Metadata Management Tutorial

Data Governance Best Practices Using erwin Metadata Management (EMM)

TABLE OF CONTENTS

TABLE OF FIGURES

1 Introduction

Data Governance can play a major role in developing a well managed data architecture, which in turn is reflected in a flexible and efficient technical architecture. A metadata management system, in particular, is a key tool in any governance process. In particular, erwin Metadata Management (EMM) provides:

- A central portal into your technical architecture and how it interrelates
- Web based tools to support development, deployment/implementation and management of a complete data architecture, such as
    - An ISO 11179 based business glossary to capture, define, maintain and implement an enterprise business glossary of terminology, data definitions, code sets, domains, validation rules, etc.
    - Semantic model harvesting from existing models and simple incorporation into glossaries and including in the semantic mappings
    - Logical/Conceptual Data Documenter to bring undocumented assets "into the fold" and provide logical names, definitions, domains, business rules and code set definitions to otherwise undocumented data stores in the technical architecture
    - Semantic Mapper to define relationships among the semantic layer assets such as glossaries and model and the technical architecture
    - Data Flow Mapper to define and/or document data processes which are or will be a part of the technical architecture
    - Full support for change management migration or forward and reverse engineering
    - Semantic usage and impact analysis and definition/lineage tracing and reporting
- Collaboration acceleration capabilities in all facets erwin Metadata Management (EMM) to facilitate administration, management and interaction within the data governance process and the entire information management and technology environment.
- Flexible data governance workflow process allowing for detailed or optional tool based enforcement of the processes and responsibilities.

While the methods and processes of data governance can have a great many features in common, in practice the actual business goals behind any data governance initiative vary widely. While many governance initiatives have the long term goal of an on-going effort to ensure that all the hard work early on is not lost, it is also critical to remember that immediate (and agile) returns on investment are provided. It can also be the case that the initiative truly is a one-time effort, solving a specific, high profile or profitable business goal, which cannot be "bogged down" by a more rigorous and hopefully future proof process or environment.

Because of these realities, erwin Metadata Management (EMM) is designed with a very flexible and optional set of organizational, inter-relationship and workflow rule and methods. It emphasizes:

- The strength of a rigorous metadata definition with flexible extensibility of attribution
- The flexibility of a very detailed and comprehensive selection of workflow rules while ensuring that they can be optionally taken "out of the way" when first bootstrapping a business glossary or attaining stretch goals
- The power of tight integration, re-use and forward-reverse engineering with the technical and semantic architecture that is the bread and butter of erwin Metadata Management (EMM).

In particular, erwin Metadata Management (EMM) focuses the data governance process in a business glossary.

This document is designed to help users navigate all of the features provided with the business glossary and determine the best practice for them, based upon their own specific requirements and current assets and capabilities. Instructions covering how to use full features of the business glossary are provided are provided in detail and through example in the Metadata Management Tutorial.

However, here is a brief overview of how the business glossary works.

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 Overview of the business glossary features

In erwin Metadata Management (EMM), a business glossary is a self-contained collection of categories and the terms sub-categories contained within each category. In turn, the terms may be semantically mapped to objects throughout the rest of the repository, such as tables and columns in a data model. Once mapped, one may perform semantic lineage traces such as definition lookups and term semantic usage across any configurations containing the business glossary, mappings and mapped objects.

Building a business glossary can be as simple as dragging in an existing well documented data model, via import from other sources via a CSV file format, or can be populated directly via the user interface as well as during the process of classifying objects in other data store models. In general, a combination of such methods are employed in conjunction with one another.

In order to ensure that the business glossary is accurate, up-to-date, available to all who need access to it and integrated properly with the rest of the metadata in the repository, erwin Metadata Management (EMM) also provides a robust collection of Data Governance tools and methodologies. The erwin Metadata Management (EMM) business glossary provides a very flexible workflow and publication process that may alternatively be quite sophisticated or quite simple depending upon one's needs. In addition, one may maintain any number of business glossaries, each with different workflow and publication characteristics.

The Business Glossary may be part of your lineage, will appear in the repository panel and when you open a Business Glossary, you will presented with a different UI than for other (harvestable) models.

1.1.1 Categories

A Business Glossary is organized into categories, which may then contain terms or other categories. Terms may be cross-linked in a wide variety of ways. Simply edit a specific term to do so.

1.1.2 Glossary Workflow

By default, a business glossary will have no workflow requirement (including no approval process). In this simple state changes made to the business glossary are reflected immediately throughout the system. In addition, in this open mode, the editing of the business glossary may be done primarily in the Metadata Explorer UI, and thus potentially open even to "crowd sourcing" of terminology, definitions, relationships and categorization.

This no workflow mode is a very useful (and very common) way for organizations to work when they are first building and populating a business glossary and related semantic mappings. However, it may also be useful in the long term for organizations that do not want or need the complexity of a workflow process. In fact, it is often sufficient to

simply limit the number of users who may edit the glossary as these people already "play well together."

If one wishes to or must enforce a workflow, the erwin Metadata Management (EMM) also provides a very flexible and complete set of possible workflow and publication processes that one may employ. Choose these processes carefully, as once selected they cannot be undone or changed.

When your organization would like to have a formal business glossary development process that involves multiple users you can enable the business glossary workflow. The workflow is prepackaged sequence of business glossary activities around term proposal, review, acceptance, publishing and depreciation. It is a flexible process that can be customized to require only publishing activity, approval with or without review, approval and review by one or multiple users, etc.

1.1.3 Workflow Roles

A user with the Administrator security role can enable the workflow and assign the following workflow roles to Categories:

- Editor
- Reviewer
- Approver
- Publisher

A workflow role can be assigned to users and applies to all terms in the category. A Category inherits all user-to-role assignments from the parent category and can have additional ones.

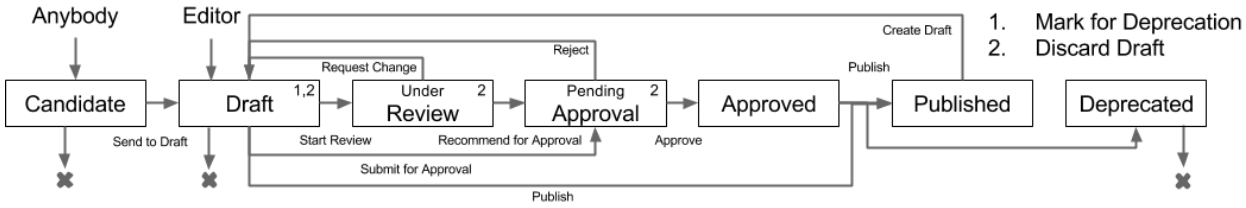
1.1.4 Stewards

Stewards are users or groups of users who are assigned as point of contact to answer questions for specific terms or entire categories. They have no special workflow role assignments based upon their stewardship assignment. However, notification of changes is based upon Stewardship assignments.

1.1.5 Workflow process options

The workflow process applies to terms, but not categories. All changes to Categories made in the Metadata Manager UI are immediately visible (published) in the Published version of the business glossary in the Published version of any containing configuration in the Metadata Explorer UI. When the workflow is enabled you cannot delete a category that has published terms.

The most complete workflow possible is in the diagram below:



You can enable the workflow when you create the business glossary or after. You cannot disable the workflow after it has been enabled.

1.1.6 Term management dashboards

Workflow driven search criteria are available allowing one to efficiently organize terms and identify what requires action at any given time. These include:

- Show all my candidate terms
- Show all my draft terms
- Show all terms under my review
- Show all terms pending my approval
- Show all term ready for me to publish

1.1.7 Workflow transition buttons

When working with individual terms which are at some point in the workflow process, workflow transition buttons prompt you with possible actions, e.g., if a term is in Draft status, then the action icons would include:

- Start Review
- Submit for Approval
- Mark for Deprecation
- Discard

1.1.8 Versions

The business glossary is one of the content types erwin Metadata Management (EMM) supports. As a content, it can have multiple versions. You can employ different business glossary workflow strategies that involve one, two and multiple versions of the business glossary. Here are some options:

- Simple - single version
- Dev vs. Prod - development and published versions
- Snapshots - historical versions

The business glossary may be part of your lineage, will appear in the repository panel, and when you open a business glossary you will presented with a different UI than for other (harvestable) models.

1.1.9 Categories

A business glossary is organized into categories, which may then contain terms or other categories. Categorization can help with:

- Subsetting by subject matter or organizational structure
- Managing stewardship assignments (at the category level)

1.1.10 Relationships

Terms may be cross-linked in a wide variety of relationship types, including:

- Synonyms
- See Also
- More General
- More Specific
- Contains
- Contained By
- Represents
- Represented By

Simply edit a specific term to do so.

1.1.11 Flexible and optional workflow process

The business glossary provides a comprehensive yet flexible and even option draft and candidate based edit/review/approve process (workflow). This workflow involves users with Editor, Reviewer or Approver responsibilities. Anyone with Administrator security role to the configuration may assign these workflow roles to users by setting these roles at the category level. This may be done for groups of users as well as individual users.

In addition, the process of publication is independent of the rest of the optional workflow steps. Publication of a draft of a term in the business glossary means that this new version is now visible to users in the Metadata Explorer UI. In this way, one has full control over publication and dissemination of terms independently of where they are in the workflow process.

Note, this means that while one can edit the business glossary in the Metadata Explorer UI when workflow is disabled.

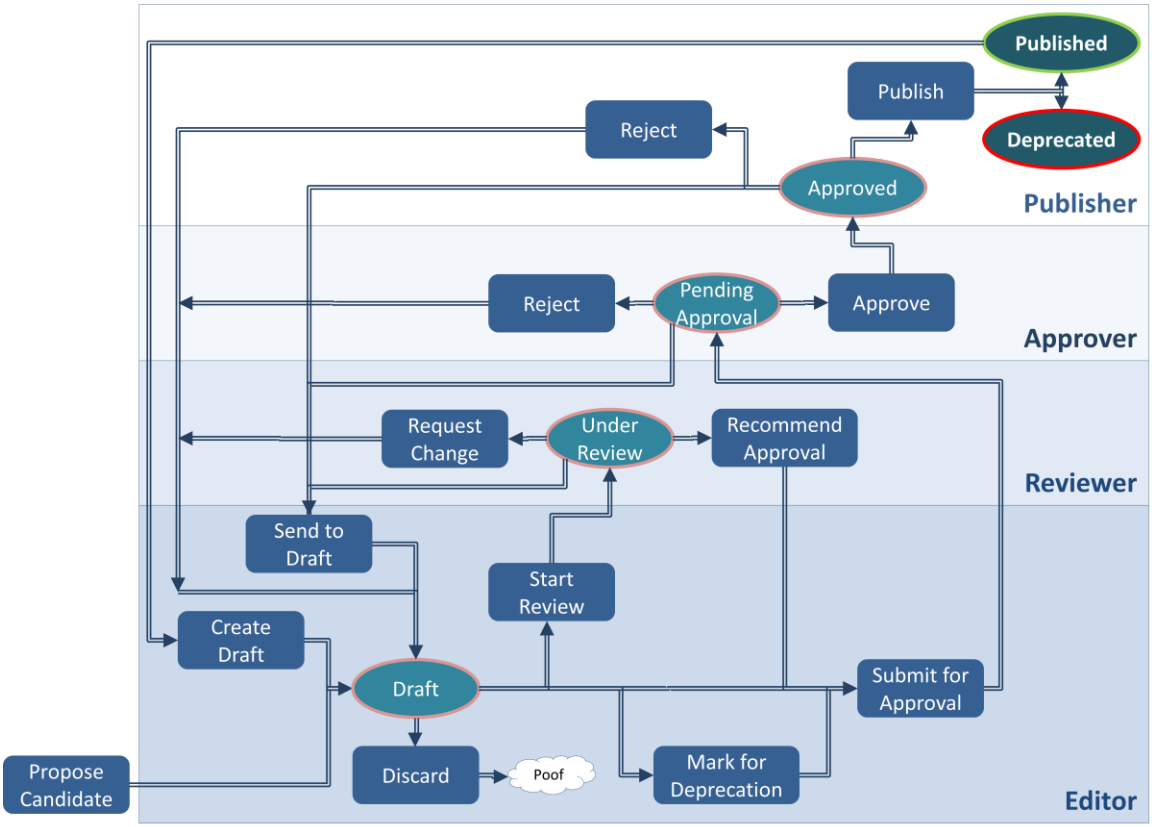


Figure 1 - Simple workflow diagram

In the above diagram

- The four workflow roles are the horizontal bands
- Activities conducted by each role (expressed in the Metadata Manager UI as workflow buttons)
- The process flow (arrow lines)
- Draft/Published/Deprecated terms are ovals.

As can be seen in this workflow diagram, the Editor role is the primary one. This is the role of the Steward.

1.1.11.1 Workflow actions

Workflow Action	Workflow Role			
	Editor	Reviewer	Approver	Publisher
Propose Candidate	X	X	X	X
Create Draft	X			
Discard	X			
Start Review	X			
Mark for Deprecation	X			
Submit for Approval	X			
Send to Draft	X			

Recommend Approval		X		
Request Change		X		
Reject (Awaiting Approval)			X	
Approve			X	
Reject (Approved)				X
Publish				X
Publish (Deprecate)				X
Create, edit or remove attributes and relations	X	X	X	
Create comments	X	X	X	X
Edit or remove comments	X	X	X	
Create, edit or remove attachments	X		X	
Assign roles to users and groups			X	
Start/stop workflows and reassign tasks	X		X	

Responsibility Details:

- An *Editor* can create, edit and delete a term. A new term has the Draft status.
- A *Reviewer* can comment on terms under development and recommend for approval or further change/review
- An *Approver* can reject or approve a term
- The *Publisher* can reject, publish or deprecate a term. Publication means that a term is visible from the Metadata Explorer UI.

1.1.11.2 Term Status

As the workflow process is separate from the publication process, there are two different status indicators for a term. Workflow status relates to the status of any draft in the workflow process.

Table of business glossary term workflow status possible values:

Workflow Status	Description
Draft	Draft term being edited before review
Under Review	Draft term being reviewed
Pending Approval	Draft term awaiting approval
Approved	Draft term approved and ready for publication
Published	Term is published and has no draft
Deprecated	Term is deprecated and cannot be used further

Figure 2 - Term workflow status table

Table of business glossary term publication status possible values:

Publication Status	Description
New Term	Candidate draft term before it is first published
Published Term	Term that has been published and the non-draft properties are visible on the Metadata Explorer UI
Deprecated Term	Term that has been deprecated after having been a published term. It is not longer visible on the Metadata

	Explorer UI

Figure 3 - Term workflow status table

Status may be assigned term by term, or at the category level (all terms contained within).

In addition, the Metadata Explorer UI only shows assets which are in Approved status, while the Metadata Manager UI allows one to search for assets by **status** and **steward**. For example:

- An *Approver* can list assets that are ready for approval, with the Reviewed status.
- An *Editor* can list assets assigned to her/him as *Steward* that are in draft mode, Candidate status.
- A *Reviewer* can list assets in a particular category or the whole business glossary that are under development, with Candidate status.

Basic Rules:

- You can edit Candidate assets only.
- An *Editor* can submit an asset for approval by setting its status to Reviewed.
- An *Editor* can change the status back to Candidate to continue editing an asset.
- Only and *Administrator* can assign a *Steward* to an asset.
- Only an *Approver* or an *Editor* can change and asset's status.

Permission		**View Reviewer**	**Update Editor**	**Administer Approver**
Term	Create	No	Yes	Yes
	Delete	No	Yes	Yes
	Edit Attributes	No	Yes*	Yes*
	Assign Steward	No	No	Yes*
	Change Status	No	Candidate<->Reviewed	Candidate, Reviewed, Approved
Category	Create	No	Yes	Yes
	Delete	No	Yes	Yes
	Edit Attributes	No	Yes*	Yes*
	Assign Steward	No	No	Yes
Domain	Create	No	Yes	Yes
	Delete	No	Yes	Yes
	Edit Attributes	No	Yes*	Yes*

Assign Steward	No	No	Yes
Comment	Yes	Yes	Yes

* When a user is steward (or when there is no steward assigned) and the Term is Candidate

Figure 4 - Permissions and role table

1.1.12 Configurations and versioning

Of course, in order to trace lineage, search, report, etc., one will want to include the business glossary and any Semantic Mappings you create inside any number of configurations.

Now, you may update these later at anytime. And, each time you edit the business glossary or Semantic Mapping, the configuration must be validated.

This can be an awkward requirement. But, keep in mind that it is only a version of the business glossary and any mappings that are included in a version of a configuration. This, one good solution is to create a new version of the configuration and business glossary and mappings for major editing (and a long approval process). This way, each time you edit the future version of business glossary or Semantic Mapping, only the newer (unpublished to business users) configuration needs to be validated.

2 Glossary Activities

There are several business glossary activates which must be considered before one can best determine the course of action to take in building and managing a business glossary and enterprise architecture. These can be grouped as follows:

- Population
  - Drag and Drop
  - Import CSV
  - While Mapping
  - While Documenting
- Categorization
  - By Responsibility
  - By Stewardship
  - By Subject Matter
- Mapping
  - Drag and drop
  - Mapping Editor
    - Mapping
    - Creating a Term
  - Documenting
    - Re-using a term
    - Creating a tem
- Governance
  - Workflow
  - Responsibilities (Roles)
  - Stewardship

2.1 Population

This activity refers to all the activities required to create terms and domains within the business glossary, fully document their properties and descriptions and interrelate them through the various relationship types.

An important consideration when producing any semantic model (including a business glossary) is to take advantage of the metadata sources with the most well defined business metadata such as descriptions, domains, business rules, relationship definitions, generalizations, etc.

2.1.1 Drag and Drop from and Imported Model

The simplest means of populating is to use an already imported or harvested data model from one of the data modeling tools that erwin Metadata Management (EMM) imports from. When populating a business glossary based upon one of these metadata sources, erwin Metadata Management (EMM) will look for all of these types of metadata and create equivalents within the business glossary. In addition, it can create a new semantic mapping and create individual mappings between the newly create business glossary equivalents and the specific metadata objects from which they were constructed. In this way, one may perform sophisticated impact (where used) and reverse lineage (show definition or meaning) type of lineage analysis.

See the erwin Metadata Management (EMM) - Data Governance Tutorial for detailed examples of populating the business glossary in this manner.

2.1.2 Import CSV

If you have an external source of business glossary metadata that is not accessible via the bridges (not harvestable), then you may use the CSV format import capability. In this case, you may import categories terms and domains and all their properties. Many of these external sources (e.g., on-line data dictionaries or other business glossary type products) will have additional properties or attribution on the objects that will become categories terms and domains in the business glossary after import. You may extend the business glossary with any number of such addition user defined properties/attributes to support these in the original source.

See the erwin Metadata Management (EMM) - Data Governance Tutorial for details about populating the business glossary in this manner.

2.1.3 While Mapping

One may also populate the business glossary in the process of mapping terms to data elements (e.g., tables, columns) in you other harvested models. The semantic mappings are also created this way. This is discussed more below under Mappings.

2.1.4 While Documenting

When using the data documenter, one can create new terms based upon the logical names and descriptions of data elements being documented. The semantic mappings are also created this way. This is discussed more below under Mappings.

2.2 Categorization

A business glossary is organized into categories, which may then contain terms, domains or other categories. Categorization can help with:

- Managing stewardship assignments (at the category level)
- Subsetting by subject matter or organizational structure

2.2.1 Managing Terms and Domains

As the Metadata Manager UI presents terms according categories first, it is especially useful to organize terms and/or domains by category so as to make update and administration easy. E.g., by category one may:

- Edit terms, domains and categories like a spreadsheet (See the erwin Metadata Management (EMM) - Data Governance Tutorial for detailed examples).
- Edit multiple (batch edit) the status and steward of terms (See the erwin Metadata Management (EMM) - Data Governance Tutorial for detailed examples).

2.2.2 Managing Terms in the Workflow

As there are three statuses, there may still be a need to scope what should be edited, reviewed and or approved at a given time. E.g., if a term is in Candidate status, this could mean that the term is being edited, is ready for review or has been reviewed but updates have not yet been applied. One way to manage this is to organize terms to be worked on by category, thus the *Editor* can notify *Reviewers* that the terms is a particular category are the ones ready for review, and thus controlling the scope.

2.2.3 Subsetting by Subject Matter or Organizational Structure

Obviously, terms and domains may be organized by subject matter or responsibility of systems within the architecture. This may also comport well with the management of terms in the workflow, above, as editing, review and approval acuities often group by subject matter.

2.3 Mapping

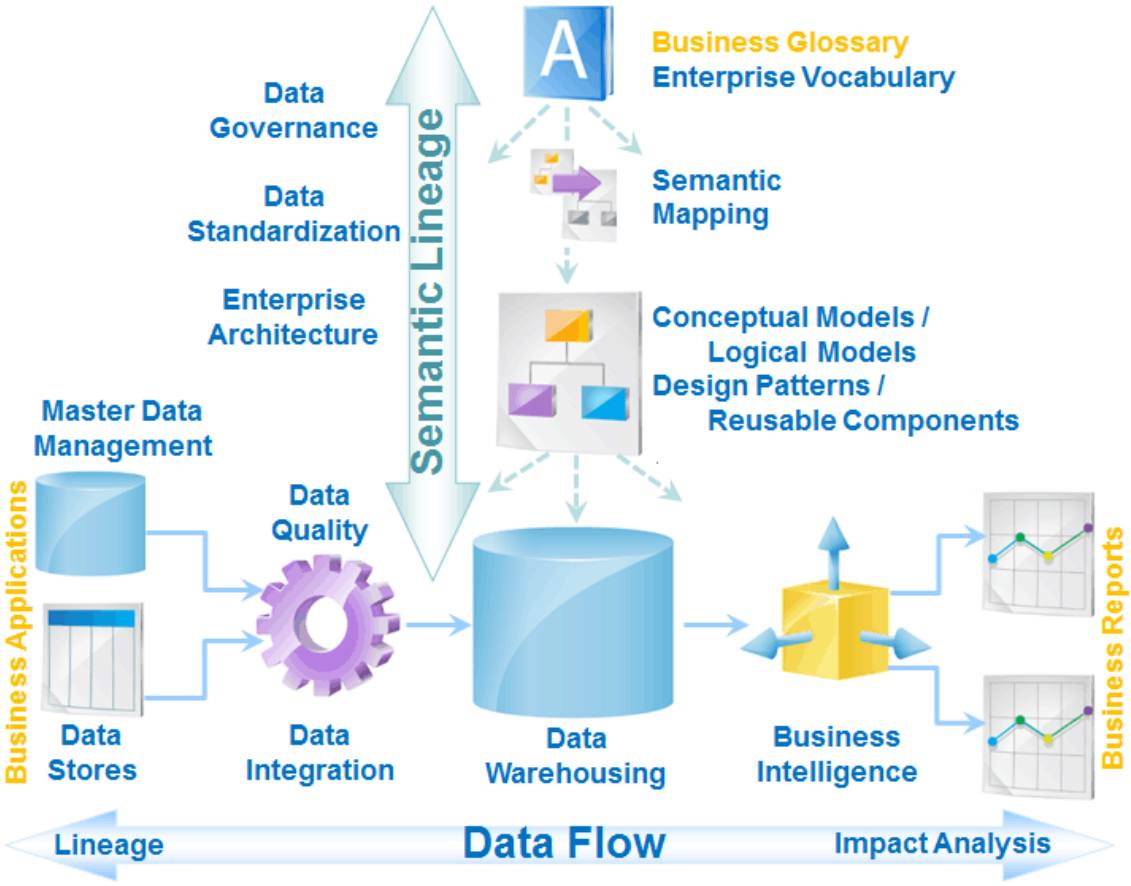


Figure 5 - Semantic mapping for lineage down to the data flow architecture

In order to facilitate nearly all of the features that end-users look for from a business glossary, it is a requirement that one populate the semantic mappings from the business glossary (terms) to models in the logical and physical architecture, including:

- Definition Lookup
- Report Lookup
- Semantic Usage
- Transitive closure

The process is generally one of:

- Creating a Semantic Mapping as a content in erwin Metadata Management (EMM)
- Linking this Semantic Mapping to both the business glossary and a model (generally the data warehouse) in your data flow architecture
- Adding the Semantic Mapping to your configuration.

From here you may open the mapping and edit by drag and drop.

The Data Warehouse is generally chose as the target of the mapping because it is the portion of your architecture which should contain data elements representing nearly all of the type of concepts used in the system and reported on.

2.3.1 Auto-Population

Again, the simplest means of populating is to use an already imported or harvested data model from one of the data modeling tools that erwin Metadata Management (EMM) imports from. When populating a business glossary based upon one of these metadata sources, erwin Metadata Management (EMM) will look for all of these types of metadata and create equivalents within the business glossary. In addition, it can create a new semantic mapping and create individual mappings between the newly create business glossary equivalents and the specific metadata objects from which they were constructed. In this way, one may perform sophisticated impact (where used) and reverse lineage (show definition or meaning) type of lineage analysis.

Note, this is especially simple as the Semantic Mapping is already created for you and connected to the correct models. In addition, all of the good data modeling design information is translated into concepts in your business glossary. Note, however, that the mapping created is to the actual model you dragged in (likely a data model of the ultimate database in your physical architecture), not any database that you may have harvested. If it is the database that you have connected (stitched) into your architecture, then you will still need to semantically map the data model to the database. Oftentimes, that is simply a stitching in the configuration manager, but can be more complicated if the model does not match the database.

See the erwin Metadata Management (EMM) - Data Governance Tutorial for detailed examples of populating the business glossary in this manner.

2.3.2 While Mapping

One may also populate the business glossary in the process of mapping terms to data elements (e.g., tables, columns) in you other harvested models. The semantic mappings are also created this way.

2.3.2.1 Simple Mapping

Of course, when editing a Semantic Mapping, one can drag and drop a term into a data element. This will create an individual map, and a term may be mapped to any number of terms.

See the erwin Metadata Management (EMM) - Data Governance Tutorial for details about mapping the business glossary in this manner.

2.3.2.2 Creating a Term

However, it is also possible to edit the mapping and drag the data element into a category in the business glossary side of the mapping. In this case, erwin Metadata Management (EMM) will create the term for you using the logical name and description in the model where the data element resides.

This method can be especially useful when wishing to add only a subset of terms based upon the other model, rather than the Auto-Population, which will create terms for everything in the model. The disadvantage as compared with the auto-populate method is that this method does not populate all the properties of the term, nor does it populate domains or inter-related terms and domains.

See the erwin Metadata Management (EMM) - Data Governance Tutorial for details about mapping the business glossary in this manner.

2.3.3 While Documenting

The Data Documenter may be intimately tied into a business glossary. In this case, one actions taken in the Data Documenter can impact both mappings and terms in the business glossary.

2.3.3.1 Reusing a Term

When reusing a term to provide documentation for a data element in a documented model, a semantic link is formed. In this way, definition looking, usage analysis, etc., will all be possible from the reused term and that data element, it up and downstream data flow and any downstream report fields.

See Section 2 in the erwin Metadata Management (EMM) – Data Store Documentation Tutorial for details about mapping the business glossary in this manner.

2.3.3.2 Creating a Term

The Data Documenter also allow one to create terms on the fly as one is providing logical names and descriptions to data elements. In this way, again a semantic link is created between the term and data element, and definition looking, usage analysis, etc., will all be possible from the reused term and that data element, it up and downstream data flow and any downstream report fields.

This can be a very powerful way of build up large sections of the business glossary as you document models. Again, the disadvantage as compared with the auto-populate method is that this method does not populate all the properties of the term, nor does it populate domains or inter-related terms and domains.

See Section 2 in the erwin Metadata Management (EMM) – Data Store Documentation Tutorial for details about mapping the business glossary in this manner.

2.4 Governance

2.4.1 Workflow

If one wishes to or must enforce a workflow, the erwin Metadata Management (EMM) also provides a very flexible and complete set of possible workflow and publication processes that one may employ. Choose these processes carefully, as once selected they cannot be undone or changed. In addition, in keeping with strict enforcement of workflow and publication (only published terms are visible in the Published version of the glossary which is a part of any number of Published configuration versions in the Metadata Explorer UI).

The workflow an organization ultimately develops can be simple or complex. The simple workflow rules that erwin Metadata Management (EMM) adheres to are very flexible and should support any number of edit/review/approve type processes.

It is certainly possible to have different groups adhering to different workflows, where the governance processes dictate how the simple erwin Metadata Management (EMM) workflow is utilized for each group.

The entire process is laid out in the erwin Metadata Management (EMM) – Metadata Management Tutorial on Data Governance Enforcement, along with step-by-step examples. Below are specific issues one may see.

2.4.2 Responsibilities (Roles)

Responsibilities are a reuse of the erwin Metadata Management (EMM) role base security and user interface assignments. As such, it is only users in the Administrator role who will be able to assign role based responsibilities. This includes assigning the appropriate privileges on the business glossary itself or the folder it is contained in. However, once in place, it should not require much maintenance.

Keep in mind, by using the Responsibilities method, the span of control is an entire business glossary at a time. Role assignments cannot be subdivided by category or term. For that, you must use the stewardship model. Of course, one may use the Responsibilities method overall but assign stewards in cases where they want to assign a span of control at the term level.

In general, what you will find is that each role will generally want to perform a limited set of searches in the business glossary in order to manage activities and identify what needs to be given attention:

- Editor:
  - Show all terms in Candidate Status
  - Show all terms in Candidate Status in a particular category
  - Show all terms in Candidate Status with a particular label
- Reviewer
  - Show all terms in Candidate Status

    - Show all terms in Candidate Status in a particular category
    - Show all terms in Candidate Status with a particular label
- Approver
    - Show all terms in Reviewed Status
    - Show all terms in Reviewed Status in a particular category
    - Show all terms in Reviewed Status with a particular label

These are detailed in the tutorial.

Now, say the Editor has applied changes to a set of terms in different categories. There can often not be enough information to distinguish which terms should be considered "being edited", "ready for review", "reviewed but comments are not yet assimilated". All those terms would be in the Candidate status. This is where organizations by category can come in handy, conducting updates and reviews and approvals according to category.

However, while that may be a good practice when first building the terminology or adding an entire section (category), it is often not practical for small number of terms being re-worked. In this latter case, it is useful to use the Label function and assign a label to terms as they move through this process. Remember, this is only for the limited cases (and numbers) where terms are being sent back to be re-worked.

2.4.3 Stewardship

Stewards are users who are responsible for the definition, purpose, and use of assets, like harvestable Models terms in the business glossary. A steward who works with the business glossary is assigned to various terms and categories. You may assign one user as a steward to an asset and category stewardship assignments are not inherited to the terms and categories contained within.

When a Steward is assigned to a term, that person has the *Editor* and *Approver* responsibilities for that asset. When an asset does not have an assigned *Steward* any *Editor* and any *Approver* can handle it.

Stewards then would have slightly different queries:

- Show all terms in Candidate Status for which I am Steward
- Show all terms in Reviewed Status for which I am Steward
- Show all terms in Candidate Status for which I am Steward with a particular label
- Show all terms in Reviewed Status for which I am Steward with a particular label.

Again, labeling may be useful in the same narrow cases, but in general when assigning a steward, that person already knows the subtle "ready for review" type statuses already and does not require labels.

3 Overview of Business Cases for the Business Glossary

Using erwin Metadata Management (EMM) a business glossary can simply be a tool for providing a list of data elements I.e., a pool of definitions one may search through and report on. In that case it is more a data dictionary than a complete business glossary.

One can take this goal one step farther and a business glossary can also include terms, domains and even business rules with descriptions, inter-relationships, code sets, etc. Again, the sole purpose could be for reference. In that case it is a business glossary. This can also be especially helpful when standardizing terminology and ultimate data elements used in your architecture.

However, even then it will not be taking advantage of the full benefits of erwin Metadata Management (EMM) and the way it allows one to integrate the business glossary into the data flow architecture.

3.1 The Power of Semantic Mapping

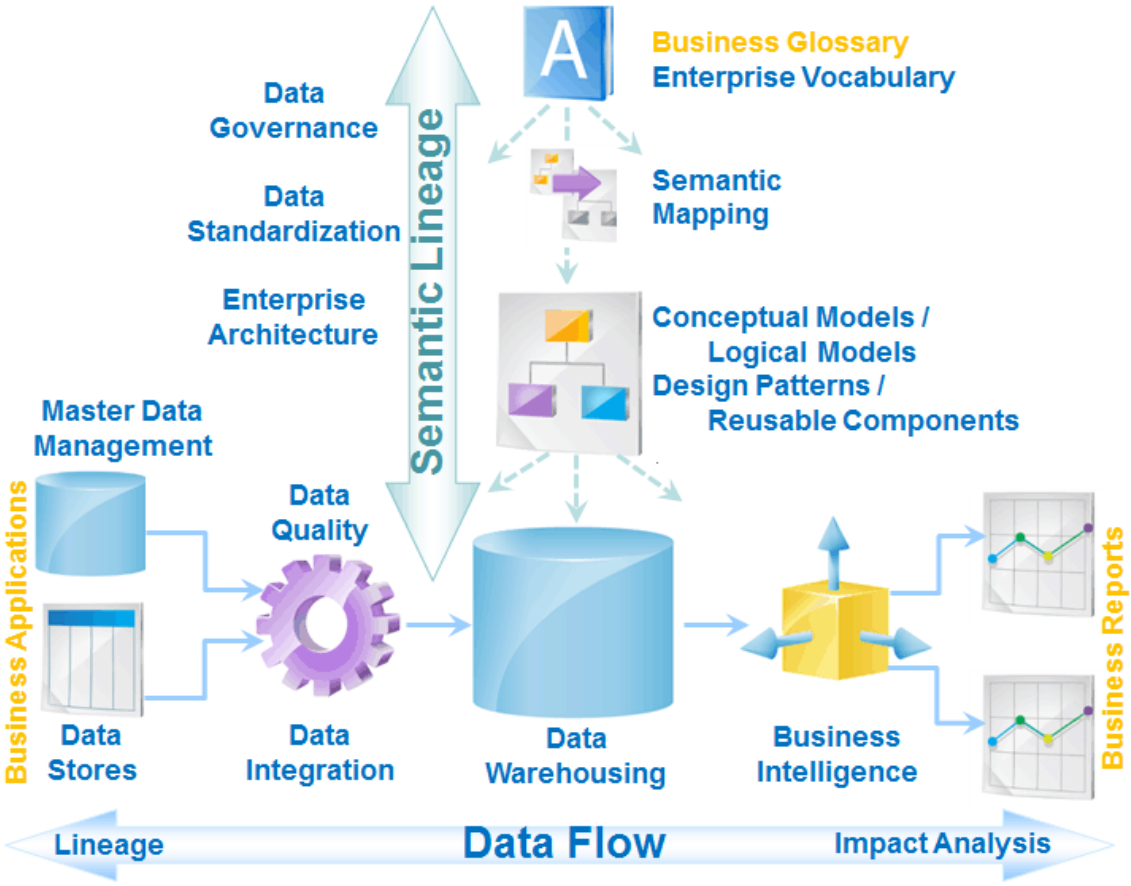


Figure 6 - Semantic mapping for lineage down to the data flow architecture

With erwin Metadata Management (EMM) one is not only able to document, relate and standardize terms in a self-contained business glossary, but one is also able to relate (map) these terms down to their representations as data elements in the data flow architecture that has been harvested and stitched.

In particular, there are three very important business user use cases which can only be addressed complete with the semantic links:

- Data element *definition lookup*
- Display a *report glossary*
- *Show related business reports*

Then, we can put it all together in a *report portal.*

Another use case relates to managing these semantic links in a simple, intuitive process directly in Metadata Explorer UI, as well as the workflow required method using the Metadata Manager UI.

Finally, two very important technical user use cases:

- Semantic usage and impact
- Transitive closure

3.1.1 Data element definition lookup

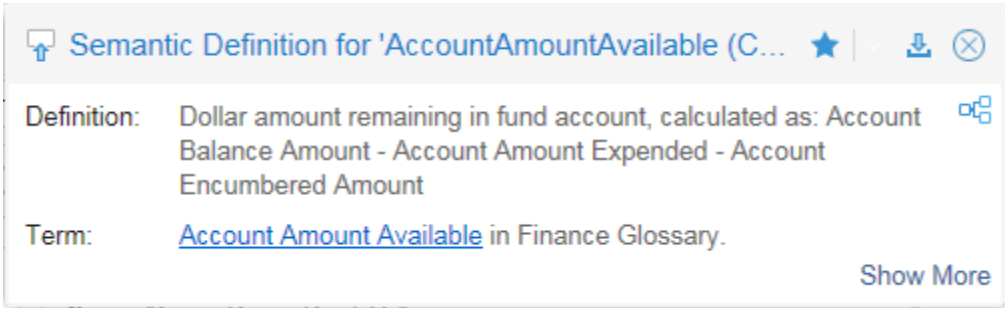


Figure 7 - Semantic Definition of AccountAmountAvailable traced back to a term in the business glossary

In this use case, one has found a data element (a column in a table in a database for example, or a field in a report) and wants to understand what it means. By defining the semantic links properly, erwin Metadata Management (EMM) can trace back through the physical data flow (as long as there is no transformation which would change the meaning) to an element that is mapped to a term in the business glossary and thus find a useful definition.

The caveat that the above only works "as long as there is no transformation which would change the meaning" implies that some subset of the fields in your reports will not provide a semantic definition. The trace will simply stop at the transformation and never get to a model (again likely the data warehouse) that has semantic lineage.

So, in addition to this method of "trace through the dataflow as long as there is no transformation which would change the meaning", there is another which is search based or name matching based. In this case, if there is a field in a report named "Net Account Amount" and it does not have a good data flow trace without transformation, one could still create a term in the business glossary named "Net Account Amount". When requesting a data element definition lookup in that case, erwin Metadata Management (EMM) will perform a search for that term and report its definition, even without a clean lineage trace. In most case, it will be necessary to fill in the blanks in some of these cases by adding terms to the business glossary.

Of course, it is quite possible that no term directly matches the report field by name. In this case, one may define a direct semantic mapping from a term in the business glossary to the field in the report. The advantage of this approach is that one may control precisely what the preferred definition will be versus the name matching method. Also, it provide a definition, even though there may not be a data flow trace that does not contain transformations. Hence, it is the preferred method for fields for which there is no equivalent in the warehouse or lake (i.e., calculated in the report) and there is no term or multiple terms that match by name.

To summarize, there are three methods used to provide an answer to a definition lookup, applied in the following order:

1. Direct semantic link from a term in a business glossary
2. Data flow lineage to the warehouse (or some other data store in the physical architecture) and semantic up to a [businessGlosssary]
3. Name match between business glossary term and field name

For completeness, though it is not a business glossary topic, the same rules above apply for fields (tables and columns) in a physical data model that may be in the physical data flow and sementic architecture. In this case, the fields (tables and columns) in a physical data model provide the same type of information (definition lookup) as terms in a business glossary.

Finally, there may be a definition for the field itself harvested from the source BI environment. This is, of course, already displayed with the name of the field, so the definition lookup does not consider it.

Please see the erwin Metadata Management (EMM) - Data Governance Tutorial for detailed examples.

3.1.2 Report Glossary

Many times one wishes to include a business glossary of all fields in a given report. This way one can quickly identify the meaning of all the fields, rather than click on each individually for a definition lookup.

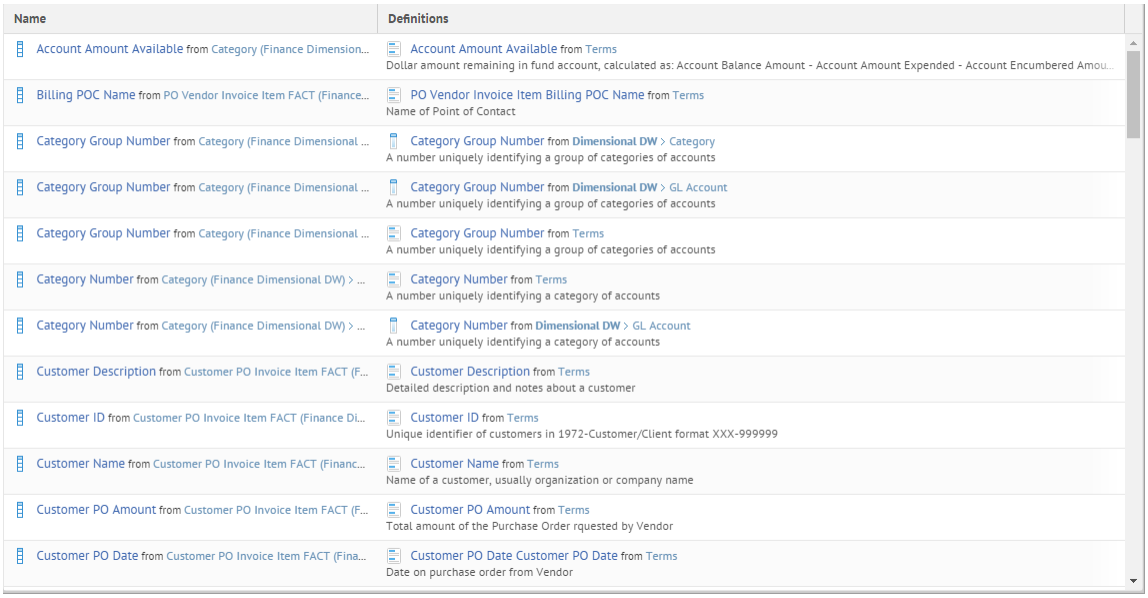


Figure 8 - Report Glossary

This can be displayed in the Metadata Explorer UI using the Report Glossary action icon. In addition, one may customize the UI so that it responds to URL that may be called from outside to produce this report (See Customization Tutorial document), even

from a third-party business intelligence tool. Full control of what is displayed (e.g., hyperlinks, icons, report field definition, etc.) may be specified in the URL or in the customized UI file.

Again, definitions are determined by the same logic provided for individual report fields:

1. Direct semantic link from a term in a business glossary
2. Data flow lineage to the warehouse (or some other data store in the physical architecture) and semantic up to a business glossary
3. Name match between business glossary term and field name

Finally, there may be a definition for the field itself harvested from the source BI environment. In this case, the definition may appear below the name of the report field.

As mentioned above, one can include or exclude as much of the displayed information as one wants to, either by updating the UI customization files or by adding display control specifications to the URL signature. Included in these are the:

- Glossary report header
- Term (or other object) name and/or definition
- Hyperlink to the term or other field with a definition
- Etc.

Please see the erwin Metadata Management (EMM) - Data Governance Tutorial for detailed examples.

3.1.3 Show related business reports

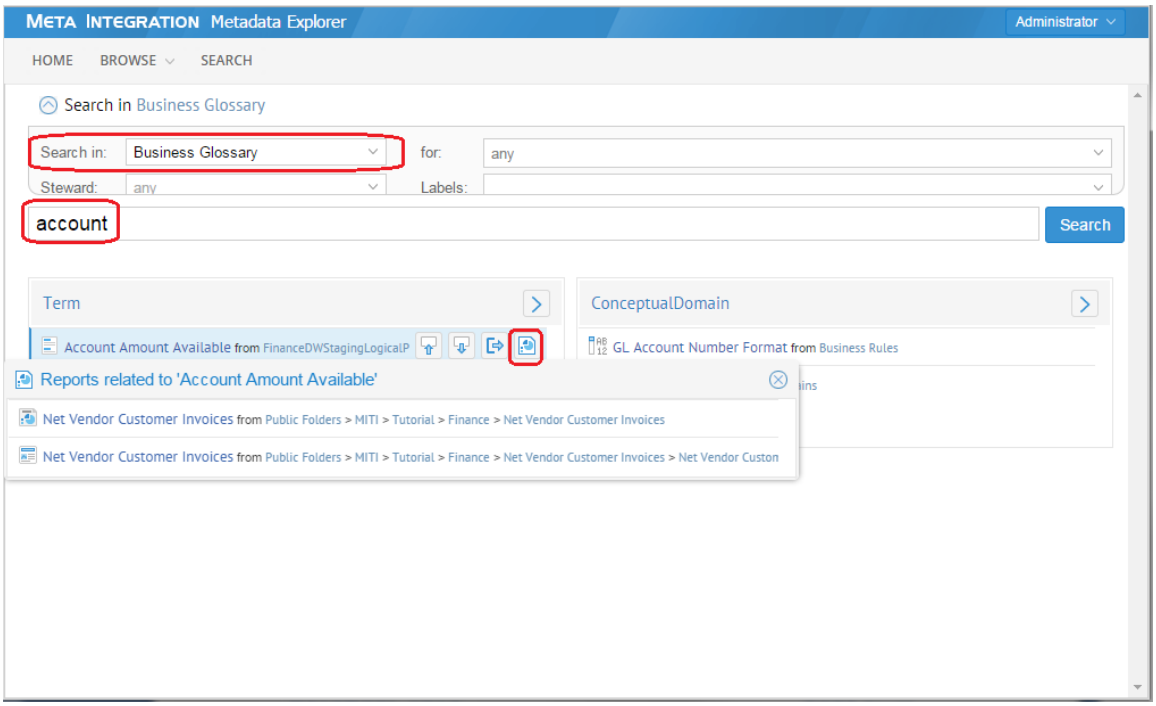


Figure 9 - Show related business reports

Once one has found a term in the business glossary with the proper meaning, an obvious question is "where is this term used". In fact, for business users, the primary request is "show me the related business reports which have a field with this definition in them."

This is the function of show related business reports. This is a very powerful use case for business users, as they are able to quickly obtain a list (and links to open) of all reports that contain information related to a given term.

By defining the semantic links properly, erwin Metadata Management (EMM) can trace down to the physical data flow to all the data elements semantically mapped to a term and then trace through the data flow to any reports which are using those data elements as fields. This function will then provide a list of all reports in what tool that make use of fields that are defined by that same term.

In this way, erwin Metadata Management (EMM) provide a catalog to ALL of your business intelligence reporting environments, independently of the underlying technology.

Please see the erwin Metadata Management (EMM) - Data Governance Tutorial for detailed examples.

3.2 Business Reporting Portal

What we have seen in this set of use cases, taken together, are what may be referred to as a *reporting portal*. From the above examples, we could see a common process for business users is to:

1. Open a report in the business glossary
2. Display the *report glossary* for that report or set of reports
3. Note the related term and the definition
4. Once there, one may then click on the hyperlink for the term
    a. Use the glossary as a reporting catalog to find other *related business reports* that are classified by that same term
    b. Navigate, browse, search through the business glossary for related terms
    c. One may also click on the hyperlink for the field to instead jump into erwin Metadata Management (EMM) details page for that element. From there one may of course
        i. Display data flow lineage back to source systems or perhaps the data lake or warehouse this information comes from
        ii. Visualize diagrams of these source or warehouse models

However, recognizing the real power of the environment, business users could in fact start directly in erwin Metadata Management (EMM) as use it as the one stop *business reporting portal* into all available terminology and concepts and thus all reporting available in the IT environment.

Commonly, then, what is done is that one customizes the erwin Metadata Management (EMM) Metadata Explorer UI homepage to provide a simple search widget into the business glossary (or multiples) that allows the user to find terminology and then, of course, related reports (simple link next to the term). One may even define the Metadata Manager UI to go directly to the report in the business intelligence tool (rather than the details page in erwin Metadata Management (EMM)) when clicking on the report name.

Please see the erwin Metadata Management (EMM) - Data Governance Tutorial for detailed examples.

3.3 On-line maintenance of the Report Glossary

The Metadata Explorer UI of erwin Metadata Management (EMM) also provides a set of features designed to support the simple contribution and maintenance of these definition lookup and report glossary results. As long as there is no workflow defined, then it becomes a very simple task to reuse and extend the terms in the business glossary, as long as one has appropriate permissions.

In this mode, users who have permissions may go to the business intelligence environment (or directly to the *business report portal*) and once they are looking at a report field details page then can click on the Glossary panel and either:

- Search for terms in the business glossary that have definitions that they can *reuse* and semantically link directly to this report field
- *Extend* the teminology by adding a term (and corresponding direct semantic link) to define the report field into the future.

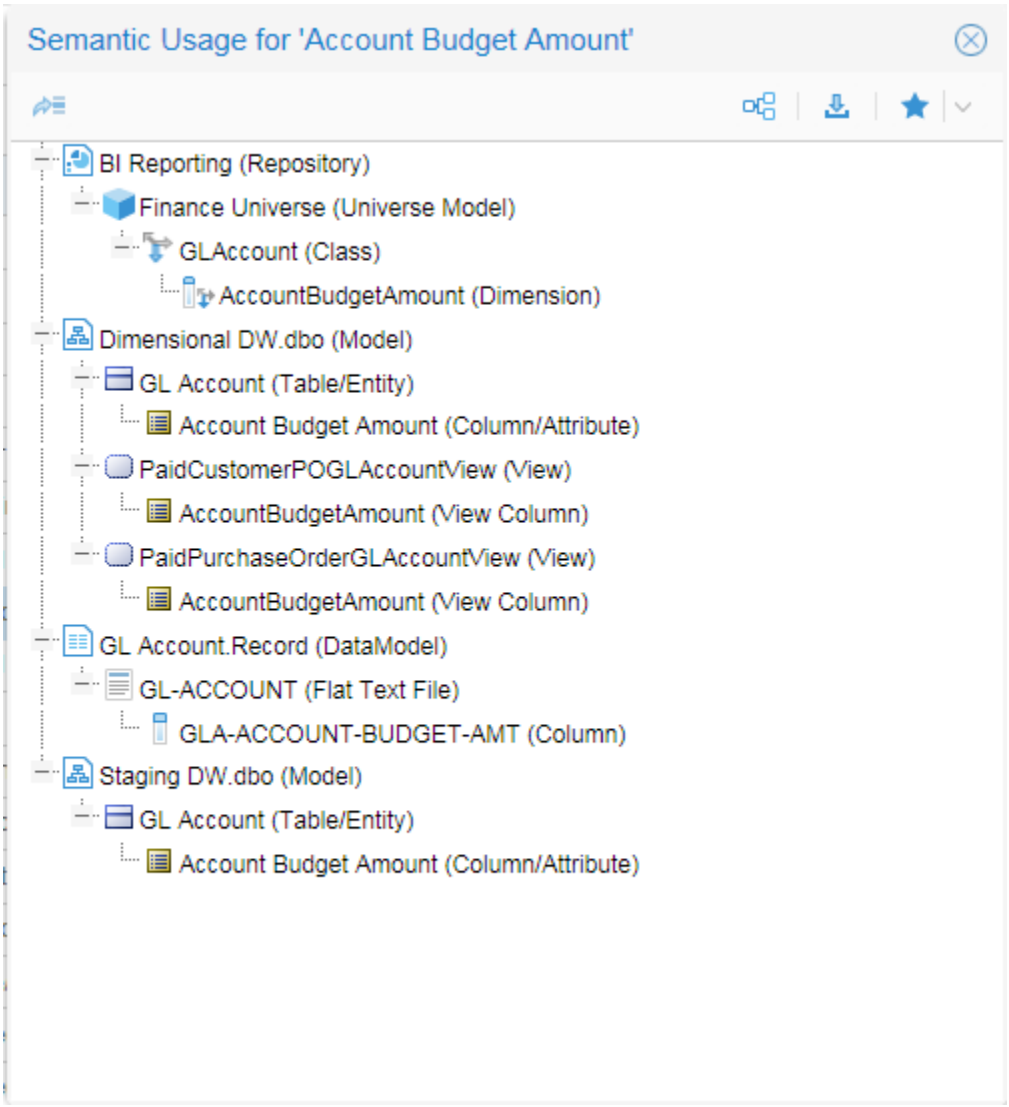
The Metadata Explorer UI is designed to make this a simple process. The only requirement (in order to see the Glossary panel in the UI) is to create and configure the semantic mapping from the business glossary to that reporting model, and add it to the configuration.

Most organizations, in this case, require that new reports be defined or *documented* in this way, either reusing the existing terminology or extending the business glossary according to the requirements of the fields in the report. Keep in mind, also, that it is quite common to semantically link to reports, groups of reports or sections of a report (graphic), and not just the fields themselves. Creating and managing terminology about reports or sections as a whole can also be quite powerful to the business user.

Please see the erwin Metadata Management (EMM) - Data Governance Tutorial for detailed examples.

3.4 Technical uses of the semantic mappings

3.4.1 Semantic usage and impact



One of the key use cases for any business glossary is to determine where a terms is "used" (has a semantic link to) by elements within the larger system configuration that is modeled in erwin Metadata Management (EMM). This result is particularly important for technical users attempting to get a handle on the cost associated with a proposed change.

By defining the semantic links properly, erwin Metadata Management (EMM) can trace down to the physical data flow to all the data elements semantically mapped to a term and then trace through forward (impact) and backward (lineage) through the data flow to see all the data elements in the system potentially impacted by a change to the term.

Please see the erwin Metadata Management (EMM) - Data Governance Tutorial for detailed examples.

3.4.2 Transitive closure

Taking advantage of both normalization (or standardization) of terms and the semantic links down into the physical architecture means that one can determine a much more complete picture of impacts due to change.

Say, the amount field in incoming transactions in your system are expanding from 8 to 10 significant digits. erwin Metadata Management (EMM) with a completely harvested and stitched configuration of your architecture can provide you with a quick answer showing you all of the fields "downstream". That simple impact analysis will not show you other parts of the system that have financial amounts that likely will need to be changed to support the larger amounts coming in, nor will it show you data elements that are related (e.g., via PK-FK) that one would likely want to change as well.

However, if you have spent the effort at normalizing all your "amount" terms into a single generalized term in the business glossary (say, "finance amount"), or if you have standardized these to a domain in the business glossary, one could:

- Perform the impact analysis
- Identify data elements in the model (like the data warehouse) where you have semantically mapped your terms to
- Look up the semantic definition of these data elements and thus the normalized term or standardized domain

Once you have that term or domain, a semantic usage report would provide a complete, architecture wide picture of what you would like wish to update to accommodate and future-proof your systems for this change. This is the result again of defining the semantic links properly, but is also payoff for your standardization efforts.

4 Best Practices Summary and Other Use Cases

Requirements by Business Glossary Feature			
Definition Lookup	**Report Lookup**	**Semantic Usage**	**Standardization**

Requires:
• Business business glossary Terms
• Either:
  • Semantic mappings to Warehouse or other strategic model
  • Direct semantic mapping to report or report field/object
  • Business Glossary Terms matching report field names
	Requires:
	• Business Glossary Terms
	• Semantic Mappings to Architecture
	and/or
		Requires:
		• Business Glossary Terms
		• Semantic Mappings to Architecture
			Requires:
			• Business Glossary Terms
			• Semantic Mappings to Architecture

Assets \ Goals	Definition Lookup	Report Lookup	Semantic Usage	Standardization
Business Business Glossary	• CSV format import • Manual Semantic linking, Data Documenter and/or term names matching report field names	• CSV format import • Manual Semantic linking, Data Documenter and/or term names matching report field names	• CSV format import • Manual Semantic linking, Data Documenter	• CSV format import • Manual Semantic linking, Data Documenter
Data Dictionary	• CSV format import • Manual Semantic linking, Data Documenter and/or term names matching report field names	• CSV format import • Manual Semantic linking, Data Documenter and/or term names matching report field names	• CSV format import • Manual Semantic linking, Data Documenter	• CSV format import • Manual Semantic linking, Data Documenter
Data Model	• Auto populate through drag and drop • Add terms with names matching report field names for those elements which are not "pass-through"	• Auto populate through drag and drop • Add terms with names matching report field names for those elements which are not "pass-through"	• Auto populate through drag and drop	• Auto populate through drag and drop
Maps well to physical DB	• Create mapping to populate business glossary	• Create mapping to populate business glossary	• Create mapping to populate business glossary	• Create mapping to populate business glossary
Does not map well to physical DB	• Data Documenter	• Data Documenter	• Data Documenter	• Data Documenter

4.1 Other Use Cases

4.1.1 Normalization / Standardization of terms

See the erwin Metadata Management (EMM) - Data Governance Tutorial for detailed examples of populating the business glossary in this manner.

4.2 Merging additional sources of terms

See the erwin Metadata Management (EMM) - Data Governance Tutorial for detailed examples of populating the business glossary in this manner.

4.3 Creating and relating domains and business rules

See the erwin Metadata Management (EMM) - Data Governance Tutorial for detailed examples of populating the business glossary in this manner.