

# Introduction to Watson Knowledge Catalog

## Introduction

This lab will introduce Watson Knowledge Catalog. Watson Knowledge Catalog is a secure enterprise catalog to discover, catalog and govern your data/models with greater efficiency. The catalog is underpinned by a central repository of metadata describing all the information managed by the platform. Users will be able to securely share data with their colleagues more easily, regardless of what the data is, where it is stored, or how they intend to use it. In this way, the intelligent asset catalog will unlock the value held within that data across user groups helping organizations use this key asset to its full potential.

## End-to-End Data Science

The general flow of the End to End Data Science PoT will be guided by the activities shown in Figure 1- End to End Flow. This lab will focus on the **Find Governed Data** activity and demonstrate **Connect to Data**.

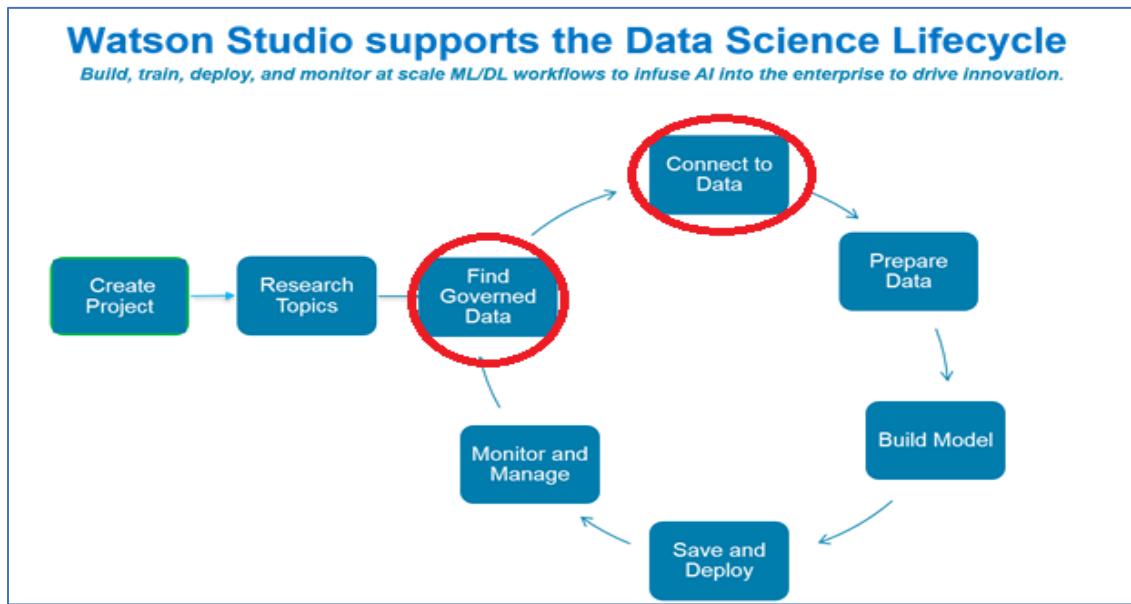


Figure 1- End to End Flow

## Objectives

The goal of the lab is for the users to gain familiarity with the features of the Watson Knowledge Catalog. We will perform the following catalog tasks:

1. Create a governed catalog
2. Add a member to the catalog
3. Add Data Assets to the catalog
4. Search the catalog
5. Edit/Review/Profile a structured Data Asset
6. Demonstrate access control features

7. Demonstrate policy creation and enforcement (optional)
8. Push the Data Assets to the project set up in Lab-1

## Download Files

### Step 1: Download the trafficking\_data.zip from the github repository

1. Click [here](#), to download the zipped file.
2. Extract the file contents. You should have two files extracted, (1) Categories.csv, and (2) Occupation.csv.

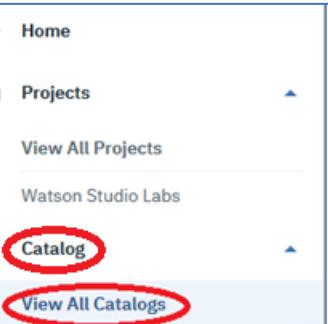
## Perform Catalog Tasks

### Step 1 – Create a Governed Catalog

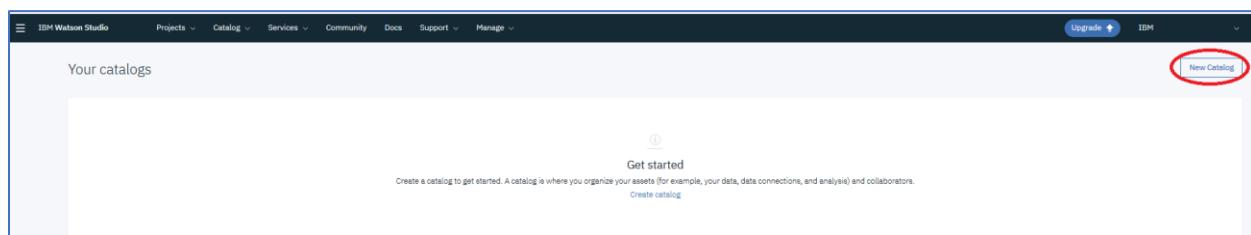
1. Click on the **IBM Watson Studio** browser tab to return to Watson Studio.
2. Click on **Watson Studio** to go to the home screen.



3. Click on  icon, click on **Catalog**, and then **View All Catalogs**.



4. Click on **New Catalog**.



5. Enter **Watson Studio Labs** for the **Catalog name**. Click on **Permanently enforce data policies**. You will get a pop-up as shown below. Click on **OK**.

New catalog

Details

Name\* Watson Studio Labs

Description

IBM Cloud Object Storage ⓘ  
This is for storing files that are associated with assets.

Object storage instance\* cloud-object-storage-gu

Or click here to create an additional instance. Click here to refresh the instances.

Can't be changed later.  
Can't be changed later. Permanently enforce or ignore policy for this catalog.

Permanently enforce data policies ?

Are you sure you want to enforce data policies for this catalog? You cannot disable policies enforcement after you create the catalog.

**Note:** By default, access to data assets in a catalog are only restricted by the privacy settings of the data assets. Privacy settings and policy rules can limit which members of the catalog can view and use the assets. You can implement data policies to restrict access to data based on the contents of the data. Data policies help you control data access and ensure that the right people can access the right data. Selecting the option to **enforce data policies** enables the enforcement of data policy rules to allow or deny access to a data asset or mask, anonymize and redact data at the data asset field level.

Setting this option for a catalog is a good best practice. Once it is enabled it cannot be undone, but it does not restrict or impede any functionality, it provides additional security measures to protect data assets.

#### 6. Click **Create** to create the catalog.

New catalog

Details

Name\* Watson Studio Labs

Description

IBM Cloud Object Storage ⓘ  
This is for storing files that are associated with assets.

Object storage instance\* cloud-object-storage-gu

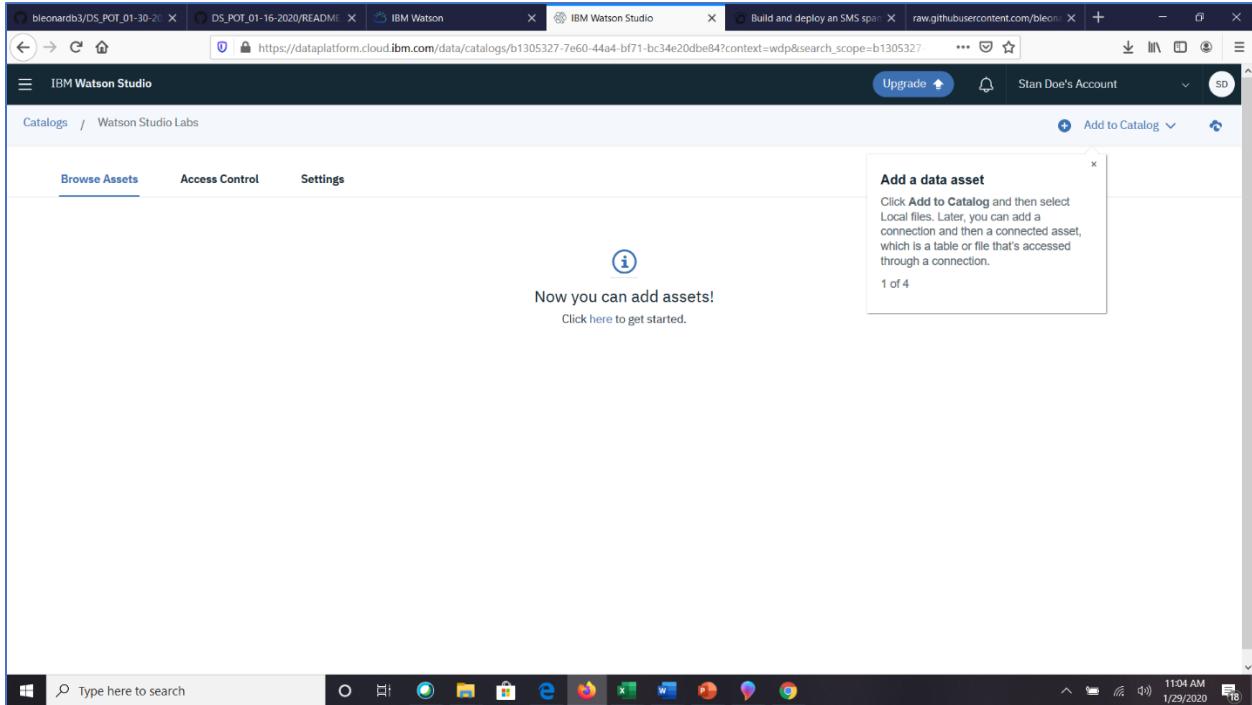
Or click here to create an additional instance. Click here to refresh the instances.

Can't be changed later.  
Can't be changed later. Permanently enforce or ignore policy for this catalog.

Permanently enforce data policies

Enforce data policies

## 7. Now you are ready to add assets!



## Step 2: Add a Member to the Catalog

To demonstrate the Catalog access control capability, we will add user ws.catalog.user@gmail.com as a member of the catalog. Note that the name of this user is “Howard Doe”.

### 1. Click on Access Control.

Name	Email	Role	Date added
George Doe	wsuser42000@gmail.com	Admin	Apr 25, 2020

### 2. Click on Add Collaborator.

3. Leave Viewer as the **Role**. Enter “Howard”. The list should populate with “Howard Doe” and the [ws.catalog.user@gmail.com](mailto:ws.catalog.user@gmail.com) e-mail address.

Add collaborators as Viewers

Having trouble adding collaborators? ⓘ X

Role:

Admin  Editor  Viewer

Collaborators (0)

Howard

Howard Doe ws.catalog.user@gmail.com

4. Click on the **Howard Doe** entry in the list box.

## Add collaborators as Viewers

Having trouble adding collaborators? [\(i\)](#)

Role:

Admin  Editor  Viewer

Collaborators (0)

Howard

Howard Doe

ws.catalog.user@gmail.com

5. Click **Add**.

## Add collaborators as Viewers

Having trouble adding collaborators? [\(i\)](#)

Role:

Admin  Editor  Viewer

Collaborators (1)

Howard Doe 

Cancel

Add

6. Howard Doe is added as a collaborator on the Watson Studio Labs catalog.

The screenshot shows the Watson Studio Labs catalog interface. The 'Access control' tab is selected. A table lists 'Collaborators' with columns for Name, Email, Role, and Date added. Two entries are shown: George Doe (Admin, April 25, 2020) and Howard Doe (Viewer, April 25, 2020). The Howard Doe row is highlighted with a red border.

## Step 3: Add Data Assets to the Catalog

In this step, we will add 4 assets; two local files, a connection, and a connected data asset. The local files we will add are the “Occupation.csv” file, and the “Categories.csv” file. We will set up a connection to a DB2 Warehouse service. We will add a connected data asset that will point to the FEMALE\_HUMAN\_TRAFFICKING table in the DB2 Warehouse service instance.

This step will introduce you to the three methods available to discover and catalog data assets; **Local files, Connected asset and Connection**. You will use these methods to catalog data assets into the newly created Knowledge Catalog and then tag them for users to easily find them, understand their content and make them available throughout IBM Watson Studio, for use during data preparation and within models, dashboards and notebooks as we will see in our lab exercises.

### Add Local Files

1. Click on **Add to Catalog** and click on **Local Files**.

The screenshot shows the Watson Studio catalog interface with the 'Add to Catalog' button highlighted. A dropdown menu is open, showing three options: 'Local files' (highlighted with a red circle), 'Connected asset', and 'Connection'.

2. Click on **browse** to navigate to the folder where the “Occupation.csv” file is stored.

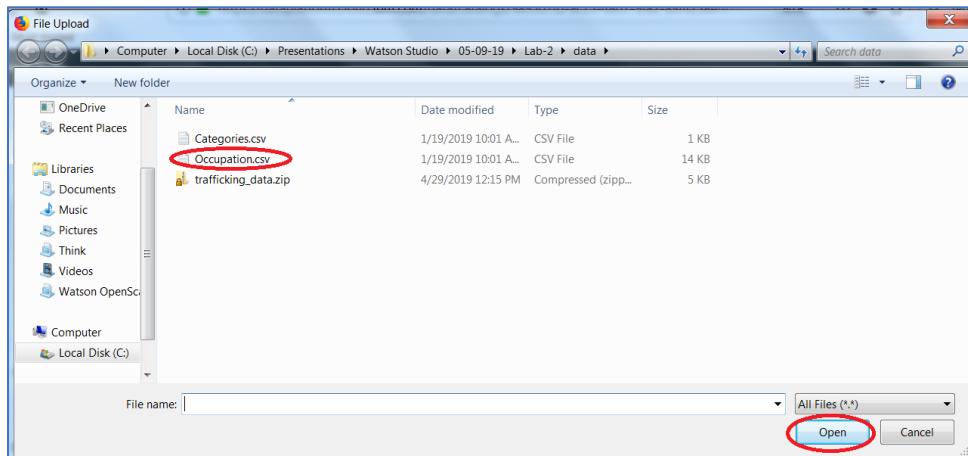
## Add data assets from local files

### Select file(s)

Drop your files here or browse your files to add new files (up to 5 GB each)  
Files for assets are saved in the catalog's associated storage.

Drag and drop file(s) here to upload

3. Click on **Occupation.csv** and then click **Open**.



4. We can add tags to the newly added data asset to more easily search for it. Add two tags, enter “occupations”, click on the + icon, and enter “jobs” and click on the + icon. Leave the **Classification** as None. Click on Private for the **Privacy**. Click on **Add**.

Catalogs / Watson Studio Labs

**Selected files (1)\***

Asset name	Format
Occupation.csv	Microsoft Excel

**Business terms**  
Search business terms

**Tags**  
**occupations** **jobs**

**Classification\***  
None

**Privacy**  
 Public  Private  
Only the asset owner and asset members can find or use the asset.

**Members**

⚠ Stay in the catalog until loading is complete! If you leave the catalog, the incomplete asset will be deleted.

Add

- The Occupation.csv asset metadata is displayed. Note a number of features on the panel. The top part of the page contains the search area. This consists of an entry field to type search text and two filters, one by data type, and one by tags (filters shown in red below). Underneath the search area, are 3 tabs: **Watson Recommended**, **Highly Rated**, and **Recently added**. The square below contains assets that are displayed when one of these tabs are selected (shown in blue below). Note this area can be collapsed. At the bottom is the list of assets (shown in green).

Catalogs / Watson Studio Labs

Browse assets Access control Settings

What assets are you looking for?

Any type Any tag Clear all

Watson recommends Highly rated Recently added

Data asset  
**Occupation.csv**

Owner: George Doe  
Added: Apr 26, 2020 9:37 AM  
Tags: occupations, jobs  
☆☆☆☆☆ 0 reviews

Showing 1 of 1 item

Name	Owner	Tags	Business terms	Type	Date added
Occupation.csv	George Doe	occupations, jobs		Data asset	Apr 26, 2020

- Follow the same procedure (1-4) above to add the local file “Categories.csv” file to the catalog. For step 4, enter the tag “job types” and then click on the “+” icon, enter the tag “job categories” and then click on the “+” icon, leave the **Classification** as None, leave the **Privacy** as Public, and then click on **Add**.

Catalogs / Watson Studio Labs

Selected files (1)\* Edit name and format [C](#)

Asset name	Format
Categories.csv	Microsoft Excel

**Business terms**  
Search business terms

**Tags**  
[job types](#) [job categories](#)

Start typing to add values

**Classification\*** ⓘ  
[None](#)

**Privacy**  
 [Public](#)  [Private](#)  
All catalog members can find and use the asset.

**Members**

⚠ Stay in the catalog until loading is complete! If you leave the catalog, the incomplete asset will be deleted.

[Cancel](#) [Add](#)

## 7. The Categories.csv asset is displayed

Catalogs / Watson Studio Labs

Browse assets Access control Settings

What assets are you looking for?

Any type Any tag Clear all

Watson recommends Highly rated Recently added

Categories.csv

Owner: George Doe  
Added: Apr 26, 2020 9:56 AM  
Tags: job types, job categories  
0 reviews

Occupation.csv

Owner: George Doe  
Added: Apr 26, 2020 9:37 AM  
Tags: occupations, jobs  
0 reviews

Showing 2 of 2 items

Name	Owner	Tags	Business terms	Type	Date added
Categories.csv	George Doe	job types, job categories		Data asset	Apr 26, 2020
Occupation.csv	George Doe	occupations, jobs		Data asset	Apr 26, 2020

## 8. Change the name of the Occupation.csv asset to be just Occupation by clicking on the Occupation.csv asset.

Showing 2 of 2 items

Name	Owner	Tags	Business terms	Type	Date added
Categories.csv	George Doe	job types, job categories		Data asset	Apr 26, 2020
Occupation.csv	George Doe	occupations, jobs		Data asset	Apr 26, 2020

9. Hover the mouse on the **Occupation.csv**. Click the “pencil” icon when it appears.

The screenshot shows the Watson Studio Labs catalog interface. A data asset named "Occupation.csv" is selected. The "Overview" tab is active, displaying the following information:

- Description:** There is no description available for this asset.
- Added:** Apr 26, 2020 9:37 AM
- Format:** Microsoft Excel
- Size:** 30 KB
- Business terms:** There are no terms available for this asset.
- Tags:** jobs, occupations
- Reviews:** 0 reviews
- Connection:** Recruitment consultant

The "Lineage" section shows a schema with 2 columns and 492 rows, including a preview of 492 rows. The data table lists various occupation names and their counts:

OCCUPATION	Code
Text	Code
Chiropodist	6
Engineer, structural	2
Engineer, land	2
Careers adviser	15
Geneticist, molecular	7
Electronics engineer	2
Editor, commissioning	4
Writer	4
Recruitment consultant	15

10. Change the name to Occupation and click **Apply**.

The screenshot shows the Watson Studio Labs catalog interface. The asset name "Occupation.csv" is highlighted in the breadcrumb navigation bar. A modal dialog box is open, allowing the user to change the name to "Occupation". The "Apply" button is circled in red.

11. Click on **Watson Studio Labs** to return to the catalog display.

The screenshot shows the Watson Studio Labs catalog interface. The breadcrumb navigation bar now shows "Watson Studio Labs / Occupation.csv", indicating the name has been successfully changed. The asset details page for "Occupation" is displayed, showing the new name in the title bar.

12. Follow steps 8-10 to change the Categories.csv asset name to be Categories.

## Add Connection

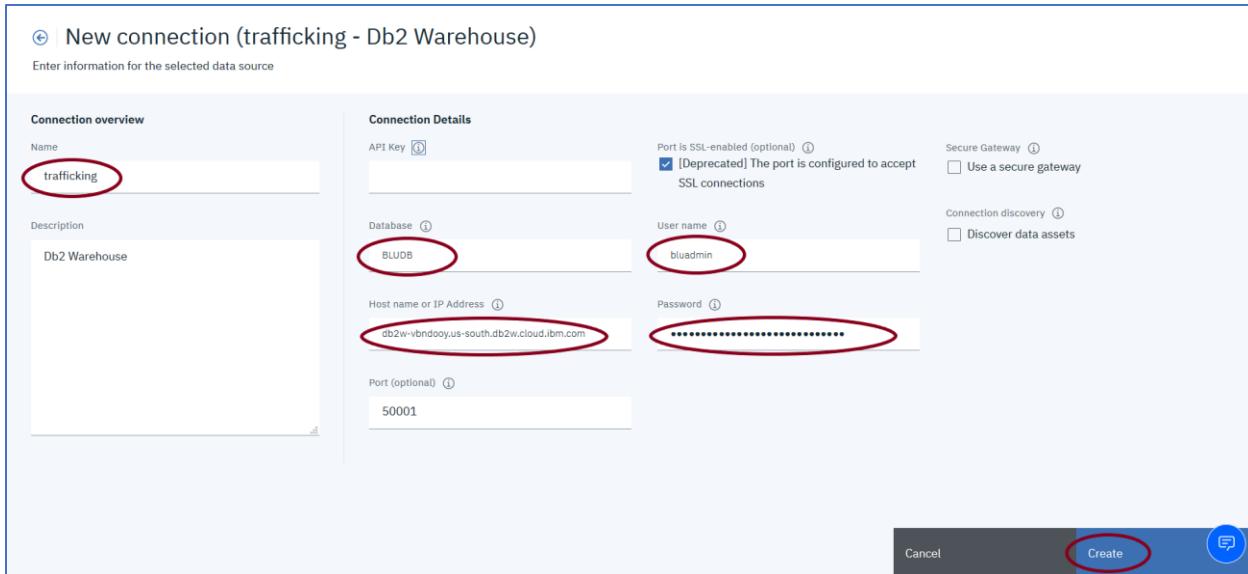
1. Click on **Add to Catalog**, and then click on **Connection**.

The screenshot shows the IBM Watson Studio interface. At the top, there's a navigation bar with 'IBM Watson Studio' and an 'Upgrade' button. Below it, a breadcrumb path says 'Catalogs / Watson Studio Labs'. On the right side, there's a vertical menu with options: 'Add to Catalog', 'Local files', 'Connected asset', and 'Connection'. The 'Connection' option is circled with a red oval.

2. The Connection page is displayed with the various connectivity options. Click on **DB2 Warehouse**.

The screenshot shows the 'New connection' page. It has a heading 'Select a data source to begin'. Under 'IBM services', there's a section for 'Db2' which includes 'Db2 for i', 'Db2 on Cloud', and 'Db2 Warehouse'. The 'Db2 Warehouse' icon is circled with a red oval. Other services listed include Analytics Engine HDFS, Cognos Analytics, Planning Analytics, Cloud Object Storage, Compose for MySQL, Db2 for z/OS, Informix, Cloudant, Db2, Db2 Hosted, and Netezza (PureData System for Analytics). Under 'Third-party services', there's a list of various databases and data stores like Amazon Redshift, Cloudera Impala, Google Cloud Storage, Microsoft SQL Server, Pivotal Greenplum, Snowflake, Amazon S3, Dropbox, Looker, MySQL, PostgreSQL, Sybase, Apache HDFS, FTP, Microsoft Azure Data Lake Store, OData, Salesforce.com, Sybase IQ, Apache Hive, Google BigQuery, Microsoft Azure SQL Database, Oracle, SAP OData, and Tableau. At the bottom right, there are 'Cancel' and 'Create' buttons, with the 'Create' button having a blue speech bubble icon next to it.

3. Enter “trafficking” for the **Name** of the connection. Leave the **API key** blank. Enter BLUDB for the **Database**. Enter bluadmin for the **User name**. Click [here](#) to obtain the DB2 Warehouse Credentials that will be needed for the **Host name** and the **Password**. Pick a random database connection from the list. Cut and paste the values for hostname and password into the corresponding entry areas on the screen. Click on **Create**.

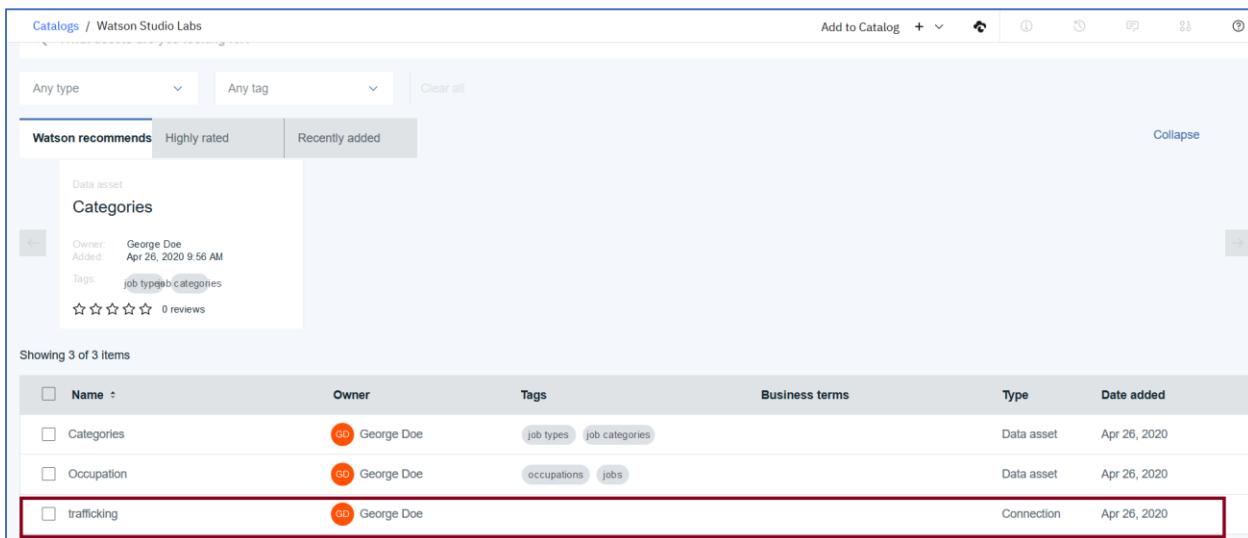
 | New connection (trafficking - Db2 Warehouse)

Enter information for the selected data source

<b>Connection overview</b>	<b>Connection Details</b>
Name <b>trafficking</b>	API Key 
Description Db2 Warehouse	Port is SSL-enabled (optional)  <input checked="" type="checkbox"/> [Deprecated] The port is configured to accept SSL connections
	Secure Gateway  <input type="checkbox"/> Use a secure gateway
	Connection discovery  <input type="checkbox"/> Discover data assets
Host name or IP Address 	User name 
db2w-vbndooy.us-south.db2w.cloud.ibm.com	bluadmin
Port (optional) 	Password 
50001	*****

Cancel **Create** 

4. The connection asset is displayed.



Catalogs / Watson Studio Labs

Add to Catalog + ⚙ ⓘ ⏷ ⏸

Any type Any tag Clear all

Watson recommends Highly rated Recently added

Categories

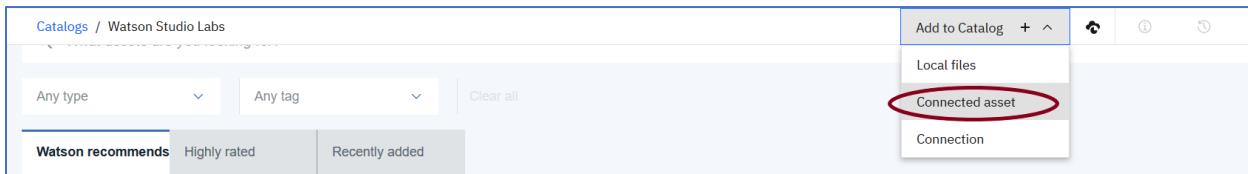
Owner: George Doe  
Added: Apr 26, 2020 9:56 AM  
Tags: job type, job categories  
0 reviews

Showing 3 of 3 items

<input type="checkbox"/> Name	Owner	Tags	Business terms	Type	Date added
<input type="checkbox"/> Categories	 George Doe	job types, job categories		Data asset	Apr 26, 2020
<input type="checkbox"/> Occupation	 George Doe	occupations, jobs		Data asset	Apr 26, 2020
<input type="checkbox"/> trafficking	 George Doe			Connection	Apr 26, 2020

## Add Connected Data

- We have added a connection to the DB2 Warehouse service. We will now add a Connected Data Asset to point to the FEMALE\_HUMAN\_TRAFFICKING table in the warehouse. Click on **Add to Catalog** and click on **Connected asset**.



Catalogs / Watson Studio Labs

Add to Catalog + ⚙ ⓘ ⏷ ⏸

Any type Any tag Clear all

Watson recommends Highly rated Recently added

Connected asset

Connection

2. Click on **Select Source**.

Add asset from connection

<b>Source*</b> <input style="border: 1px solid #ccc; border-radius: 5px; padding: 2px 10px; width: 100%;" type="button" value="Select source"/>	<b>Name*</b> <input type="text" value="Asset name"/> <input type="text" value="Description"/>
--	---

- Click on the **trafficking** connection, click on the **TRAFFICKING Schema name** click on **FEMALE\_HUMAN\_TRAFFICKING**, and then click **Select**.

Catalogs / Watson Studio Labs

Connections	trafficking	TRAFFICKING
Connections (1)	Schemas (7)	Tables (1)
<b>trafficking</b>	> AUDIT BLUADMIN DB2INST1 IBM_SAILFISH NULLIDR1 NULLIDRA <b>TRAFFICKING</b>	> <b>FEMALE_HUMAN_TRAFFICKING</b>

**Select**

- Enter female\_human\_trafficking for **Name**, enter “trafficking” for **Tags**, click the “+” icon, enter “female human trafficking for **Tags**, click the “+” icon. Leave **Classification** as None. Scroll down as needed to click on Private, for **Privacy**, and click on + **icon** to the right of **Members**. The privacy setting of “Private” specifies that only catalog members on the access control list for this data asset will have access to the data asset.

Catalogs / Watson Studio Labs

Source Db2 Warehouse Change source

Name\* female\_human\_trafficking

Description

Business terms Search business terms

Tags trafficking female human traff... Start typing to add values

Classification\* None

Privacy Public Private Only the asset owner and members can find or use the asset.

Members

Cancel Add

5. Enter Howard in the **Members** field. The list of matching users is displayed in the member list. Click on **Howard Doe**.

Add members

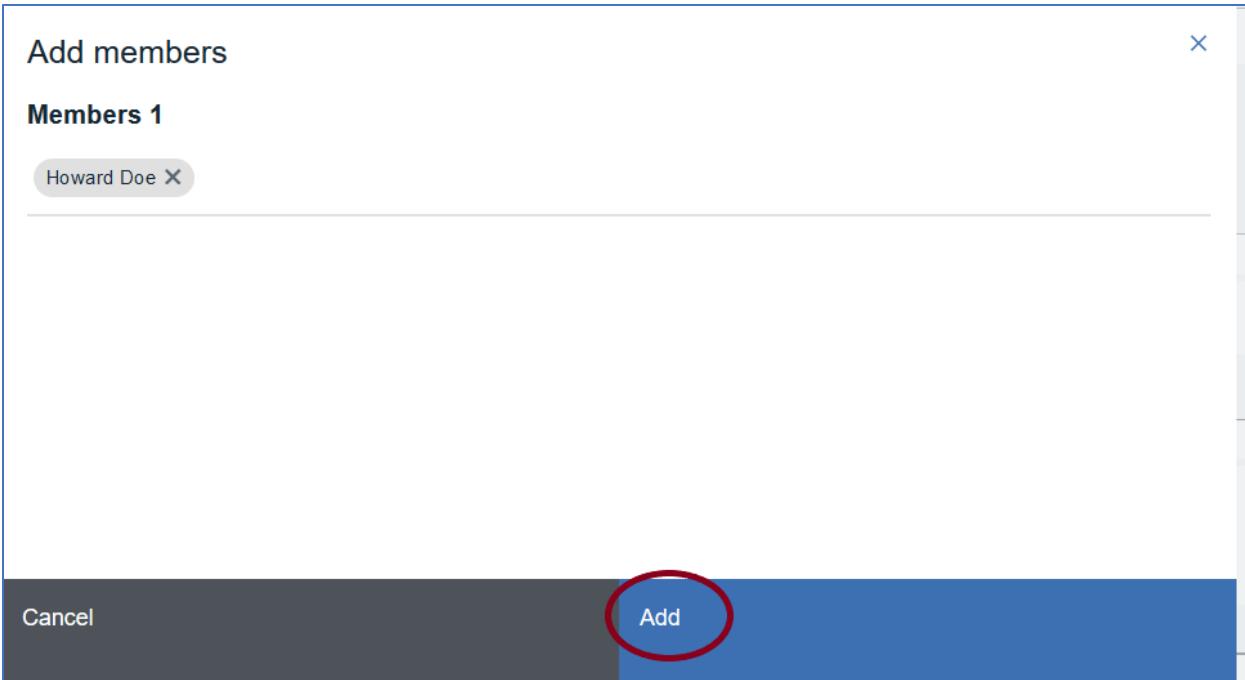
Members 0

Howard

Howard Doe ws.catalog.user@gmail.com

Cancel Add

6. Click on **Add** to add Howard Doe to the access control list for the female\_human\_trafficking asset.



7. Click on **Add**.

This screenshot shows a detailed "Add" dialog form:

- Name\***: female\_human\_trafficking
- Description**: (empty text area)
- Business terms**: Search business terms (empty text area)
- Tags**: trafficking X, female human traffi... X  
Start typing to add values (with a blue plus icon)
- Classification\***: None (with a dropdown arrow icon)

At the bottom right of the dialog, there are "Cancel" and "Add" buttons, with the "Add" button being circled in red.

8. All of the assets have now been added to the catalog. The catalog display should appear as below.

The screenshot shows the Watson Studio Labs Catalog interface. At the top, there are search filters for 'Any type' and 'Any tag', and a 'Clear all' button. Below these are two tabs: 'Watson recommends' (selected) and 'Highly rated'. A 'Recently added' section follows. The main area displays four items:

Name	Owner	Tags	Type	Date added
Categories	George Doe	job types, job categories	Data asset	Apr 26, 2020
Occupation	George Doe	occupations, jobs	Data asset	Apr 26, 2020
female_human_trafficking	George Doe	trafficking, female human trafficking	Data asset	Apr 26, 2020

The last row, 'female\_human\_trafficking', is highlighted with a red box. Below the table, a 'Showing 4 of 4 items' message is visible. At the bottom right, there is a 'Collapse' link.

## Step 4: Search the Catalog

Several options exist for searching the assets in the catalog. We can enter text in the search input text field and the system will match against asset names and tags. We can then further filter the search results by selecting the Asset Type, and/or the Tags. Or we can directly select the Asset Type or Tags without entering anything in the search text field.

1. Click on **Watson Studio Labs** to return to the Catalog home screen.

The screenshot shows the top navigation bar of the Watson Studio Labs interface. It includes links for 'IBM Watson Studio', 'Projects', 'Catalog' (which is highlighted in blue and has a red circle around it), 'Services', 'Community', 'Docs', 'Support', and 'Manage'. Below the navigation bar, the breadcrumb path 'Catalogs / Watson Studio Labs' is shown, with 'Watson Studio Labs' also having a red circle around it. At the bottom of the page, there are three buttons: 'Browse Assets' (underlined in blue), 'Access Control', and 'Settings'.

2. Enter **job** in the **search text field**, and click on **Collapse**.

Catalogs / Watson Studio Labs

Add to Catalog + ⚙️ 🌐 📁 📤 📈 🎯

## Watson Studio Labs

Browse assets Access control Settings

🔍 job

Any type Any tag Clear all

Watson recommends Highly rated Recently added

**Categories**

Owner	Added
George Doe	Apr 26, 2020 9:56 AM

Tags: job typeab categories

⭐⭐⭐⭐⭐ 0 reviews

**trafficking**

Owner	Added
George Doe	Apr 26, 2020 11:48 AM

⭐⭐⭐⭐⭐ 0 reviews

**Collapse**

Showing 2 of 2 items

3. Data assets are displayed that contain the word or letters **job** in their Name, Description, or tags.

Showing 2 of 2 items

<input type="checkbox"/> Name	Owner	Tags	Business terms	Type	Date added
<input type="checkbox"/> Categories	 George Doe	<a href="#">job types</a> <a href="#">job categories</a>		Data asset	Apr 26, 2020
<input type="checkbox"/> Occupation	 George Doe	<a href="#">occupations</a> <a href="#">jobs</a>		Data asset	Apr 26, 2020

4. Clear the **search text field**. Click on **Any Type**. Click on the check box next to **Data Asset**.

The screenshot shows a search interface with a search bar at the top containing the placeholder "What assets are you looking for?". Below the search bar are two dropdown menus: "Any type" and "Any tag". The "Any type" dropdown has a red oval highlighting the text "Any type". A sub-menu is open under "Any type", showing options like "Connection" and "Data asset". The "Data asset" option is also highlighted with a red oval. At the bottom of this sub-menu, it says "Showing Data asset 1 items".

5. The search results list the 3 Data Assets. The 1 Connected asset, and the 2 uploaded files.  
Note: Other asset types that can be cataloged include, Connections, Models, Notebooks, and Dashboards.

The screenshot shows the search results page. At the top, there is a search bar with the placeholder "What assets are you looking for?", a "Type" filter set to "1 x Type", an "Any tag" filter, and a "Clear all" button. Below the filters, a search result summary says "Showing 3 of 3 items". A table lists three data assets:

<input type="checkbox"/>	Name	Owner	Tags	Business terms	Type	Date added
<input type="checkbox"/>	Categories	George Doe	job types job categories		Data asset	Apr 26, 2020
<input type="checkbox"/>	Occupation	George Doe	occupations jobs		Data asset	Apr 26, 2020
<input type="checkbox"/>	female_human_trafficking	George Doe	trafficking female human trafficking		Data asset	Apr 26, 2020

6. Click on **Any Tag**. Click on the check box next to **job types**.

What assets are you looking for?

Type

Any tag

Clear all

Data asset

Showing 3 of 3 items

Name

Categories

Occupation

female human...

job categories

job types

jobs

occupations

trafficking

7. The Categories Data Asset is listed.

What assets are you looking for?

Type

Tag

Clear all

Data asset

job types

Show 1 of 1 item

Name	Owner	Tags	Business terms	Type	Date added
Categories	George Doe	job types, job categories		Data asset	Apr 26, 2020

## Step 5: Edit/Review/Profile Data Asset

In this step, we will drill down on the female\_human\_trafficking data asset to show how to edit the metadata describing the asset, to provide a rating and review of the asset, and to profile the asset.

1. Click on **Clear all**.

Q. What assets are you looking for?

Type  Tag  **Clear all**

Data asset  job types

Showing 1 of 1 item

<input type="checkbox"/> Name	Owner	Tags	Business terms	Type	Date added
<input type="checkbox"/> Categories	George Doe	job types job categories		Data asset	Apr 26, 2020

2. Click on the female\_human\_trafficking asset.

Showing 4 of 4 items

<input type="checkbox"/> Name	Owner	Tags	Business terms	Type	Date added
<input type="checkbox"/> Categories	George Doe	job types job categories		Data asset	Apr 26, 2020
<input type="checkbox"/> trafficking	George Doe			Connection	Apr 26, 2020
<input checked="" type="checkbox"/> female_human_trafficking	George Doe	trafficking male human trafficking		Data asset	Apr 26, 2020
<input type="checkbox"/> Occupation	George Doe	occupations jobs		Data asset	Apr 26, 2020

3. The metadata and data contents of the asset are displayed as shown below. Toward the top of the screen, are 5 tabs, highlighted in blue below. The panel shown below corresponds to the **Overview** tab. The **Access** tab allows the user to edit the access control list for this asset. The **Review** tab enables the user to provide a rating and descriptive review of the asset. These social features aid in searching for relevant and useful assets. The **Profile** tab invokes an auto-profiling capability where the columns of the data asset are classified based on machine learning scoring into one of many classifiers that are built into the platform. This capability will be shown below. Finally, the **Lineage** tab tracks the changes made to the data asset over its lifecycle. This capability is not available in the lite catalog offering so will not be exercised in this lab.

Catalogs / Watson Studio Labs / female\_human\_trafficking Add to Catalog +

DATA ASSET **female\_human\_trafficking** Remove Download Add to project +

Overview Access Review Profile Lineage

**Description**  
There is no description available for this asset.

**Added:** Apr 26, 2020 2:18 PM **Format:** application/octet-stream

**Business terms**  
There are no terms available for this asset.

**Tags**  
female human trafficking trafficking

**Reviews**  
☆☆☆☆☆ 0 reviews

**Connection**  
**Source:** trafficking  
**Source type:** DB2 Warehouse

**Schema:** 26 Columns  
Preview: 1000 rows Last refresh: 2 minutes ago

INTERNAL_ID	VETTING_LEVEL	DESCRIPTION	NAME	GENDER	BIRTH_DATE	BIRTH_COUNTRY	BIRTH_COUNTRY_CODE
338	100	NA	Meghan Moses	F	1994-08-27	Ghana	GH
339	30	NA	Trace Carr	F	2001-11-30	Ghana	GH
340	10	NA	Ami Casey Wood	F	1983-11-05	Ghana	GH
341	30	NA	Mellinda Kimm Hi	F	1980-01-16	Brazil	BR
342	10	NA	Linda Tucker	F	1995-01-14	Brazil	BR
343	100	NA	Tamara Palmer	F	1987-05-01	Ghana	GH
344	30	NA	Brandy Scott	F	1999-08-09	Ghana	GH
345	100	NA	Nann Steffy Willia	F	2001-06-07	Ghana	GH
346	30	NA	Jesie Molly Stafft	F	1970-05-02	Bangladesh	BD
347	100	NA	Elizabeth Ronnie	F	1995-05-07	Ghana	GH
348	20	NA	Malissa Rodier	F	2004-09-24	China	CH

The actions circled in green will **Remove** the asset from the catalog, **Download** the asset to the local file system (grayed out), or add the asset to a catalog enabled project (**Add to Project**). Hovering over the **Name** of the data asset, the **Description**, the **Business Terms**, or the **Tags** will popup an Edit icon that allows the user to edit these entities.

#### 4. Click on the **Review** tab.

Catalogs / Watson Studio Labs / female\_human\_trafficking

DATA ASSET

**female\_human\_trafficking**

Remove Download Add to project

Overview Access **Review** Profile Lineage

Description Schema: 26 Columns  
There is no description available for this asset. Preview: 1000 rows

Last refresh: 2 minutes ago

#### 5. Click on the fourth star to rate this asset. Enter “Good data set for demonstrating data profiling capabilities” in the **My Review** text field. Click on the **Submit** button.

DATA ASSET

**female\_human\_trafficking**

Remove Download

Overview Access **Review** Profile Lineage

**Overall rating**  
0.0  
☆☆☆☆☆ 0 reviews

**Review summary**  
5 (0)  
4 (0)  
3 (0)  
2 (0)  
1 (0)

Your review

George Doe Apr 26, 2020  
★★★★☆

Good data set for demonstrating data profiling capabilities

Cancel **Submit**

#### 6. The review of the asset is displayed. Click on the **Profile** tab.

Catalogs / Watson Studio Labs / female\_human\_trafficking

DATA ASSET

**female\_human\_trafficking**

Remove Download Add to project

Overview Access **Review** Profile Lineage

**Overall rating**  
4.0  
☆☆☆☆☆ 1 review

**Review summary**  
5 (0)  
4 (1)

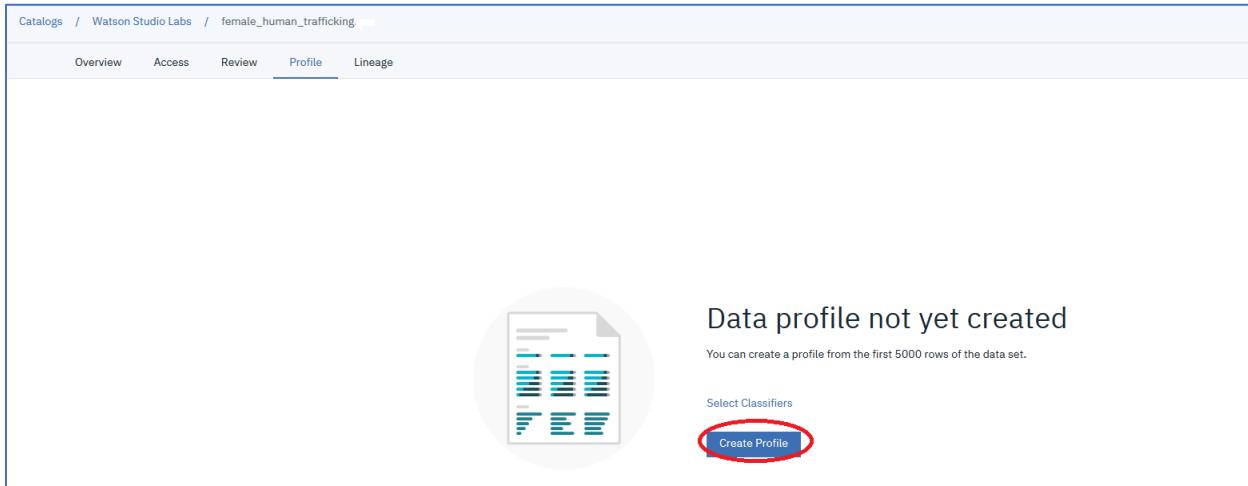
Your review

George Doe Apr 26, 2020  
★★★★☆

Good data set for demonstrating data profiling capabilities

Successfully added your review.

7. Click on **Create Profile**. If a profile was already created, it would be displayed instead of this prompt screen. In that case, skip to step 9.



Catalogs / Watson Studio Labs / female\_human\_trafficking

Overview Access Review **Profile** Lineage

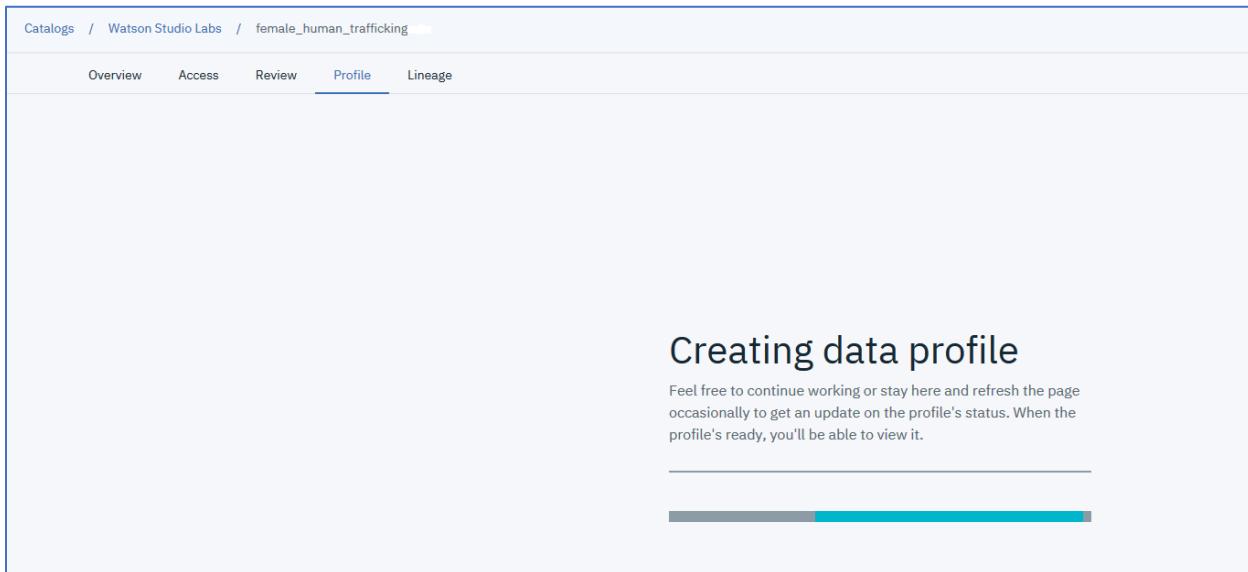
Data profile not yet created

You can create a profile from the first 5000 rows of the data set.

Select Classifiers

**Create Profile**

8. This step could take some time. Please wait a minute or so, and then click on browser refresh.



Catalogs / Watson Studio Labs / female\_human\_trafficking

Overview Access Review **Profile** Lineage

**Creating data profile**

Feel free to continue working or stay here and refresh the page occasionally to get an update on the profile's status. When the profile's ready, you'll be able to view it.

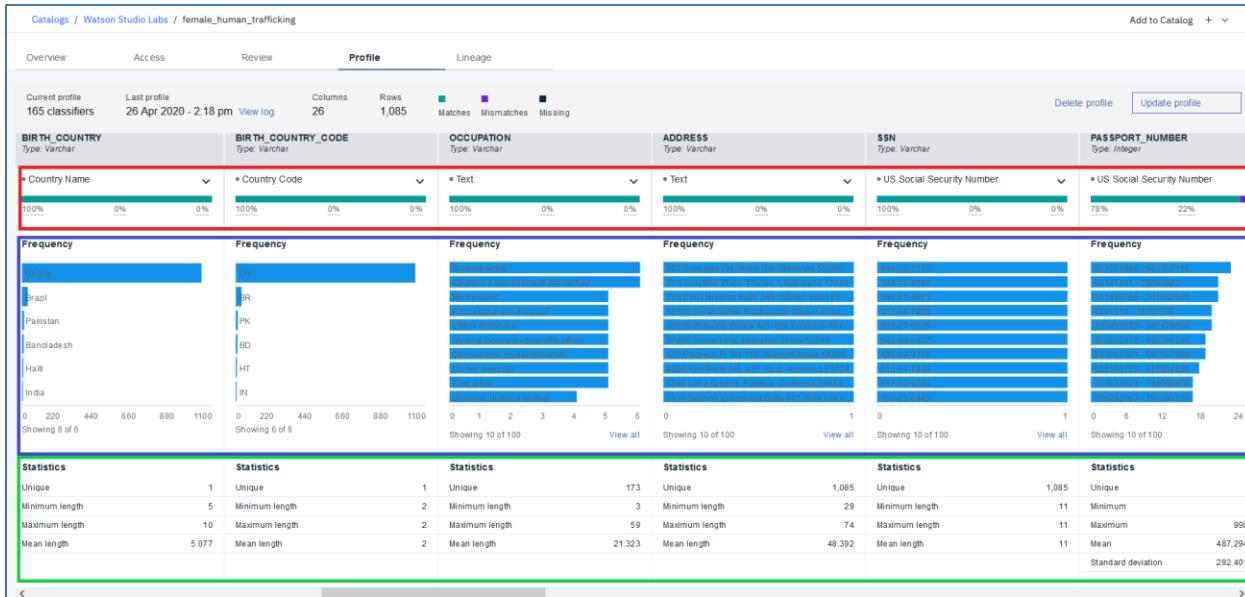
---

9. The profile information is displayed below using several screenshots, based on horizontal scrolling, to display many of the columns of the data asset. The highlighted red box contains the inferred column classifications. The highlighted blue box contains counts of the number of occurrences of each value (for discrete fields), and a histogram for continuous fields. The green box contains statistics calculated from the contents of the columns.

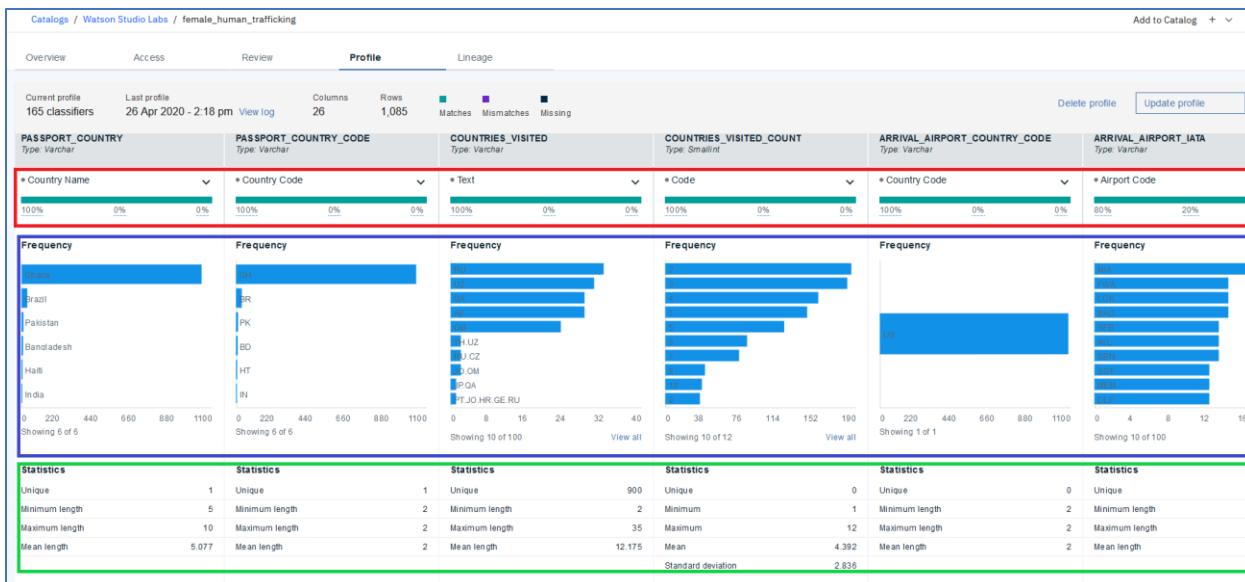
The screenshot shows the IBM Watson Studio interface with the 'female\_human\_trafficking' catalog selected. The 'Profile' tab is active. At the top, it shows '165 classifiers' and 'Last profile 26 Apr 2020 - 2:18 pm'. Below this, the 'INTERNAL\_ID' column is highlighted with a red box, showing its type as 'Smallint' and a frequency chart for values ranging from 1000-7000. The 'NAME' column is also highlighted with a red box, showing its type as 'Varchar' and a frequency chart for person names. The 'BIRTH\_DATE' column is highlighted with a green box, showing its type as 'Date' and a frequency chart for dates. Other columns like 'VETTING\_LEVEL', 'DESCRIPTION', 'GENDER', and 'NAME' are also visible with their respective profiles and frequency charts.

The classifications of the columns are listed below:

- INTERNAL\_ID – Identifier
- VETTING\_LEVEL – Code
- DESCRIPTION – Country Code – this is an error due to the description input as NA. We can correct this misclassification.
- NAME: Person Name – the classification accurately determines that this field is the name of a person.
- GENDER: Gender – the classification accurately determines that this field is the gender of a person
- BIRTH\_DATE: Date of Birth – the classification accurately determines that this field represents a date of birth.



- BIRTH\_COUNTRY: Country Name- correctly classified as a country
- BIRTH\_COUNTRY\_CODE – Country Code - correctly classified
- OCCUPATION – Text – there is not a classifier for occupation, so it determines that this is a text string.
- ADDRESS: Text
- SSN – US Social Security Number – correctly classifies the column as a social security number.
- PASSPORT\_NUMBER – incorrectly classifies as a social security number. This can be corrected as we will see below.



- PASSPORT\_COUNTRY - Country Name- correctly classified as a country
- PASSPORT\_COUNTRY\_CODE: Country Code - correctly classified

- COUNTRIES\_VISITED – Text
- COUNTRY\_VISITED\_COUNT – Code – this is an error. It's not a code. This can be re-classified if necessary.
- ARRIVAL\_AIRPORT\_COUNTRY\_CODE – Country Code – correctly classified.
- ARRIVAL\_AIRPORT\_IATA: Airport Code – the classification accurately determines that this field represents an Airport Code.

There are more columns in the data set but this gives a representative sample of the profiling capability. The classification of the columns becomes important when we create and enforce security policies.

10. Let's fix the classification for passport. Click  icon next to the **US Social Security Number** underneath the **PASSPORT\_NUMBER** column. Click on **View All**.



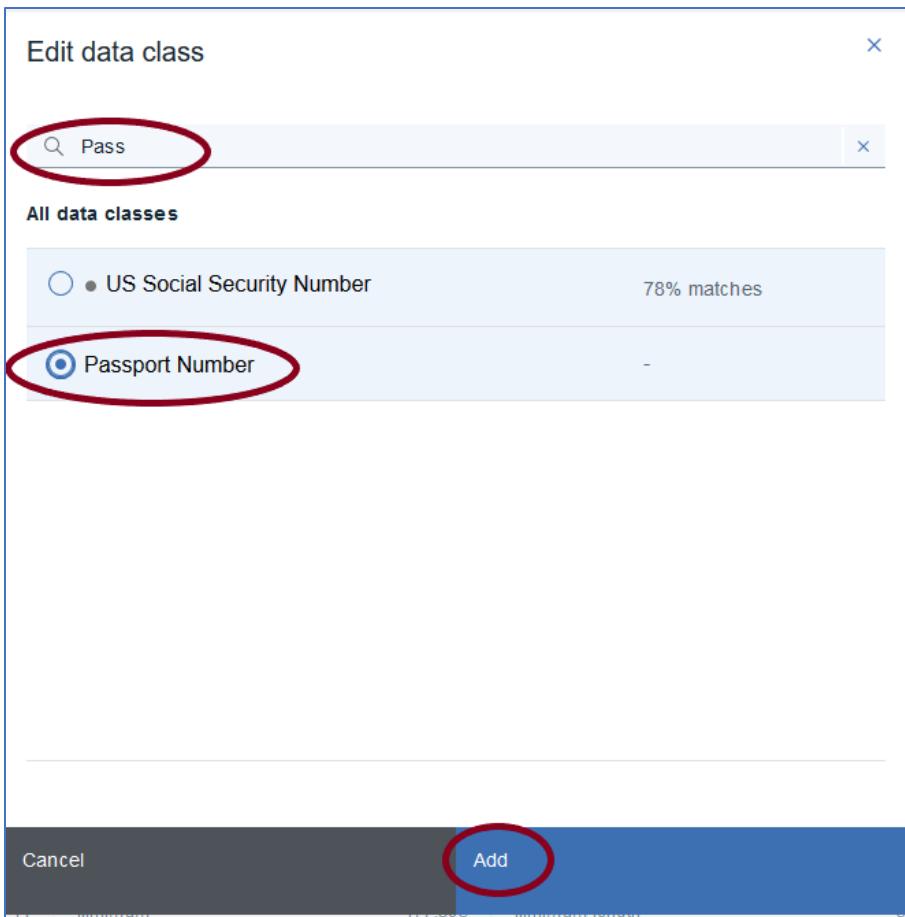
**PASSPORT\_NUMBER**  
Type: Integer

- US Social Security Number 
- ✓ US Social Security Num... 78% matches
- Not classified 21% matches
- Canadian Social Insura... 7% matches
- Routing Transit Number 3% matches

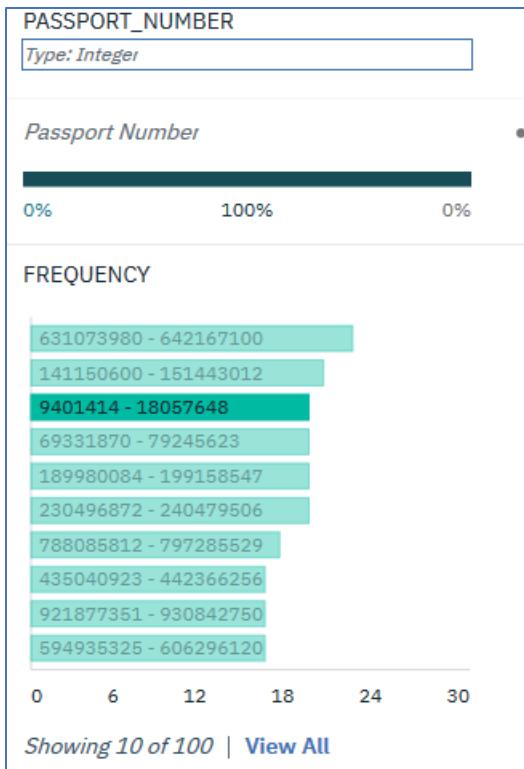
[View all](#)

9401414 - 18057648

11. Type “Pass” in the **Search** field. Click on **Passport Number** radio button. Click on **Add**.



12. The column classification has been changed.



## Step 6: Access Control

All Catalog members can access assets with a privacy setting of Public. Assets that have a privacy setting of Private are secured based on the asset access control list. A summary of the access control settings for our Watson Studio Labs catalog and the contained Data Assets is shown below.

Watson Studio Lab Catalog members:

1. Catalog owner (logged on user)
2. [ws.catalog.user@gmail.com](mailto:ws.catalog.user@gmail.com)

Data Asset Privacy:

1. female\_human\_trafficking (Private) ACL Members:
  - a. asset owner (logged on user)
  - b. [ws.catalog.user@gmail.com](mailto:ws.catalog.user@gmail.com)
2. Occupations (Private) ACL Members:
  - a. asset owner (logged on user)
3. Categories (Public)
4. trafficking (Connection) (Public)

We would expect that if we log in as [ws.catalog.user@gmail.com](mailto:ws.catalog.user@gmail.com), that only 3 Data Assets would appear. The two public Data Assets, and the Private Data Asset where [ws.catalog.user@gmail](mailto:ws.catalog.user@gmail.com) is on the access control list. This user would not have access to the Occupations asset, since it is a

Private asset, and the user is NOT on the access control list for this asset. To verify complete the following steps.

1. Open a different browser (if you are using FireFox, then open Chrome or vice versa).
2. Type in **dataplatform.ibm.com** for the URL to navigate to the Watson Studio signin.
3. Click on **Log in**



4. Login using [ws.catalog.user@gmail.com](#) for the user name and click **Continue**.

A screenshot of a login form titled "Log in to IBM". The form has a light blue header. Below it, there is a text input field labeled "IBMid" with the placeholder "ws.catalog.user@gmail.com" (which is circled in red). To the right of the input field is a link "Forgot IBMid?". Below the input field is a checkbox labeled "Remember me" with a blue "i" icon next to it. At the bottom of the form is a large blue button with the word "Continue" in white (which is circled in red). Below the form, there is a horizontal line and the text "Don't have an account? [Create an IBMid](#)".

5. Enter **wsPassw0rd** for the **Password** and click **Log in**.

# Log in to IBM

Logging in as ws.catalog.user@gmail.com [Not you?](#)

Password

[Forgot password?](#)

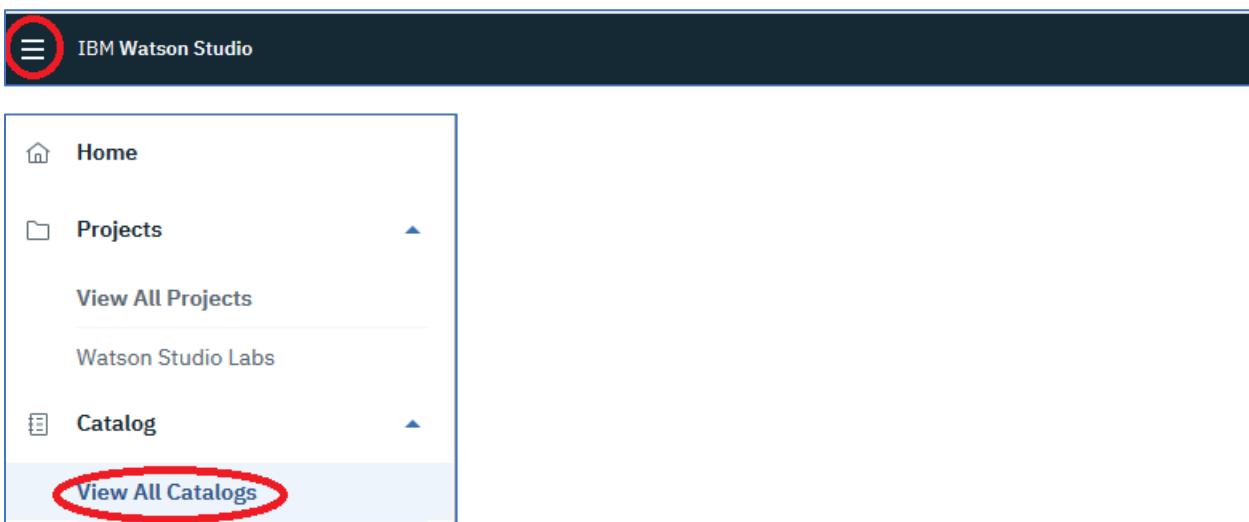
 

Remember me [\(i\)](#)

**Log in**

Don't have an account? [Create an IBMid](#)

6. Click on  icon. Click on **Catalog** and then **View All Catalogs**



The image shows the IBM Watson Studio interface. At the top, there is a dark header bar with the "IBM Watson Studio" logo on the left. Below the header is a light blue sidebar containing the following navigation items:

-  Home
-  Projects
  - [View All Projects](#)
  - Watson Studio Labs
-  Catalog
  - [View All Catalogs](#)

7. Scroll down to find the **Watson Studio Labs** catalog that corresponds to your name. You can also type <Ctrl><F> and enter your name to search. Note that there will be 1 catalog per lab participant. Click on the **Watson Studio Labs** catalog corresponding to your user name.

The screenshot shows the IBM Watson Studio interface with a dark header bar. The main content area is titled 'Your catalogs'. A catalog entry for 'Watson Studio Labs' is displayed, with the catalog name 'Watson Studio Labs' circled in red. Below the catalog name, account details are shown: Account Name: IBM, Creator: DSX56000 User, Date created: Jan 20, 2019 5:33 PM. There is a 'New Catalog' button in the top right corner.

8. The Watson Studio Labs catalog is displayed. Three Data Assets are listed as expected.

The screenshot shows the 'Watson Studio Labs' catalog page. At the top, there are sections for 'Watson Recommends', 'Highly Rated', and 'Recently Added'. Below these are two cards: 'Data Asset Categories' and 'Connection trafficking'. Underneath these cards, a list of items is shown with a red border around the entire list. The list includes three items: 'Categories' (Owner: Andrew Doe, Added: Apr 30, 2019 9:03 PM, Tags: job types, job categories), 'female\_human\_trafficking' (Owner: Andrew Doe, Added: Apr 30, 2019 9:26 PM, Tags: trafficking, female huma...), and 'trafficking' (Owner: Andrew Doe, Added: Apr 30, 2019 9:26 PM, Tags: trafficking).

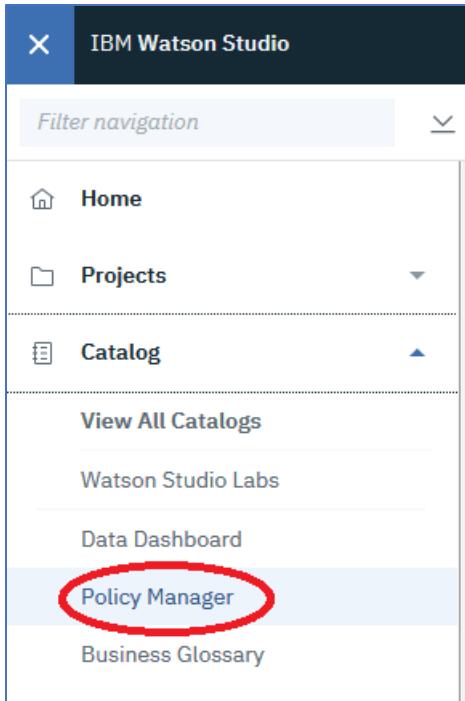
9. Leave this browser at this screen for Step 7 below.

### Step 7: Create and Enforce Policy (optional)

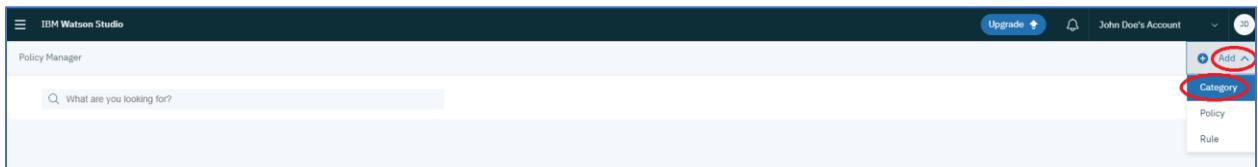
In this step, we will create a policy that will anonymize the SSN and Passport fields when user [ws.catalog.user@gmail.com](mailto:ws.catalog.user@gmail.com) tries to access the female\_human\_trafficking data asset.

1. **RETURN TO THE BROWSER WHERE YOUR** Watson Studio id is logged in.
2. Click on the icon. Click on **Catalog** and then **Policy Manager**.

The screenshot shows the IBM Watson Studio interface with a dark header bar. The navigation menu on the left includes 'IBM Watson Studio', 'Projects', 'Catalog', 'Services', 'Community', 'Docs', 'Support', and 'Manage'. The 'Catalog' and 'Policy Manager' options are highlighted in the dropdown menus. The main content area is currently empty.



3. Click on **Add** and then **Category**.



4. Type in **Privacy** for the **Name** of the Category, and optionally a **Description**, and click on **Create**.

New category

Details

Name\*  
Privacy (circled in red)

Description  
Policies that enforce privacy (circled in red)

Cancel Create (circled in red)

A screenshot of the 'New category' creation form. It has fields for 'Name\*' and 'Description'. The 'Name\*' field contains 'Privacy' (circled in red). The 'Description' field contains 'Policies that enforce privacy' (circled in red). At the bottom right are 'Cancel' and 'Create' buttons, with 'Create' being circled in red.

5. Click on **Add** and then **Policy** to add a policy to the Privacy Category.

6. Type in PII for the Policy **Name**, type in “Limit access to PII data” for the Policy **Description**, click on **create a new rule**.

7. Enter “Redact Government ID” for the rule **Name**, enter “Redact a Government ID” for the **Business Definition**, for the rule click in the entry fields to select the following:

If **Data Class** contains any  
**Government Identities**

**And**

If **User Name** contains any  
**Howard Doe** (make sure you select Howard Doe from the list!)

then **anonymize data** where **in columns containing**  
**Government Identities**

The screenshot shows the IBM Watson Studio Rule builder interface. On the left, there's a 'Details' panel with fields for 'Name' (Redact Government ID) and 'Type' (Access). Below that is a 'Business definition' section with the same text. On the right, the 'Rule builder' panel contains two conditions under 'Criteria'. Condition 1 is 'If Data class contains any Government Identities'. Condition 2 is 'If User Name contains any Howard Doe'. Both conditions have red circles around 'Data class' and 'User Name'. Below the conditions is an 'Action' section with 'then anonymize data in columns containing Government Identities'. A red circle highlights 'anonymize data'. At the bottom right of the dialog are 'Cancel' and 'Create' buttons.

8. Scroll down and select how to anonymize data: **Substitute**, and then click **Create**.

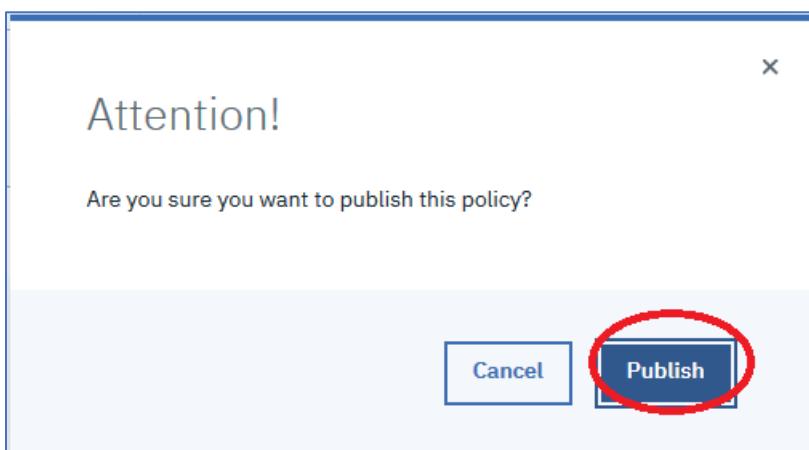
The screenshot shows the 'Anonymize data' configuration dialog. It has three options: 'Redact', 'Substitute' (which is circled in red), and 'Mask'. Below the 'Substitute' option, there's a note: 'Replace data with values that don't match the original format.' At the bottom of the dialog are 'Cancel' and 'Create' buttons, with 'Create' also circled in red.

9. Click on **Publish**.

New policy

<b>Details</b>	<b>Rules</b> <small>1</small>	<small>Add</small>
<b>Name*</b> <input type="text" value="PII"/>		
<b>Type*</b> <input type="text" value="Access"/>		
<b>Category*</b> <input type="text" value="Privacy"/>		
<b>Description</b> <input type="text" value="Limit access to PII Data"/>		
<small>Cancel</small> <small>Save as draft</small> <b>Publish</b>		

10. Click on **Publish**.



11. The PII policy has been added.

Policy Manager / Privacy		<small>Add</small>
<b>Category</b> Privacy		
<b>Description</b> Policies that enforce privacy		
<small>Creator: DSX59000 User    Created: Feb 5, 2019    Contains: 1 published policies and 1 rules</small>		
<input type="text" value="Q. What are you looking for?"/> <input type="text" value="NAME"/> > PII	<small>TYPE</small>	<small>STATUS</small>
	Policy	Published
		1 Item
		Feb 5, 2019
		<small>LAST MODIFIED</small>

12. Click on the  icon, then click on **Catalog** and then **Watson Studio Labs**.



Filter navigation

- [Home](#)
- [Projects](#)
- [Catalog](#)
- [View All Catalogs](#)
- [Watson Studio Labs](#)
- [Data Dashboard](#)
- [Policy Manager](#)
- [Business Glossary](#)

13. Switch back to the browser with the [ws.catalog.user@gmail.com](mailto:ws.catalog.user@gmail.com) user logged in. Click on the **female\_human\_trafficking** asset.

Available Assets

Showing 3 of 3 assets

NAME	OWNER	TAGS	TYPE	DATE ADDED
Categories.csv	DSX59000 User		Data Asset	Jan 31, 2019
female_human_trafficking	DSX59000 User	trafficking female human trafficking	Data Asset	Jan 31, 2019
trafficking	DSX59000 User		Connection	Jan 31, 2019

14. The **female\_human\_trafficking** asset is displayed. Wait for the Data Masking to complete.

Catalogs / Watson Studio Labs / female\_human\_trafficking

DATA ASSET

**female\_human\_trafficking**

Overview    Access    Review    Profile    Lineage

Description  
There is no description available for this asset.

Added: Apr 26, 2020 2:18 PM  
Format: application/octet-stream  
Size: 321 KB

Business terms  
There are no terms available for this asset.

Tags  
female human trafficking, trafficking

Reviews  
★ ★ ★ ★ ☆ 1 review

Classification

Schema: 26 Columns | 1085 rows | 2 Columns masked

Last refresh: just now

INTERNAL... Smallint VETTING\_L... Smallint DESCRIP... String NAME String GENDER String BIRTH\_D... Date BIRTH\_COU... String BIRTH\_COUNTRY\_C... String OCCU... String

Identif... Code Countr... Perso... Gender Date o... Country N... Country Code Text

Data masking in progress

This asset is being masked by the data enforcement rule: PII. You can wait here to see a preview of the asset or we can notify you when the preview is ready.

Notify Me

15. Note that 2 columns are anonymized.

Added:	Apr 26, 2020 2:18 PM	OCCUPAT...	ADDRESS	SSN	PASSPORT_NUMB.	PASSPORT_COU...	PASSPORT_COUNTR...	COUNTRIES_VIS...
Format:	application/octet-stream	String	String	String	String	String	String	String
Size:	321 KB	Text	Text	US So...	Passport Number	Country Name	Country Code	Text
<b>Business terms</b>	There are no terms available for this asset.							
<b>Tags</b>	<a href="#">female human trafficking</a> <a href="#">trafficking</a>							
<b>Reviews</b>	1 review							
<b>Connection</b>								
<b>Source:</b>	Watson Studio Labs_DataCatalog							
<b>Source type:</b>	Cloud Object Storage							
<b>Classification</b>	None							

16. Close the browser tab for user ws.catalog.user@gmail.com.

## Step 8: Push the data assets to the project

This step will add the cataloged data assets to our Watson Studio project to be used in subsequent labs. Return to your Watson Studio account.

1. Click on **Watson Studio Labs** to go to the Catalog home screen.

2. Click on the checkbox to the left of the **female\_human\_trafficking** asset and the **trafficking** asset. Click on **Add to Project**.

Name :	Owner	Tags	Business terms	Type	Date added
<input type="checkbox"/> Categories	George Doe	job types job categories		Data asset	Apr 26, 2020
<input type="checkbox"/> Occupation	George Doe	occupations jobs		Data asset	Apr 26, 2020
<input checked="" type="checkbox"/> female_human_trafficking	George Doe	trafficking female human trafficking		Data asset	Apr 26, 2020
<input checked="" type="checkbox"/> trafficking	George Doe			Connection	Apr 26, 2020

3. Select the Watson Studio Labs as the **Target**. Click on **Add**.

Catalogs / Watson Studio Labs

Add to project

Target: Watson Studio Labs

Selected assets (2)

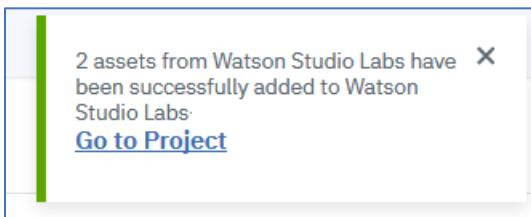
Asset name	Catalog	Connection
female_human_trafficking	Watson Studio Labs	trafficking
trafficking	Watson Studio Labs	

Connections to be added (1)

trafficking	Watson Studio Labs	Db2 Warehouse
-------------	--------------------	---------------

Cancel Add

4. Wait for the message that 2 assets have been added to the project in the top right.



5. Click on **Categories**, and **Occupation**, and then click again on **Add to Project**.

Add to project Remove

2 of 4 items selected Cancel

Name	Owner	Tags	Business terms	Type	Date added
<input checked="" type="checkbox"/> Categories	George Doe	job types, job categories		Data asset	Apr 26, 2020
<input checked="" type="checkbox"/> Occupation	George Doe	occupations, jobs		Data asset	Apr 26, 2020
<input type="checkbox"/> female_human_trafficking	George Doe	trafficking, female human trafficking		Data asset	Apr 26, 2020
<input type="checkbox"/> trafficking	George Doe			Connection	Apr 26, 2020

6. Select **Watson Studio Labs** as the **Target** and click **Add**.

Catalogs / Watson Studio Labs

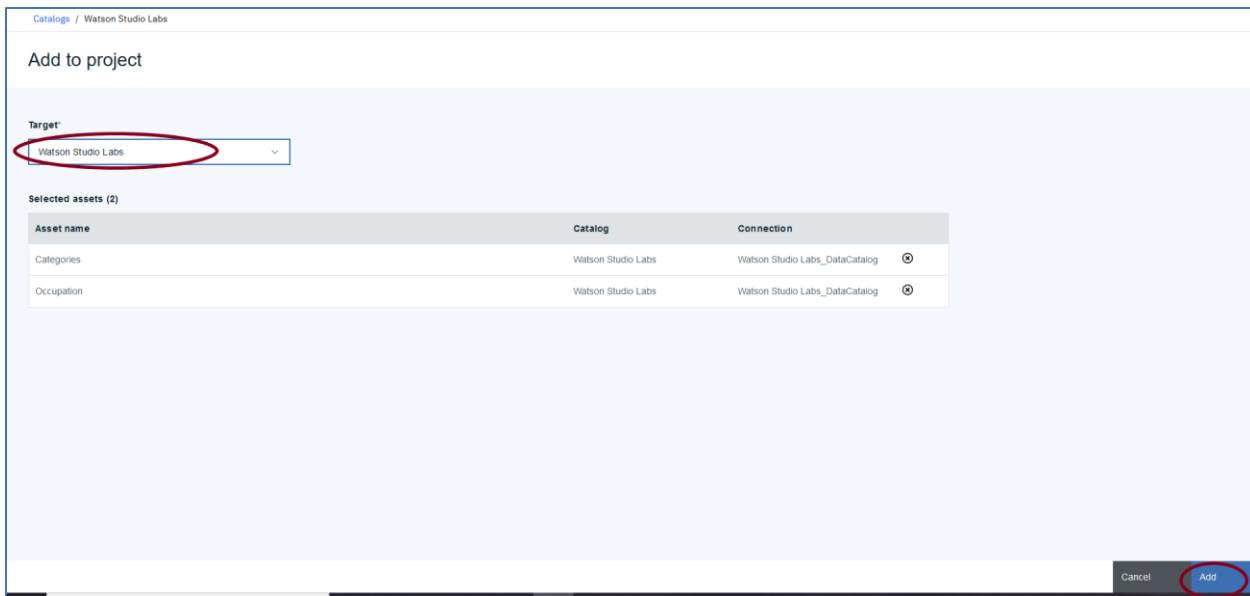
Add to project

Target: Watson Studio Labs

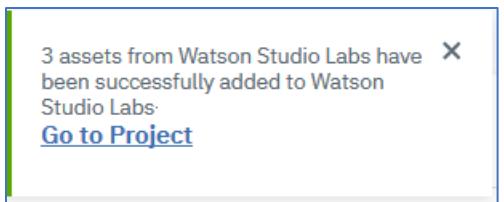
Selected assets (2)

Asset name	Catalog	Connection
Categories	Watson Studio Labs	Watson Studio Labs_DataCatalog
Occupation	Watson Studio Labs	Watson Studio Labs_DataCatalog

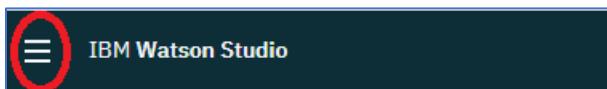
Cancel Add



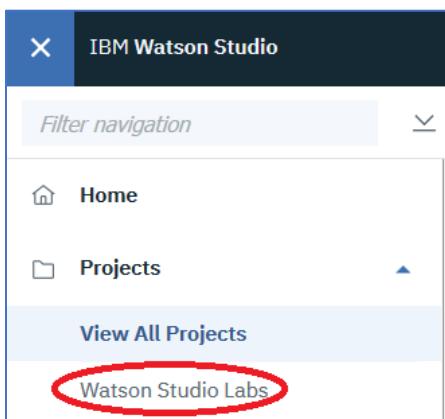
7. Wait for the message that **3** assets have been added to the project in the top right. Two data assets and a connection to the catalog object storage.



8. Click on the hamburger icon .



9. Click on **Watson Studio Labs** under **Projects**.



10. Click on the Assets tab.

The screenshot shows the IBM Watson Studio interface. At the top, there is a navigation bar with links for 'IBM Watson Studio', 'Projects', 'Catalog', 'Services', 'Community', 'Docs', 'Support', and 'Manage'. On the right side of the top bar, there are buttons for 'Upgrade', 'Launch IDE', and '+ Add to project'. Below the top bar, the URL 'https://dataplatform.cloud.ibm.com/projects/014024b5-6551-42d4-90f0-59edc2cc556d/assets' is visible. The main content area has tabs for 'Overview', 'Assets' (which is circled in red), 'Environments', 'Bookmarks', 'Deployments', 'Access Control', and 'Settings'. The 'Assets' tab is currently selected.

11. The Data assets are now incorporated in the project.

The screenshot shows the 'Assets' tab in the IBM Watson Studio interface. The left sidebar has a section titled 'Data assets' with a sub-section 'Occupations'. The main content area displays a table of data assets:

NAME	TYPE	CREATED BY	LAST MODIFIED	ACTIONS
female_human_trafficking	Data Asset	DSX56000 User	23 Jan 2019, 5:04:39 pm	⋮
Occupations	Data Asset	DSX56000 User	23 Jan 2019, 5:04:38 pm	⋮
Categories	Data Asset	DSX56000 User	23 Jan 2019, 5:04:38 pm	⋮
Occupations	Connection	DSX56000 User	23 Jan 2019, 5:04:37 pm	⋮

On the right side of the interface, there is a file upload area with the placeholder text 'Drop files here or browse for files to upload.' Below the table, there is a brightness slider set to 15. The bottom of the screen shows the Windows taskbar with various pinned icons.

## You have completed Lab-2!

- ✓ Created a governed catalog
- ✓ Added a member to the catalog
- ✓ Added Data Assets to the catalog
- ✓ Searched the catalog
- ✓ Edited/Reviewed/Profiled a structured Data Asset
- ✓ Demonstrated access control features
- ✓ Demonstrated policy creation and enforcement (optional)
- ✓ Pushed the Data Assets to the project set up in Lab-1