

# Lineage Models

## Page Contents

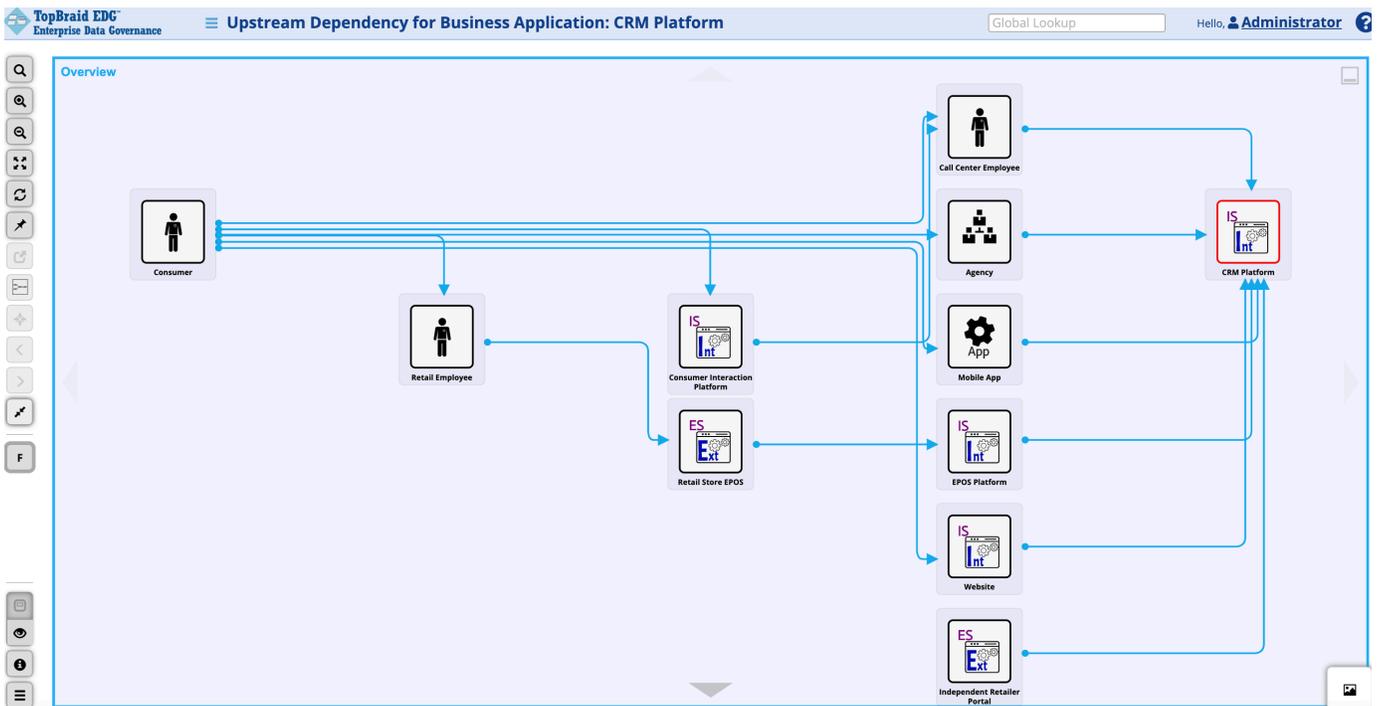
- 1 Overview of Lineage Models
- 2 Lineage Models Home
  - 2.1 Prerequisites: Licensing and Enablement
  - 2.2 Listing of Lineage Models
  - 2.3 Create New Lineage Model
    - 2.3.1 Search Indexing
    - 2.3.2 URI Construction Rules
    - 2.3.3 Includes

## Overview of Lineage Models

Data lineage describes what happens to data as it goes through diverse processes.

If assets you are governing are connected using relationships that describe data lineage, you will be able to click on a given asset and see its lineage or impact across enterprise by selecting the corresponding option from the visualization menu. Lineage will show specifics of how data flows between data sources and applications to users, in support of business activities and functions, and the enablement of enterprise capabilities. Information is presented using an interactive diagram called LineageGram.

One example of such diagram is shown below where we see business (aka logical) lineage showing all systems that feed into one organization's CRM Platform in a context of a product registration activity and associated information.



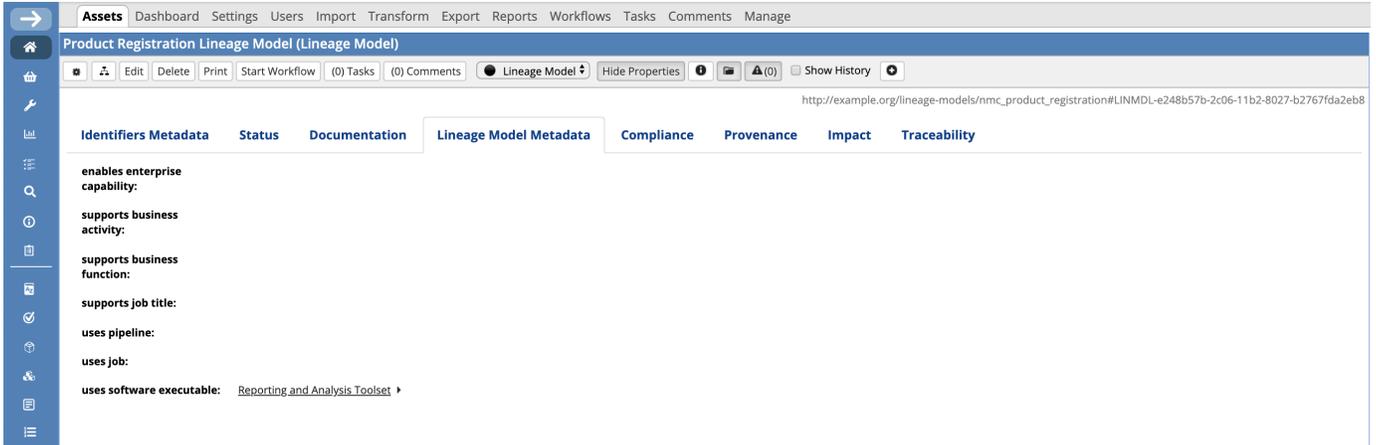
It is fairly common in an enterprise to have many different connections between individual assets where connections belong to different contexts serving different purposes. For example:

- CRM system is fed with customer information not only as a result of product registration, but also as a result of other CRM processes and activities such as marketing or customer support
- Similarly, employee information, for example, may flow between HR related applications in the context of HR processes and it may also flow between applications that are used for customer support or customer acquisition in the context of CRM processes.

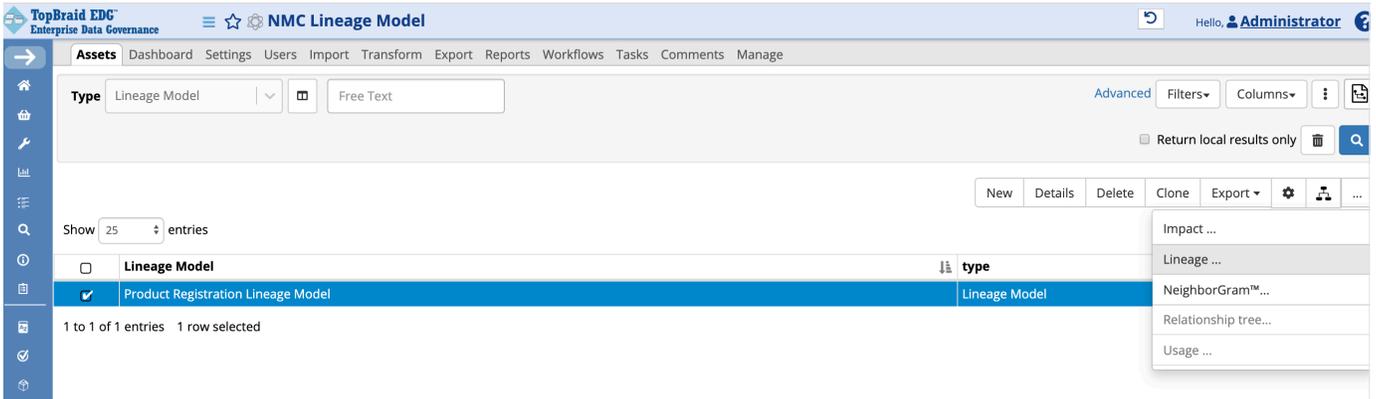
These are different enterprise flows. If they were all captured and we did not specify a context of interest, asking for a lineage of a system or a data source or a data element would display not only the dependencies in play for product registration, but also many other different links and feeds making it hard for users to understand lineage as it relates to a given business activity. Seeing this fuller picture can be important,

especially for impact analysis, and EDG provides it. However, users analyzing data lineage will often want a focused and bounded exploration of lineage. To support this, EDG lets you create Lineage Models as separate asset collections to capture contextualized lineage relationships. The role of the Lineage asset collections is to contain the context specific relationships between data, applications and other assets. Each collection can store lineage for one or more enterprise flows. It often make sense to keep lineage for related processes (e.g., HR) within the same collection.

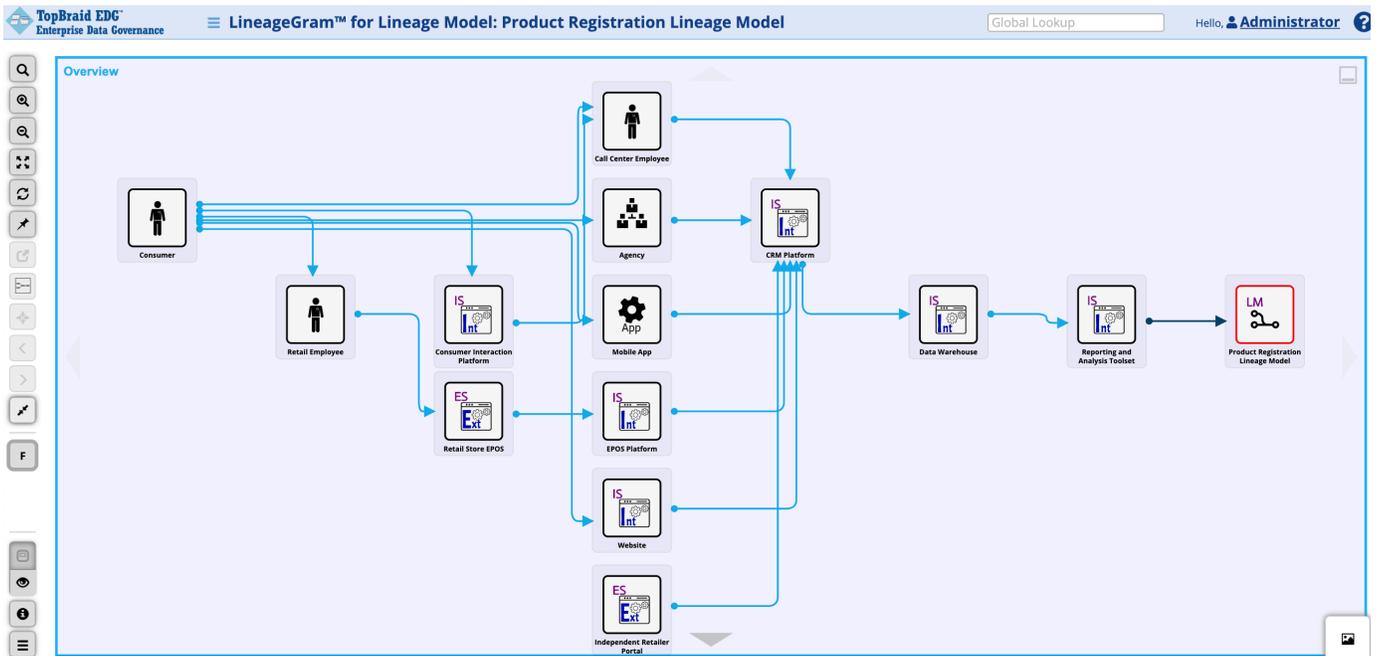
EDG also includes an asset type called Lineage Model. It shares the name with the asset collection that is intended to store lineage information, but it is an entity of its own that is created within the collection. Its use is optional and its role is to provide a convenient “starting point” for users who want to explore lineage. A Lineage Model asset is linked to assets that participate in the lineage using one of the relationships that have been designed for this purpose, for example, “uses software executable”, as is shown in the screenshot below



Only the last application or a data source in the lineage chain needs to be linked to a Lineage Model asset - as shown above. When users click on a Lineage Model asset collection, they will see the Lineage Model assets presented in a table. They can select one and choose an option to show the Lineage diagram.



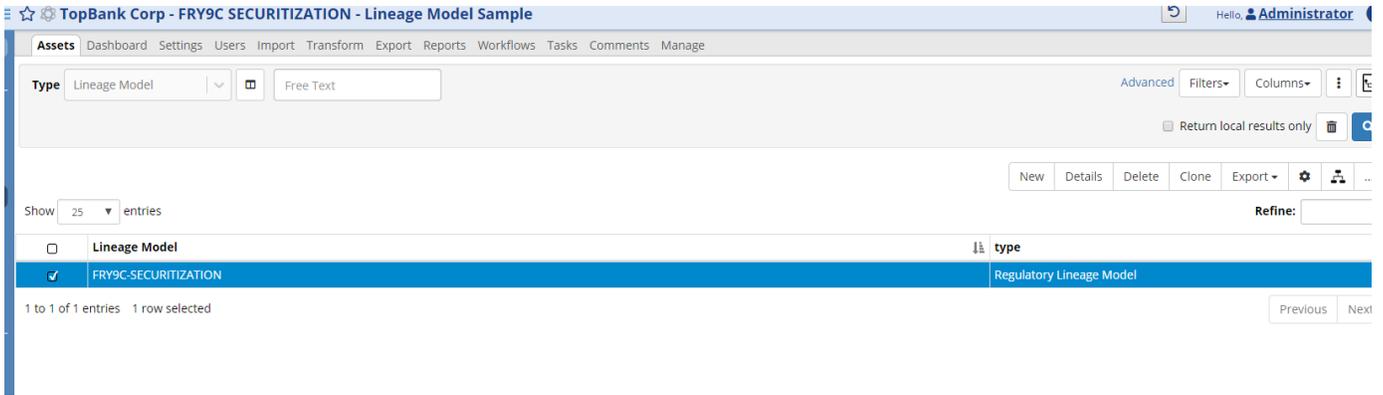
With this, we see a full flow for product registration information, from the beginning to the end. As shown below, it does not stop with the CRM Platform, but continues into a Data Warehouse and ultimately a Reporting and Analysis Toolset.



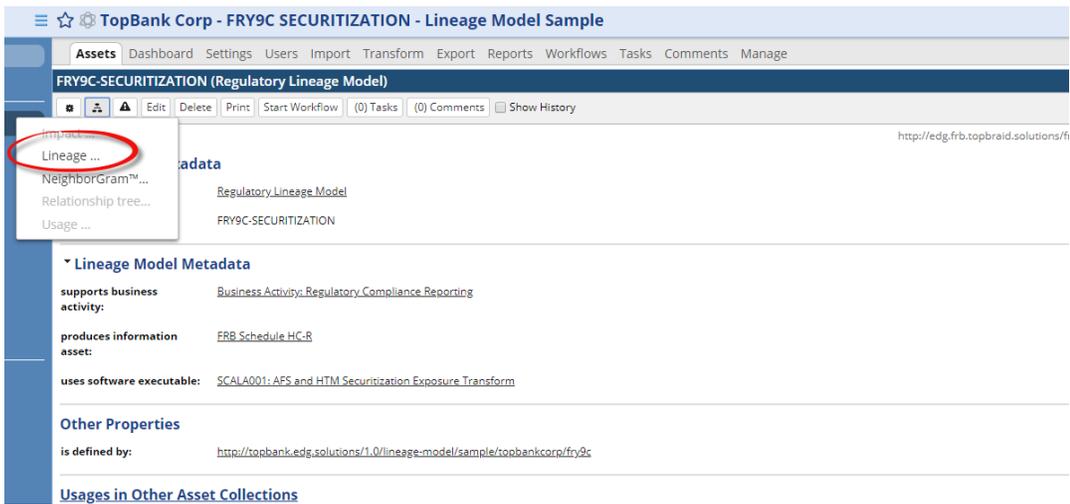
To see a sample Lineage Model in EDG, create a new Lineage Model Asset Collection, select import RDF from the Import tab and import this attached file:



Once data is loaded, click on the Assets tab and double click on the asset FRY9C-SECURITIZATION. You will see the following table.



In the visualizations menu, select Lineage:



You can also select Lineage by simply highlighting FRY9C-SECURITIZATION asset in the table and using the visualizations menu link above the table's header.

Unlike the product registration lineage shown above, FRY9C-SECURITIZATION example shows a more detailed data flow that goes beyond the business lineage and includes specific data sources and data elements. EDG can capture and connect lineage at different levels.

To learn more about the full scope of lineage supported by EDG, what kind of relationships are used to capture lineage information and to understand the capabilities of the LineageGram, click on the Lineage Model link in the blue left hand side navigation bar and navigate to the interactive tutorial by following the link in the "A tutorial explaining EDG's visualization of Lineage Models can be accessed at this page" text.

TopBraid EDG Enterprise Data Governance | Lineage Models | Global Lookup | Hello, Administrator

[Create New Lineage Model](#)

A 'Lineage Model' is a governance asset that expresses how enterprise capabilities, business functions and software systems are dependent on data flows, software jobs and data transformations across data sources and sinks. A tutorial explaining EDG's visualization of Lineage Models can be accessed at [this page](#).

## Lineage Models

Show 25 entries | Refine:

Collection Name	Subject Area	Created By	Creation Date	Last Changed By	Last Changed On	Permission	Description
<a href="#">FRY-9C</a>		Administrator	24/06/2019 02:52:31 PM			manager	
<a href="#">Im_tests</a>		Administrator	24/06/2019 03:01:44 PM			manager	
<a href="#">NMC Lineage Model</a>	NMC	Administrator	last Monday 11:13:39 AM	Administrator	last Monday 12:04:39 PM	manager	

1 to 3 of 3 entries | [Previous](#) | [Next](#)

Start New Workflow

## Lineage Models Home

Selecting the **Lineage Models** link in the left-navigation pane of TopBraid EDG lists all of the Lineage Model collections currently available to the user and, it allows authorized users to create new ones.

## Prerequisites: Licensing and Enablement

The availability of any [collection type](#) (including Lineage Models and *customer-defined* types) is determined by what is (a) licensed and (b) configured under *Server Administration*. To install a license or to view the currently licensed features, see [Setup > Product Registration](#). To configure which licensed collection types are currently enabled or disabled, see [EDG Configuration Parameters > Configure Asset Collection Types](#). For general licensing information, see the TopQuadrant website, which describes the [TopBraid products](#) and the [data governance packages](#) that determine the available collection types.

## Listing of Lineage Models

This home view shows a table with all Lineage Models that you can access in some way. For each collection some brief metadata is available in columns of the table. Columns are sortable and you can filter content of the table by typing search strings in the **Refine** field at the upper right of the table. To access an asset collection, click on its link.

To create an asset collection click on **Create New Lineage Model** button.

You can also select an asset collection in the table and start a workflow for it.

This page provides a focused view on Lineage Models. To see a view of all asset collections, irrespective of their type, that you have a governance role for, click on your **User Name** in the upper right corner of the page. To see all asset collections you have access to organized by their subject areas, click on the **Governance Areas** link in the left hand-side vertical Navigation Bar.



If a Lineage Model is either missing or it is lacking expected features in your views, you or your security role(s) may lack proper permissions for the Lineage Model. A manager of the Lineage Model can give you the needed permissions via its utilities' **Users** settings. For background information, see [Asset Collection Permissions: Viewer, Editor, and Manager](#).

Another possible cause of a missing feature is that it requires administrative setup to become active. See [EDG Administration](#) for relevant within-application settings and/or see other [EDG Administrator Guide](#) documents for relevant external installation and integration setup.

## Create New Lineage Model

The **Lineage Models > Create New Lineage Model** link opens a form with fields used to define the new Lineage Model. Note that you can also create a Lineage Model by using a Create link in the [Governance Areas](#) page.



Nobody will have a link for creating any asset collection until an administrator configures EDG's persistence technology as documented in [Server Administration: Teamwork Platform Parameters: Application data storage](#). Additionally, each user will not have a create link unless the user or their role has a *Create* permission for the EDG *Repositories* project as documented in [EDG Rights Management](#).

The **Create** dialog creates a new Lineage Model and automatically grants the Creator a Manager's permission for it. When Lineage Model creation starts from the **Governance Areas** page, new Lineage Model is automatically associated with the selected area. When Lineage Model creation starts from the Lineage Models home page, new Lineage Model is not connected to any governance area. To change this after creation, update in utilities: **Settings > Metadata > Edit > subject area**.

The **Create** dialog box asks for the Lineage Model's **Label** (name), its **Default namespace** and, optionally, a **Description**. The default namespace will be used to construct URIs (unique identifiers) for the resources in the Lineage Model. EDG will automatically pre-populate the default namespace based on the system-wide, configurable settings. Creator can change it. Recommended practice for all collection types is to use a '/' (slash) at the end of the default namespace. For ontologies, it is typical to use '#' (pound sign). However, '/' can be used as well.

<b>Label</b>	<input type="text"/>
<b>Default namespace</b>	<input type="text" value="http://example.org/edg-repositories-1.0.0-2014-01-01/new#"/>
<b>Description</b>	<input type="text"/>
<b>Data Platform</b>	<input type="checkbox"/> <i>Server-installed option to share this graph across all platform servers.</i>

## Search Indexing

If using Search the EDG with Lucene indexing (the default option), the **Create** dialog will offer an option to add this collection to the index. You can later change this setting at the **Manage** tab.

**Include this asset collection in the index for Search the EDG**

## URI Construction Rules

The **Create** dialog will also offer you an option to specify URI generation patterns for instances for each newly created graph. There are 3 different options to choose from:

New Instances Class Prefix:

New Instances Construct Method:

New Instances User Cannot Modify URI:

- **New Instances Class Prefix** has a selection of: *Default, name and acronym*. Default means that after a namespace the local name will be added. In case of name and pattern the name or pattern will be added in between the namespace and local name.

Acronym property is set for most of the asset collections in EDG, however, if you can't find acronym property in your model an extra setup needs to be done. By default ontologies don't have the `edg:acronym` property defined in a model. It requires creating a property SHAPE on an `edg:acronym` (<<http://edg.topbraid.solutions/model/acronym>> property. The shape would need to associate the property with all instances of RDFS class (in a Manage tab change Root Class of Hierarchy to `rdfs:Class`, add `edg:acronym` property to `rdfs:Class` with the appropriate SHAPE defined, then change the root class to `owl:Thing` and start adding acronym to a class for which instances will be using acronym in their URI pattern)

- **New Instances Construct Method** allows to choose from: *Default, label, uuid, counter, or custom*. In this case, the local name would be either predefined by the label, uuid or counter. The Default option will be a label. Custom option allows creation of a completely custom URI (namespace and local name).

To use custom instances construct method you'd need to provide a custom method of URI creation. It requires creating a new SWP file in IDE, importing `swa.ui.ttlx`, and overwriting '`swa:createResourceDialogUsingCustom`' class accordingly.

- **New Instances User Cannot Modify URI:** *Default, true or false* provides the option to allow modification to the URI during the creation of assets in the asset collection. By default, Create dialog for assets will allow users to modify URI

This setup can later be changed. The option is available in the **Manage tab**. This however wouldn't change the already created instance URIs, but will take effect for the new ones.

## Includes

Collections often have natural relationships to other collections (e.g., each reference dataset's *main entity* class is defined in an included ontology). Any collection using outside resources needs to *include* the collections that contain them. Some inclusions might be required while others might merely be permitted. For example:

- Taxonomies *always* include the SKOS ontology, and they *may* include other taxonomies.
- As mentioned above, each reference dataset and data graph must include at least one ontology to define the dataset's entities.
- Glossaries always include the pre-defined EDG ontology that describes business glossary terms.
- Catalogs of data assets always include the pre-defined EDG ontology describing data assets and are expected to include definitions of relevant physical datatypes.

These requirements can be further configured. When creating a collection, any required reference to another collection will either be handled automatically or be presented for selection.

**Data Asset Model to include:**

**Technical Assets Model to include:**

**Enterprise Assets Model to include:**

**Data/Technical/Enterprise Assets Models to include:** These asset models provide the contents of the information flows. If they are not selected, it is assumed that you will be storing relevant data, technical and enterprise assets directly in a lineage model.