

Reference Datasets

Page Contents

- 1 Overview of Reference Datasets
- 2 Reference Datasets Home
 - 2.1 Prerequisites: Licensing and Enablement
 - 2.2 Listing of Reference Datasets
 - 2.3 Create New Reference Dataset
 - 2.3.1
 - 2.3.2 The ontology for the main entity (ME) class

Overview of Reference Datasets

Reference datasets contain standardized data or codes, which typically are used by various applications as lists or tables. In fact, they are often called "code tables." An individual code table may seem like a simple thing, but a well-managed collection of code tables and related reference data spread across an enterprise is a resource that can bring great value to that enterprise—or cause great problems if it is not well maintained. EDG lets you control your reference data so that you can put it to work for you as efficiently as possible.

EDG datasets are much more than just flat code tables. Reference data in different datasets can have relationships. For example, as currencies are associated with countries, currency codes have a relationship (connection) to country codes. Reference datasets can also model structural relationships in data, such as hierarchies of industrial categories, locations, or product types. Finally, you can capture any additional information you need to have about each code. And reference datasets themselves provide a lot of rich information or metadata such as the source of a dataset, how it is managed, where it is being used, and the meaning of each data field.

The tabular-editors of EDG collections (for searching, viewing, and editing assets) requires the underlying schema to be backed by SHACL. To migrate a collection's included ontologies to a SHACL basis, see: [Ontology Utilities > Convert OWL Axioms to SHACL Constraints](#).

Reference datasets no longer allow classes without a primary key property to be used as their main entity.



For additional perspectives and details on *reference data management* and related topics, see these [TopQuadrant whitepapers](#).

Reference datasets are used with *ontologies*, which define the data schema (classes, properties, relationships, constraints) of the reference dataset items. For example, you might define a class (or entity) called *Gender* in an ontology and then, in a reference dataset that uses this ontology, enter the values *Male* and *Female* as instances of this list. Ontologies thus define the data attributes for each entity and the relationships between entities.

TopBraid EDG makes it possible for you to:

- **Reduce independent maintenance of code tables:** If different departments use the same code table, they may be maintaining individual copies of it on spreadsheets being emailed around to each other. When they all use the same copy, changes are coordinated, and they can be confident that they're using the right codes.
- **Reduce data quality problems due to coding errors:** Workers who don't have access to recent, correct codes can't always enter the proper values, and improper values can lead to lost revenue.
- **Reduce the cost of designing code tables for databases:** When new code tables have similarities or other relationships to other tables, these relationships can be leveraged in the design of the new tables. Well-organized, searchable metadata about which applications use which code tables also makes it easier to coordinate new and legacy tables.
- **Reduce data integration issues due to inconsistent codes:** The inconsistencies caused by maintaining multiple copies of the same code tables, or by using copies that were updated at different times, can lead to problems when combining datasets that reference these tables. Consistent tables mean easier data integration.
- **Make informed decisions based on code table data:** Code table entries are often cryptic abbreviations, leaving people to guess about their meaning and appropriateness for which ones to use when. Metadata such as definitions and provenance information ensure that people will use the right codes in the right places.

Reference Datasets Home

Selecting the **Reference Datasets** link in the left-navigation pane of TopBraid EDG lists all of the Reference Dataset 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 Reference Datasets 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 Reference Datasets

This home view shows a table with all Reference Datasets 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 Reference Dataset** button.

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

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

The **Reference Datasets > Create New Reference Dataset** link opens a form with fields used to define the new Reference Dataset. Note that you can also create a Reference Dataset 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 box asks for the Reference Dataset's **Label** (name) and, optionally, a **Description**.

Create New Reference Dataset

This creates a new Reference Dataset with yourself as the manager.

Label

Description

Ontology to Include

Data Platform (dataset) *Server-installed option to share this graph across all platform servers.*

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

If using Search the EDG with Lucene indexing (the default option), an option exists on create to add this collection to the index. This is the same as selecting it in Search the EDG configurations with the default property selectors.

The ontology for the *main entity* (ME) class

Each reference dataset needs an ontology class to act as its *main entity*, which will be the class of the dataset's reference instances. From the existing ontologies listed for **Ontology to Include**, select the ontology that contains the class to be used as the new main entity.

After submitting the creation form, the ME class itself can be designated either via (1) the dataset's utilities: **Settings > Metadata > Edit > Overview > main entity (class) drop-down selection** or via (2) a form prompt that appears when the dataset is first edited. The main entity class must have exactly one property designated as the *primary key* - see [Ontology View or Edit: Setting a primary key for a class](#).