

Adding asset types, their assets, and their data flows into the catalog

Contents

- Managing additional information asset types and assets3
- Required skills and roles.....3
- Overview.....3
- Example Scenario5
- Defining asset type bundles6
- Registration of asset type bundles7
- Generating asset import documents and flow import documents.....8
- Importing assets and flows by using the REST API9
- Exporting assets and flows by using REST API..... 10
- Updating asset type bundles 11
- Multiple locales (languages)..... 11
- Removing registered information asset types from the catalog..... 12
- Related information 12

The information in this document is meaningful ONLY in the context of examples that are in <http://www.ibm.com/support/docview.wss?uid=swg21699130>

Managing additional information asset types and assets

You can extend the model of IBM® InfoSphere® Information Governance Catalog and add asset types and metadata from other products to give a holistic representation of those assets in the catalog. The main reason to add new asset types is to enrich your lineage reports.

Required skills and roles

The following skills are needed to register new information asset types into the model:

- Familiarity with modeling: classes, references, attributes
- Ability to write code that can do the following actions:
 - Build XMLs with ID references
 - Use RESTful APIs

You must have the Information Governance Catalog Information Asset Administrator role to work with information asset bundles in InfoSphere Information Governance Catalog.

Overview

Until now, you could use only the following means to populate and extend the catalog metadata:

- IBM InfoSphere Metadata Asset Manager to import metadata from design tools, business intelligence tools, databases, and files into the metadata repository of IBM InfoSphere Information Server.
- Bridges and connectors to import metadata from third-party tools and from databases and files.
- Importing extended data sources from a CSV file or from an ISX file by using the istool command-line utility.

These methods bring a limited number of known asset types into the catalog.

Now, when you need to bring metadata from other products into the catalog, you can dynamically register the additional asset types in the catalog. The group of asset types from a single product or tool is called an *asset type bundle*. The format of an asset type bundle is a compressed .zip file that contains:

- Descriptor XML file for the asset types
- Asset type icons
- Label property files

The bundle defines the asset types and how to display them in the catalog.

Registration is the process of calling IBM InfoSphere Information Governance Catalog REST API to add the asset types that are defined in the asset type bundle.

When you register the asset type bundle by using IBM InfoSphere Information Governance Catalog REST API, the new asset types and assets are accessible only through the catalog and not by other applications in InfoSphere Information Server.

The registered asset types with their icons, names, and their imported assets remain in the catalog even after in-place upgrades of InfoSphere Information Server and InfoSphere Information Governance Catalog. However, if you migrate your catalog to a different instance of InfoSphere Information Governance Catalog, you must register the asset bundles in the new instance.

Asset import is the process of calling InfoSphere Information Governance Catalog REST API, typically with generated XMLs, to add asset metadata that conforms to the previously registered asset types.

Flow import is the process of calling InfoSphere Information Governance Catalog REST API, typically with generated XMLs, to add data flow metadata that connects previously imported assets. The assets in such data flows can be of core asset types or of bundle-defined types alike. For the optimal user experience, the generation and import of asset XMLs and flow XMLs happen automatically, or with as little user intervention as possible.

The new assets behave mostly like other information assets in the catalog. You can assign terms, stewards, and information governance rules to the new information assets. You can browse, query, and run lineage reports on them. You can also place them in collections.

Registration of asset types offers the following advantages over extension mapping documents and extended data sources:

- Metadata of more asset types can be included in the catalog, with their specific icons and labels
- More detailed data flows can be expressed in lineage reports
- Data flows can be of types other than just “user-defined”, such as “operational”
- You do not need to wait for a new release of InfoSphere Information Governance Catalog or InfoSphere Metadata Asset Manager that can work with new asset types

Example Scenario

Bank ABCD must meet Basel 3 compliance requirements. Data lineage reports must reflect the data flows of all products that are used by the bank. XYZ is a BI product. Bank management wants to publish XYZ metadata to the catalog to comply with regulations.

Bob is an Integration Developer with knowledge of XYZ. As the Integration Developer, Bob is responsible to extract relevant metadata from XYZ and import it into the catalog.

Erin is an information architect at Bank ABCD. The information architect is responsible for the review and reporting of the Data Quality and Governance Compliance report for the different systems at Bank ABCD.

Bob and Erin take the following steps to get the metadata of their XYZ assets into the catalog:

1. Bob extends the catalog to support XYZ data sources by doing these steps:
 - a. Authors an asset type descriptor XML that defines the XYZ-specific asset types for the catalog.
 - b. Creates a bundle, which is a zip file that contains the asset type descriptor, the asset type icons, and label property files for various locales (languages or countries).
 - c. Registers the asset types that are in the zip file by using InfoSphere Information Governance Catalog REST API.
2. Bob populates the catalog with XYZ assets and data flows by doing these steps:
 - a. Develops an application that monitors XYZ, extracts the metadata of the relevant assets, and generates XMLs with that metadata, in the format that is expected by InfoSphere Information Governance Catalog REST API.
 - b. The application uploads the XYZ metadata to the catalog by using InfoSphere Information Governance Catalog REST API.
3. Erin inspects the XYZ asset types and their assets in the catalog by doing these steps:
 - a. Browses the new XYZ asset types in the Browse All window.
 - b. Inspects the assets of the new types in their Details page.
 - c. Obtains data lineage reports with data flows from, and through, the XYZ assets.

Defining asset type bundles

Construct the asset type bundle. See the files in the section “Material to construct asset type bundles” in this [technote](#) for examples of how to structure an asset bundle, with its labels and icons.

The bundle zip for each asset family is structured as follows:

```
asset_type_descriptor.xml
i18n (folder)
  labels.properties (English, default)
  labels_en.properties
  labels_es.properties
  labels_fr.properties
  ...
icons (folder)
  <class1-locale>-icon.gif      (16x16)
  <class1-locale>-bigIcon.gif  (32x32)
  <class2-locale>-icon.gif      (16x16)
  <class2-locale>-bigIcon.gif  (32x32)
```

In the file called “asset_type_descriptor.xml”, define the classes and attributes and how they are to be shown in the catalog.

For each asset, include one icon in 16 x 16 pixel and one icon in 32 x 32 pixel sizes. The icons can be in .gif or in .png formats, but the icon file name must have the suffix “.gif”.

The icons are related to asset types by convention:

- <class1-locale>-icon.gif (16x16)
- <class1-locale>-bigIcon.gif (32x32)

Example: For a class with locale="Process", there must be two icons in the 'icons' folder of the bundle:

- Process-icon.gif (16x16)
- Process-bigIcon.gif (32x32)

Add a file called “label_<language-id>.properties” for each locale for which you provide a translation for labels of your asset types. For example, “label_fr.properties” for the French language locale. Use Unicode codes for non-ASCII characters. For example: \u00fc for the German u-Umlaut.

Compress all files and icons into a zip file. Contents of this file will be uploaded in section “Registration of asset type bundles”. See the contents of the DataMass.zip file in this [technote](#) as an example of the final compressed file.

Each new asset type bundle must have a unique bundle identifier.

Tip: For easy management of compressed zip files, give the zip file the name of the product or tool. For example, you might name the zip file with asset types from XYZ to be “XYZ.zip”.

Registration of asset type bundles

1. Open two browser tabs:
 - a. InfoSphere Information Governance Catalog at <https://<host>:<port>/ibm/iis/igc/>, and log on as a user who has the Information Governance Catalog Information Asset Administrator role.
 - b. InfoSphere Information Governance Catalog REST API at <https://<host>:<port>/ibm/iis/igc-rest-explorer>

<host>:<port> is the address and port number where InfoSphere Information Governance Catalog is installed.
2. List asset type bundles that are currently registered in the catalog by doing these steps in InfoSphere Information Governance Catalog REST API:
 - a. Expand **bundles: Asset bundle management**.
 - b. Expand **GET /bundles/**.
 - c. Click **Try it out!** IDs of asset bundles that are currently registered in the catalog are listed in the Response Body section.
3. Register each new asset type bundle by doing these steps in InfoSphere Information Governance Catalog REST API:
 - a. Expand **bundles: Asset bundle management**.
 - b. Expand **POST /bundles/**.
 - c. In the **File** field, browse to the zip file that you constructed, and then click **Try it out!**

If the Response Body section shows "[]" and the Response Code is 200, then the assets were imported into the catalog successfully. Otherwise, follow the instructions in the error messages such as correcting the bundle descriptor and then trying to register again. Alternatively, you might need to first unregister an already registered asset type bundle (see section "Removing registered information asset types from the catalog".)
4. Inspect the newly registered asset types in the catalog by doing these steps in InfoSphere Information Governance Catalog:
 - a. In the **Catalog** tab, refresh (F5).
 - b. Click **Information Assets**.
 - c. Click **Browse All**. Scroll down until you see the newly registered asset types.
 - d. Note that there are no assets yet of the new types, when you click on them.

Generating asset import documents and flow import documents

You can learn about the format of asset import documents and flow import documents from these documented examples:

- documented-ASSET-import-sample1.xml
- documented-FLOW-import-sample1.xml

Also, from the more compact examples:

- Job-assetdoc.xml
- Job-design-flowdoc.xml

You can use the XML generation technology of your choice. Some options are:

- JAXB, Castor
- DOM
- XML templates
- XSLT (applicable when your tool already outputs the relevant metadata in a structurally similar XML format)

Importing assets and flows by using the REST API

For the optimal user experience, the generation and import of asset XMLs and flow XMLs will happen automatically, or with as little user intervention as possible. Therefore, the steps below are just for experimenting. In a real deployment, they will typically be done using programmatic REST API calls.

1. Populate the newly registered asset types with assets by doing these steps in InfoSphere Information Governance Catalog REST API:
 - a. Expand **bundles: Asset bundle management**.
 - b. Expand **POST /bundles/assets**.
 - c. In the **String** field, paste the XML content of an **asset** document (use Job-assetdoc.xml or documented-ASSET-import-sample1.xml as examples), and then click **Try it out!**

If the Response Code is 200, then the assets were imported into the catalog successfully. Otherwise, follow the instructions in the error messages, and then repeat this step.

2. Inspect the new assets by doing any of the following actions in InfoSphere Information Governance Catalog:
 - o Browse to the assets of the new asset type by doing these steps:
 - a. Go to the **Catalog** tab.
 - b. Click **Information Assets**.
 - c. Click **Browse All**, and then scroll down to the new asset types. You can narrow down the type list by typing ahead, in the text field on the top.
 - d. Click an asset type to list its assets. Note the newly imported assets.
 - o Query the assets of the new asset type by doing these steps:
 - a. Go to the **Catalog** tab.
 - b. Click **Queries**.
 - c. Click **New**.
 - d. Select one of the new asset types, select some of its properties, and then click **Run**.

3. Upload the data flows of the new assets by doing these steps in InfoSphere Information Governance Catalog REST API:
 - a. Expand **flows: Lineage flow graph operations**.
 - b. Expand **POST /flow/upload**.
 - c. In the **String** field, paste the XML content of a **flow** document (use Job-design-flowdoc.xml or documented-FLOW-import-sample1.xml as examples), and then click **Try it out!**

If the Response Code is 200, then the flows were imported into the catalog successfully. Otherwise, follow the instructions in the error messages, and then repeat this step.

4. Inspect the data flows by running a lineage report on a previously imported asset that participates in the uploaded flows. For example, run lineage on the imported job “AddressFormatter” or “EmailJoiner”.

In the Details page of the previously imported asset, expand **Usage Information**, and then check that the participating assets are correct. Note the virtual assets.

Exporting assets and flows by using REST API

You can export registered asset types by using the **GET/bundles/assets** command. By default, whole asset family is exported. You can also specify classes to export only selected assets from a family.

Exporting asset family

1. By using **GET/bundles** command, look for the asset family name that you want to export.
2. In the **family** parameter, specify the name that you found in the previous step.
3. Click **Try it out!**
4. Save the .xml file with the results.

Exporting selected assets

1. By using **GET/bundles** command, look for the asset family name that you want to export.
2. In the **family** parameter, specify the name that you found in the previous step.
3. By using **GET/types** command, look for asset type ids that you want to export. Look for the values of the "_id" parameter.
4. In the **classes** parameter, specify the values that you found in the previous step. The values must be separated by a comma and a space.
5. Click **Try it out!**
6. Save the .xml file with the results.

Exporting data flows (supported since 11.7.1, and 11.5.0.2 Service Pack 6)

1. Expand flows: Lineage flow graph operations.
2. Expand GET /flows/.
3. In the assets field, specify the RID of a bundle asset for which you want to export flows.
4. Click **Try it out!**
5. Save the .xml file with the results.

After you save the exported file, you can edit its content and re-import it by using the **POST/bundles/assets** command. The following rules apply when you re-import assets:

- If there are no conflicts between the existing and imported asset values, the content of the imported file is merged with the existing values.
- If there is a conflict between the existing and imported asset values, the imported values overwrite the existing values.
- If you want to import references to terms, labels, or information governance rules, and such assets do not exist in Information Governance Catalog, they are created during the import process.

Updating asset type bundles

You can update registered asset types such that the imported assets and flows are preserved, as long as there are no structurally destructive changes to the previously registered asset types. This is useful when you need to add types or attributes, or change the way they are presented in the catalog.

Do these steps in InfoSphere Information Governance Catalog REST API:

- a. Expand **bundles: Asset bundle management**.
- b. Expand **PUT /bundles/**.
- c. In the **File** field, browse to the zip file, and then click **Try it out!**

If the Response Body section shows "[]" and the Response Code is 200, then the asset type bundle was imported into the catalog successfully. Otherwise, follow the instructions in the error messages, and then repeat this step.

Verify that your changes in the updated registered asset type bundle get reflected in InfoSphere Information Governance Catalog.

Multiple locales (languages)

If you need your assets to display in additional locales (languages, translations), then the bundle zip file must have a label property file for each locale, in the **i18n** folder. For example:

- labels_fr.properties for French
- labels_es.properties for Spanish

You can start out with just the default locale and add more label property files later on. Then, you can update the bundle in the catalog whenever you have additional translations.

When the catalog displays labels for asset types defined in bundles, it looks for the label property file for the language that the browser is using. If it cannot find a properties file for that language, it uses whatever labels are defined in the default locale.

Removing registered information asset types from the catalog

You can remove registered asset type bundles from the catalog. When you remove a bundle from the catalog, all of its asset types, assets, and data flows are removed from the catalog. Links to other assets, stewards, collections, and so forth are also removed.

Do these steps in IBM® InfoSphere® Information Governance Catalog REST API:

1. Expand **bundles: Asset bundle management**.
2. Expand **GET/bundles/**.
3. Click **Try it out!** to get a list of all bundle IDs whose asset types are registered in the catalog.
4. Expand **DELETE/bundles/{id}**.
5. In the **ID** field, type in the bundle ID whose asset types you want to remove from the catalog.
6. Click **Try it out!**
7. Refresh your browser window by pressing **F5**.

If the Response Body section shows “[]” and the Response Code is 200, then the registered asset types were removed from the catalog successfully. Otherwise, follow the instructions in the error messages.

8. Check that the registered asset types are not in the catalog. Do these steps in IBM InfoSphere Information Governance Catalog:
 - a. In the **Catalog tab, refresh (F5)**.
 - b. Click **Information Assets**.
 - c. Click **Browse All**.
 - d. Check that the asset types are no longer there.

Related information

For further information, see:

- [Sample REST API calls and use case descriptions](#)
- [Tips, tricks, and time-savers](#)