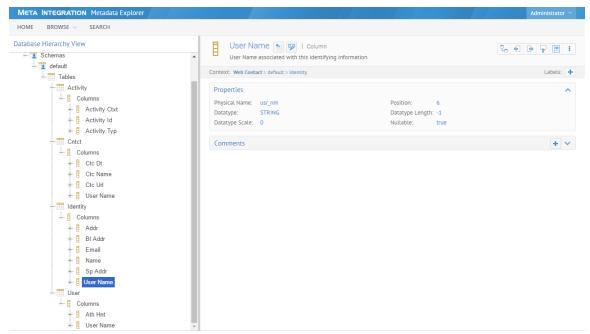


Figure 24 - Naming standards updates applied

Note, while a business glossary term can have only one abbreviation and one alternative abbreviation. Thus, a term can be used to map only two abbreviations to the same word. To get around this limitation, one can use multiple business glossary categories to define multiple terms by the same name to carry more than two abbreviations.

Please finish up documenting the model with the following names and descriptions:

Physical Name	(Business) Name	Description
act	Activity	Activity on the web site
act_ctxt	Activity Context	Context within the activity was captured
act_id	Activity Identifier	Unique identifier for activity on the web site
act_typ	Activity Type	Type of activity captured on the web site
cntct	Contact	Basic contact information associated with an activity
ctc_dt	Contact Date	Date when contact was first made
ctc_nm	Contact Name	Name entered during web contact
ctc_url	Contact URL	URL of contact
usr_nm	User Name	User name associated to this contact, if there is one
idnty	Identity	Identifying information to be associated with a User
addr	Address	Address information
bl_addr	Billing Address	Billing address
email	Email	Email address
Nm	Name	Full Name
sp_addr	Shipping Address	Shipping address
usr_nm	User Name	User Name associated with this identifying information
usr	User	Registered user
auth_hnt	Authentication Hint	Hint to remind user of password info
usr_nm	User Name	Name uniquely identifying the user

#### 2.4 Subject Areas and Diagrams

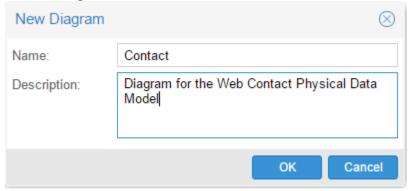
To document relationships and better understand interrelated objects the application allows to organize and view objects graphically using diagrams.

A diagram is a graphical view of a model. It shows tables, their columns and relationships between tables.

A model may contain zero or more diagrams. Diagrams reside in Subject Areas. A model may contain one or more Subject Areas. Subject Areas are like folders and can be hierarchically organized.

#### 2.4.1 Creating a Diagram

Open the Web Contact model, expand the Subject Areas panel, click on the Diagrams Subject Area, expand the Diagrams panel, then click on the plus sign (Add) icon, naming the new diagram "Contact":



Create new diagram Figure 25 -

Now, click on the Contact diagram and click the Start Editing button.

Then, drag and drop each of the tables listed on the left into the diagram on the right:

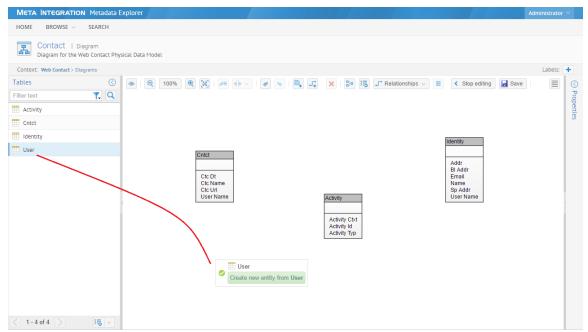


Figure 26 - Tables added to Contact Diagram

### 2.4.2 Editing a Diagram

Click on Display Preferences:

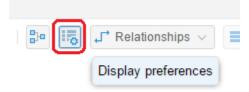


Diagram prope	erties			$\otimes$
General Di	splay format	Default Entity Display	Options	
Notation type:	Object Model	ing UML V		
Diagram Background color:	#FFFFFF	~		
Default Entity display type:	Attribute	~		
☐ Show lines o	n top			
		Apply	OK	Cancel

Figure 27 - Display preferences dialog

From the above, you can see that one can control notation, font, style, color, etc..

Now, click on the New Relationship icon and click on the Identity table and then click on the User table:

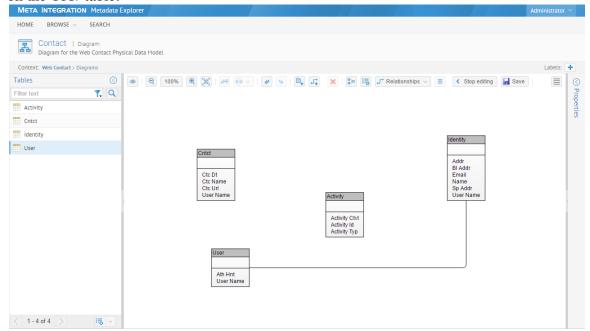


Figure 28 - Relationship from Identity to User

And we have a relationship. You may edit the relationship details. Double-click on the relationship line:

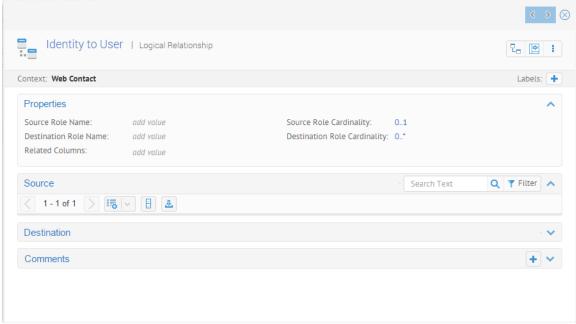


Figure 29 - Edit Relationship dialog

Define the relationship as shown below:

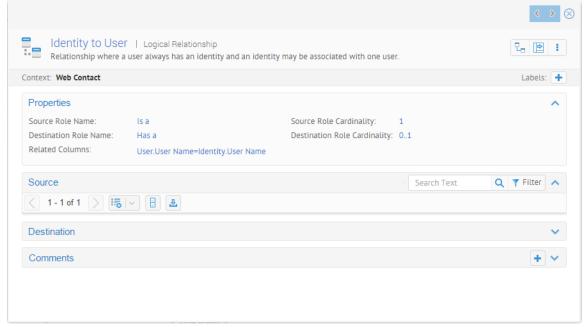


Figure 30 - Relationship join condition

One can also add annotations. Click on the New shape icon:

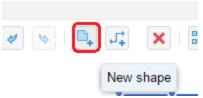


Figure 31 - New shape

And place it as shown:

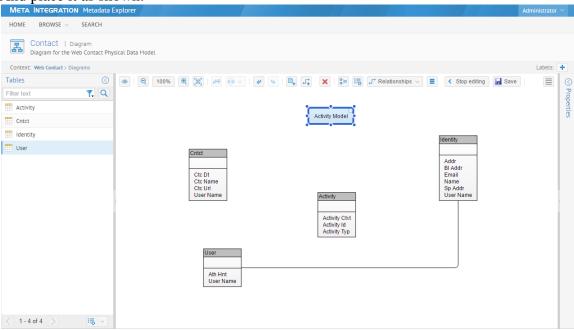


Figure 32 - Edited annotation

Now, try out the different features and update the diagram to conform with the below:

