

# Big Data Assets

## Page Contents

- 1 Overview of Big Data Assets
- 2 Big Data Assets Home
  - 2.1 Prerequisites: Licensing and Enablement
  - 2.2 Listing of Big Data Assets
  - 2.3 Create New Big Data Assets Collection
    - 2.3.1 Search Indexing
    - 2.3.2 URI Construction Rules
    - 2.3.3 Includes

## Overview of Big Data Assets

Big Data Assets specify the data structures, jobs, nodes and other software and hardware components that make up a big data ecosystem.

## Big Data Assets Home

Selecting the **Big Data Assets** link in the left-navigation pane of TopBraid EDG lists all of the Big Data Assets Collection 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 Big Data Assets 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 Big Data Assets

This home view shows a table with all Big Data Assets 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 Big Data Assets Collection** button.

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

This page provides a focused view on Big Data Assets. 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 Big Data Assets Collection is either missing or it is lacking expected features in your views, you or your security role(s) may lack proper permissions for the Big Data Assets Collection. A manager of the Big Data Assets Collection 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 Big Data Assets Collection

The **Big Data Assets > Create New Big Data Assets Collection** link opens a form with fields used to define the new Big Data Assets Collection. Note that you can also create a Big Data Assets Collection 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 Big Data Assets Collection and automatically grants the Creator a Manager's permission for it. When Big Data Assets Collection creation starts from the **Governance Areas** page, new Big Data Assets Collection is automatically associated with the selected area. When Big Data Assets Collection creation starts from the Big Data Assets home page, new Big Data Assets Collection 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 Big Data Assets Collection'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 Big Data Assets Collection. 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.

Label

Default namespace

Description

Data Platform

*Server-installed option to share this graph across all platform servers.*

## 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.

You will not be asked to provide any inclusions for a new Big Data Assets collection on create.