

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 workshop 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**.

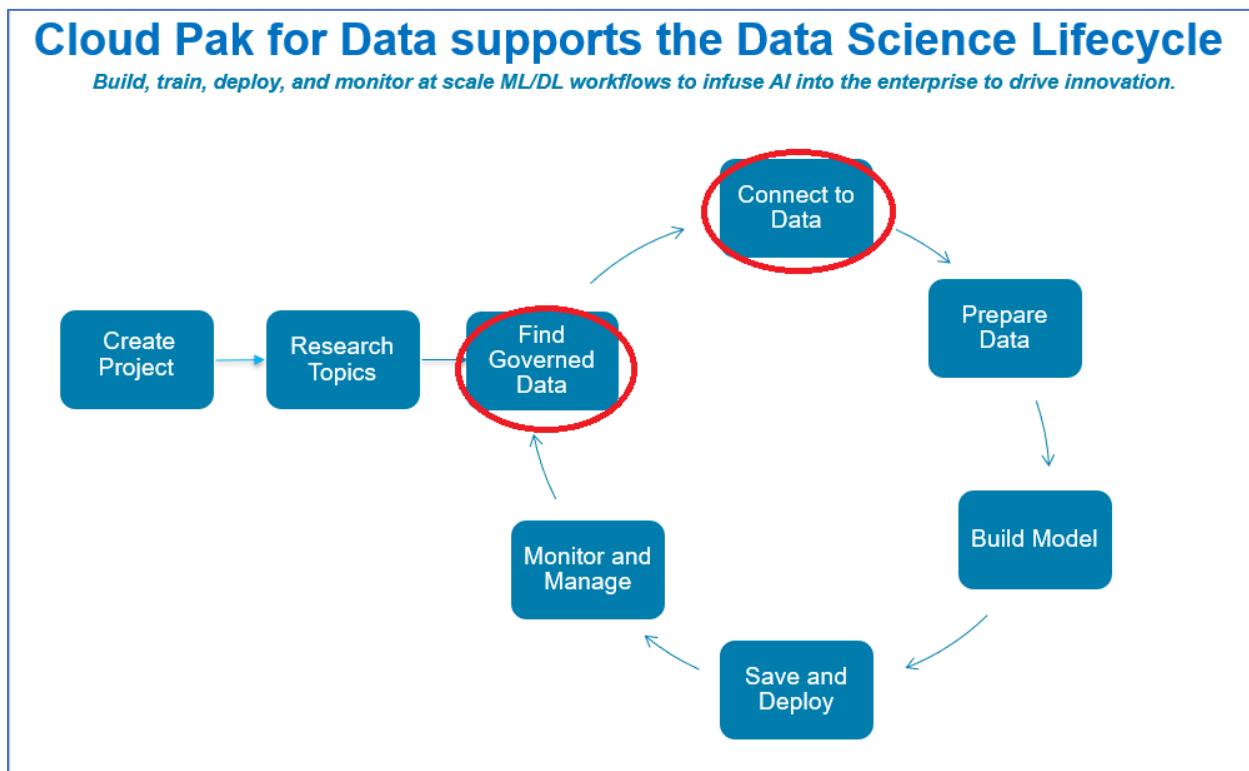


Figure 1- End to End Flow

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 Platform access connections

2. Create a governed catalog
3. Add a member to the catalog
4. Add Data Assets to the catalog
5. Push the Data Assets to the project set up in Lab-1
6. Search the catalog
7. Edit/Review/Profile a structured Data Asset
8. Demonstrate access control features

Create Platform Asset Connections Catalog

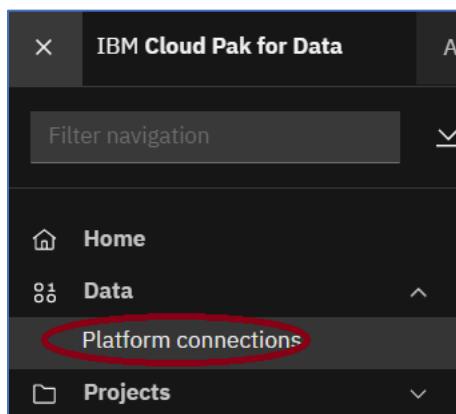
The Platform assets catalog enables sharing of connections across your organization. Any user who you add as a collaborator to the catalog can see these connections. However, only users with the credentials for the data source can use a connection.

You can add an unlimited number of collaborators and connection assets to the Platform assets catalog.

The Platform assets catalog is part of the Cloud Pak for Data as a Service platform and does not require the Watson Knowledge Catalog service. However, if you have the Watson Knowledge Catalog service, you can add other types of assets to the Platform assets catalog besides connection assets.

Step 1: Create Platform Connections Catalog

1. Click on the hamburger  icon, click on **Data**, and then **Platform connections**.



2. Click **Create Catalog**.

Platform connections



Share connections in the Platform assets catalog

To share platform-level connections with users on the platform, create the Platform assets catalog. Then, add your platform users and your platform level connections to the catalog.

[Learn more](#)



3. Select the object storage (if you have more than one instance defined) and click **Create**.
The catalog name is hard-coded to be *Platform assets catalog*.

New catalog

Details

Name: Platform assets catalog

Description: This catalog stores assets that are shared across the platform.

IBM Cloud Object Storage ⓘ
This service stores the files associated with assets in the catalog.

Object storage instance: CloudObjectStorage

Create

4. The Platform asset catalog has been successfully created.

The screenshot shows the 'Platform assets catalog successfully created!' message at the top. Below it, the 'Access control' tab is selected. A yellow banner says 'Add your platform users to the catalog. You can add individual users as collaborators or add access groups.' The 'Collaborators' section lists one user: Adrian Doe (wsuser70000@gmail.com) with Admin role added on Aug 11, 2021.

Step 2: Add Collaborator

Let's add ws.catalog.user@gmail.com as a collaborator with Viewer access.

1. Click Add collaborators.

The screenshot shows the 'Access control' tab selected. The 'Collaborators' section has a red circle around the 'Add collaborators' button.

2. Select **Viewer** as the **Role**. Enter ws.catalog.user@gmail.com and it will add **Howard Doe** to the list of collaborators. Click Add.

The screenshot shows the 'Add collaborators as Viewers' dialog. It has a message about trouble adding collaborators. Under 'Role', 'Viewer' is selected. In the 'Collaborators (1)' list, 'Howard Doe' is listed with a red circle around it. At the bottom, the 'Add' button is highlighted with a red circle.

- Howard Doe is added to the list of collaborators.

The screenshot shows a table with columns: Name, Email, Role, and Date added. There are two rows:

Name	Email	Role	Date added
HD Howard Doe	ws.catalog.user@gmail.com	Viewer	Aug 11, 2021
AD Adrian Doe	wsuser70000@gmail.com	Admin	Aug 11, 2021

Create Platform Connections

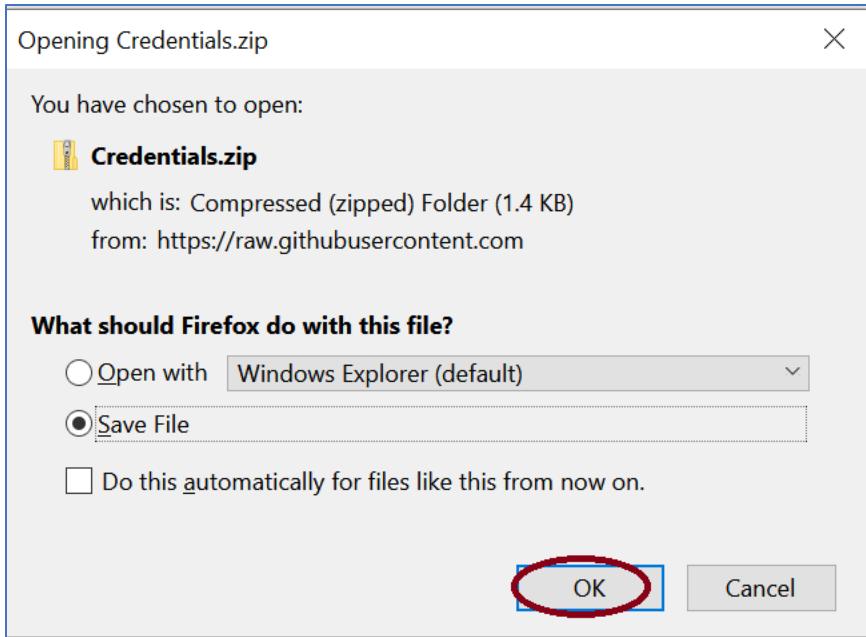
We are going to add two connections. One to a MongoDB instance on the IBM Cloud, and one to a DB2 Warehouse on Cloud instance. We will get the connection information by downloading the Credentials.zip file which contains credentials needed to establish the two connections.

Step 1: Download the Credentials.zip file

- Click [here](#) to download the Credentials.zip file. Then click on **Download**.

The screenshot shows a GitHub repository page for 'bleonardb3 / DS_POT_08-26-2021'. The 'Code' tab is selected. A single file, 'Credentials.zip', is listed in the main area. At the bottom right of the file card, there is a 'Download' button, which is circled in red.

- Select the **Save File** option and click **OK**.



3. Extract the files from Credentials.zip. There should be two files, one with credentials for the MongoDB connection, and one with credentials for the DB2 Warehouse connection.

Step 2: Add MongoDB connection.

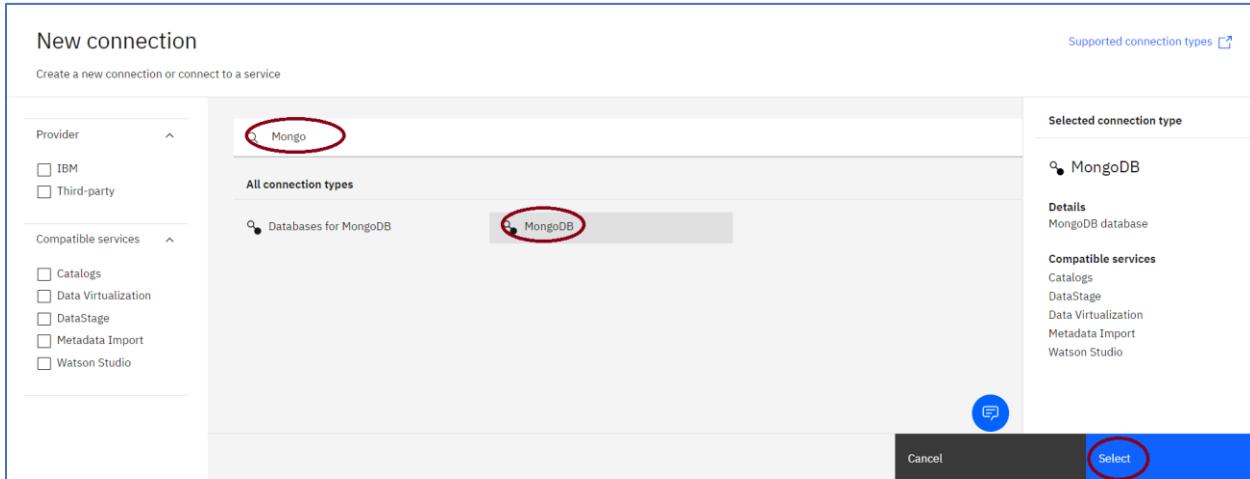
1. Click on **Connections**.

Name	Email	Role	Date added
Howard Doe	ws.catalog.user@gmail.com	Viewer	Aug 11, 2021
Adrian Doe	wsuser7000@gmail.com	Admin	Aug 11, 2021

2. Click **New connection**.

Find connections **New connection**

3. Type **Mongo** into the **Search** field, then click on **MongoDB**, then click **Select**.



4. Enter a **Trafficking_Mongo** for the **Name** of the connection.

The screenshot shows the 'Create connection: MongoDB' interface. It has a 'Connection overview' section with a 'Name' field containing 'Trafficking_Mongo'. Below it is a 'Description' section with a text input field. At the bottom is a 'Connection details' section.

5. Scroll down and copy the **Database name**, **Port number**, **Hostname**, and **Authentication database name** from the MongoDB credentials file into the corresponding fields below.

Database* ⓘ

Hostname or IP address* ⓘ

Port* ⓘ

Credentials

Authentication database ⓘ

6. Scroll down and copy **Username** and **Password** from the MongoDB Credentials file, check the **Port is SSL enabled** checkbox, and copy the **SSL certificate** from the MongoDB Credentials file, and click **Test connection**.

Create connection: MongoDB

Enter the connection information.

Test connection

Username* ⓘ

Password* ⓘ

Certificates

Port is SSL-enabled ⓘ

SSL certificate ⓘ

```
-----BEGIN CERTIFICATE-----  
MIIDZCCAfegAwIBAgIDANEH5by2/kzHMA0GCSqGSIb3DQEBCwUAMBAxHDAsBgNV  
BAMME0ICTSB0bg91ZCBEYRHyMfFzXKtwIhNNTgwNjI1MTQyOTAwWWhNMjgwNjly  
MTQyOTAwWjAeMRwwGgYDVQQDBNQJkogQ2vdWQgRGFOYWJhczVzMIBjJANBgkq  
hkiG9w0DRA0EEAAOCDA08AMTIBCaICCA0EABlaaOGycEdGaeMlm0lfMPn1OhoqvBa1
```

7. The MongoDB connection is successfully tested. If not, check the credentials to make sure they were added correctly. Click **Create**.

Enter the connection information.

✔ The test was successful.
 Click Create to save the connection information.

Password* ⓘ <input type="password" value="*****"/>	@
Certificates	
<input checked="" type="checkbox"/> Port is SSL-enabled ⓘ	
<small>SSL certificate ⓘ</small> <small>-----BEGIN CERTIFICATE-----</small> <small>MIIIDzCCAfegAwIBAgIJANEH58y2/kzHMA0GCSqGSIb3DQEBCwUAMB4xHDAABgNV</small> <small>BAMMEoIiTSBDbg91ZCBEYXRhymFzZXMuWhhNMjgNjI1MTQyOTAwWhcNMjgwNjIy</small> <small>MTQyOTAwWjAeMRwwGgYDVQQDDBNQk0gQ2xvdWQgRGF0YWJhc2VzMIIBIjANBgkq</small> <small>hkiG9w0BACFEAACAOARAMITRCakCAoFA8lnaGzxFdGneMlmoffMPnDhodRn1</small> <small>-----END CERTIFICATE-----</small>	
<small>Hostname in SSL certificate ⓘ</small> <input type="text"/>	
Back Create	

8. The **Trafficking_Mongo** connection is added.

Platform connections					Supported connection types ⓘ			
Connections		Access control						
Connected data sources								
Manage existing data source connections or create new connections that can be used across the platform.								
Filter by:	All types							
<input type="text"/> Find connections					New connection +			
Name	Type	Created by	Modified by	Last updated				
Trafficking_Mongo	MongoDB	Avril Doe	Avril Doe	Aug 21, 2021	Delete			

Step 3: Add DB2 Warehouse Connection

1. Click New connections.

Platform connections					Supported connection types ⓘ			
Connections		Access control						
Connected data sources								
Manage existing data source connections or create new connections that can be used across the platform.								
Filter by:	All types							
<input type="text"/> Find connections					New connection +			
Name	Type	Created by	Modified by	Last updated				
Trafficking_Mongo	MongoDB	Avril Doe	Avril Doe	Aug 21, 2021	Delete			

2. Click IBM, Watson Studio, DB2 Warehouse, and click Select.

Create a new connection or connect to a service

Provider: IBM

Compatible services: Watson Studio

Selected connection type: Db2 Warehouse

Details: Db2 Warehouse

Compatible services: Catalogs, DataStage, Data Virtualization, Metadata Import, Watson Studio

3. Enter Trafficking_DB2WH for the Name of the connection.

Create connection: Db2 Warehouse

Enter the connection information.

Connection overview

Name: Trafficking_DB2WH

Description: Connection description

Show more ▾

Connection details

4. Scroll down and copy **db**, **hostname**, and **port** from the DB2 Warehouses Credentials file to the corresponding fields as shown below. Note, you can use either the first or second set of credentials in the DB2 Warehouse Credentials file.

Connection details

Database* ⓘ

put in the db name

Hostname or IP address* ⓘ

put in the host name

Port ⓘ

50001

5. Scroll down and copy **username** and **password** from the DB2 Warehouse Credentials file to the corresponding fields in the panel, and check the **Port is SSL enabled** checkbox and click **Test connection**.

Create connection: Db2 Warehouse

Enter the connection information.

Test connection

Credentials

Username* ⓘ

bluadmin

Password* ⓘ

.....

Certificates

Port is SSL-enabled ⓘ

6. **The test was successful** message should be displayed. If not, check to make sure the credentials were entered correctly. Click **Create** to create the DB2 Warehouse connection.

The test was successful.
Click Create to save the connection information.

Credentials

Username* ⓘ
bluadmin

Password* ⓘ
.....

Certificates

Port is SSL-enabled ⓘ

SSL certificate ⓘ

Back Create

7. **Trafficking_DB2WH** is added to the list of connections.

Connected data sources
Manage existing data source connections or create new connections that can be used across the platform.

Filter by: All types ▾

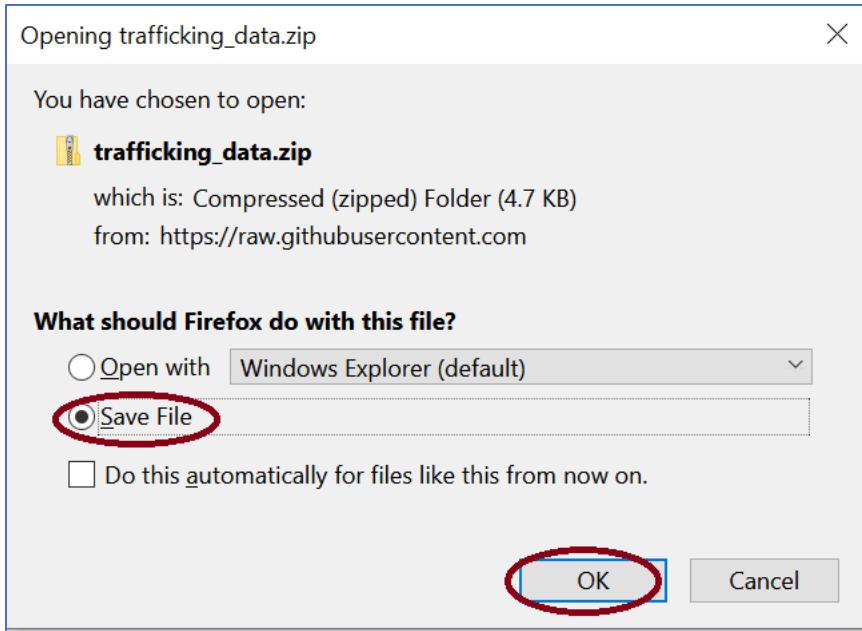
Find connections New connection +

Name	Type	Created by	Modified by	Last updated	Actions
Trafficking_DB2WH	Db2 Warehouse	Avril Doe	Avril Doe	Aug 21, 2021	
Trafficking_Mongo	MongoDB	Avril Doe	Avril Doe	Aug 21, 2021	

Download Files

Step 1: Download the [trafficking_data.zip](#) from the github repository

1. Click [here](#), to download the zipped file. Select the **Save File** option and click **OK**.



2. Navigate to where the zip file is saved. Extract the file contents. You should have two files extracted, (1) Categories.csv, and (2) Occupation.csv.

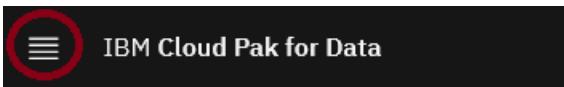
Perform Catalog Tasks

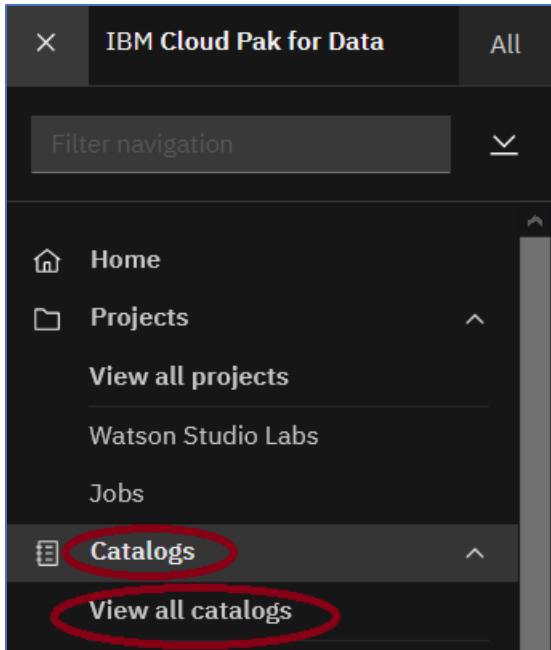
Step 1 – Create a Governed Catalog

1. Click on the **IBM Cloud Pak for Data** browser tab to return to **Cloud Pak for Data**.
2. Click on **Cloud Pak for Data** to go to the home screen.



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





4. Click on **Create Catalog**.

A screenshot of the 'Your catalogs' page. It shows a single catalog entry: 'Platform assets catalog' created by 'Admin'. Below it, there is account information: 'Account name: Avril Doe's Account', 'Creator: Avril Doe', and 'Date created: Aug 21, 2021 11:04 AM'. To the right of the catalog list is a 'Create Catalog' button, which is circled in red.

5. Enter **Watson Studio Labs** for the **Catalog name**. Select an Object Storage instance if you have more than one. 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!

Catalogs / Watson Studio Labs

Watson Studio Labs

Browse assets Access control Settings

You haven't added any assets yet.
To add a new asset, click [Add to catalog](#)

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.

Watson Studio Labs

Browse assets Access control Settings

Collaborators

Add users with roles so they can administer the catalog, add and edit catalog assets, or view and use catalog assets.

Q Who are you looking for? Show: All Add collaborators +

Name	Email	Role	Date added
<input type="checkbox"/> Horatio Doe	wsuser63000@gmail.com	Admin	May 15, 2021

2. Click on Add Collaborators.

Watson Studio Labs

Browse assets Access control Settings

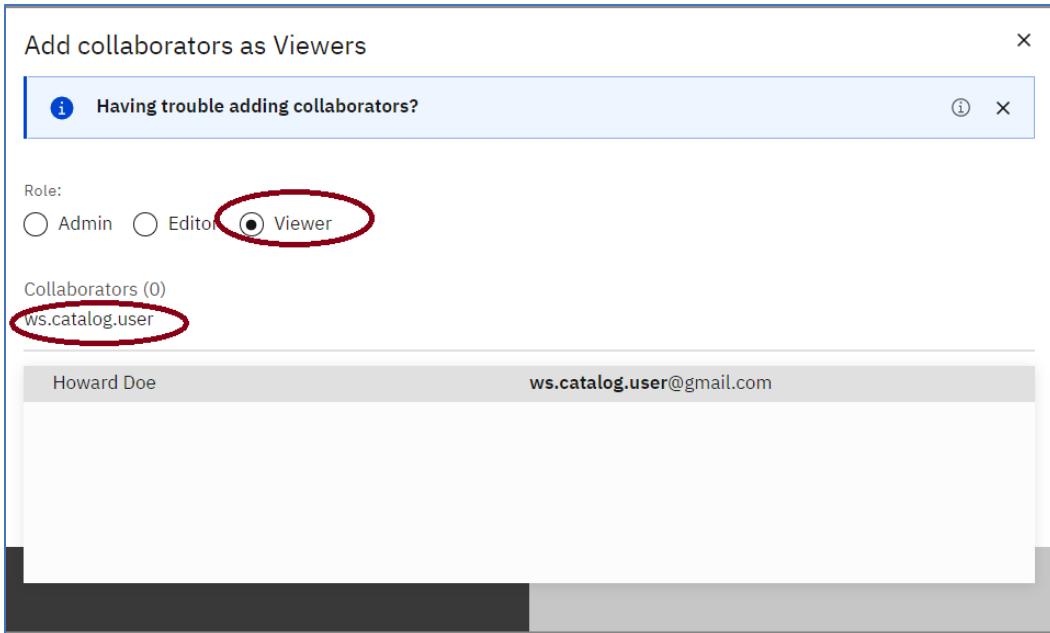
Collaborators

Add users with roles so they can administer the catalog, add and edit catalog assets, or view and use catalog assets.

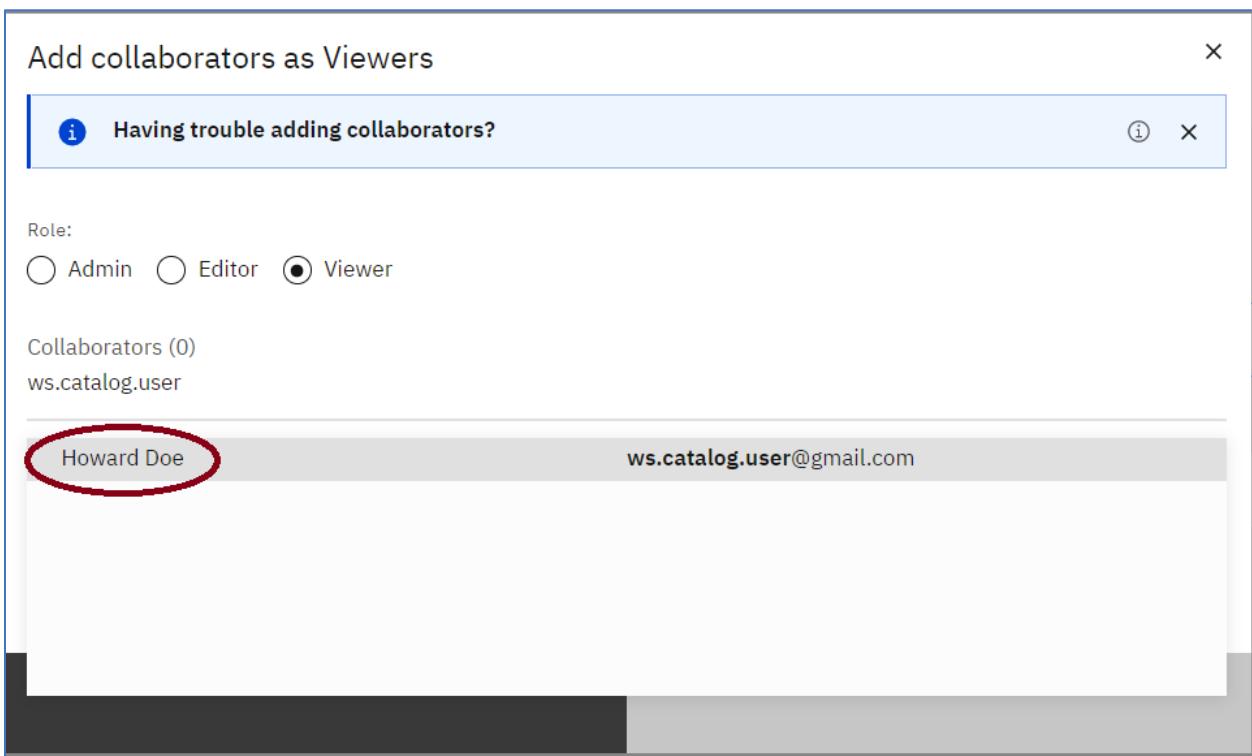
Q Who are you looking for? Show: All Add collaborators +

Name	Email	Role	Date added
<input type="checkbox"/> Horatio Doe	wsuser63000@gmail.com	Admin	May 15, 2021

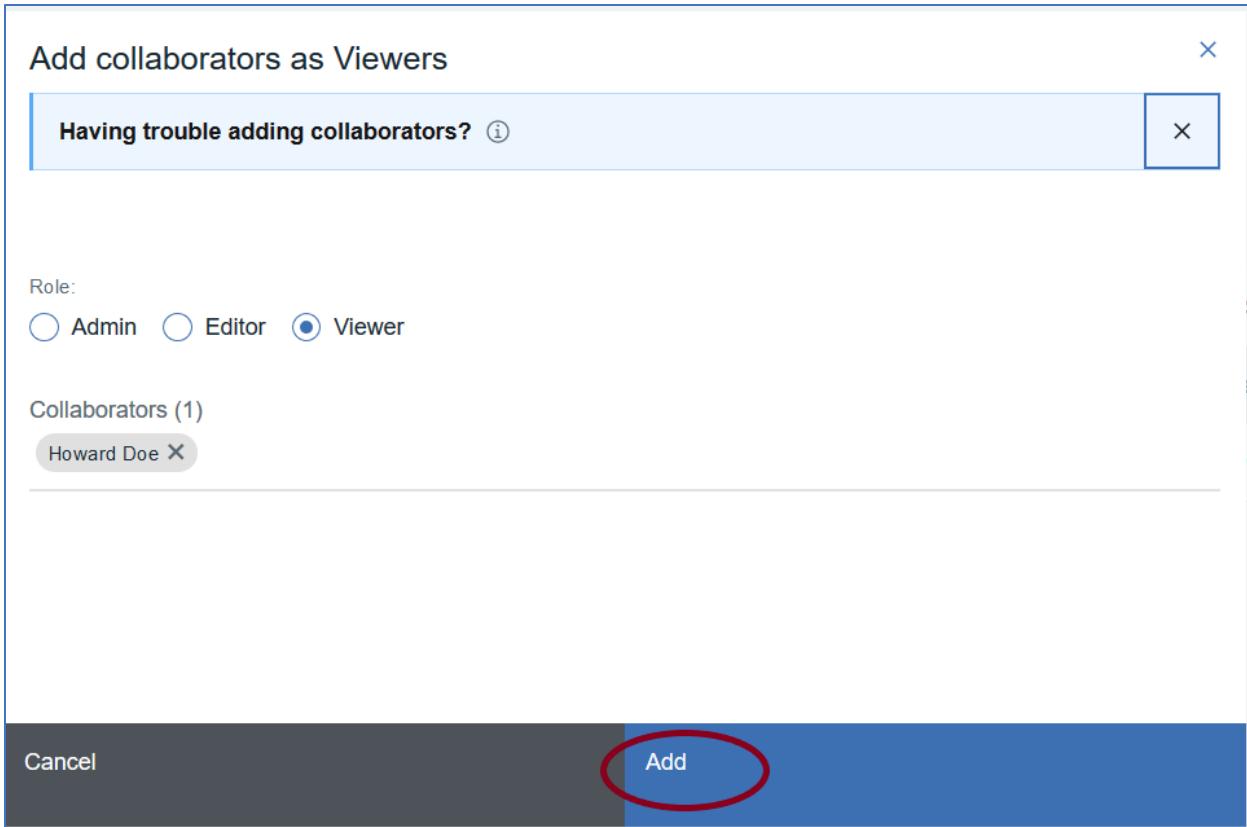
3. Leave **Viewer** as the **Role**. Start typing “ws.catalog.user@gmail.com”. The list should populate with “Howard Doe” and the ws.catalog.user@gmail.com e-mail address.



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



5. Click **Add**.



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

The screenshot shows the 'Watson Studio Labs' interface with the 'Access control' tab selected. Under the 'Collaborators' section, it says 'Add users with roles so they can administer the catalog, add and edit catalog assets, or view and use catalog assets.' A search bar 'Who are you looking for?' is present. A 'Show: All' dropdown and an 'Add collaborators +' button are also visible. The table lists two users:

Name	Email	Role	Date added
Howard Doe	ws.catalog.user@gmail.com	Viewer	Aug 21, 2021
Avril Doe	wsuser71000@gmail.com	Admin	Aug 21, 2021

Both rows are highlighted with a red border.

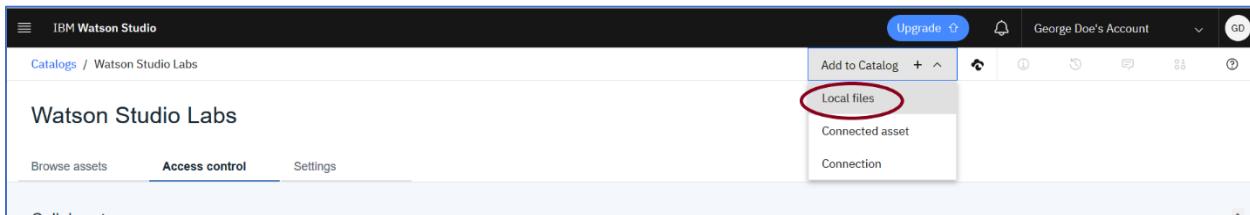
Step 3: Add Data Assets to the Catalog

In this step, we will add 6 assets: two local files, two connections, and two connected data assets. The local files we will add are the “Occupation.csv” file, and the “Categories.csv” file. We will add the two connections from the Platform access connections catalog. We will add a connected data asset that will point to the FHT_DB2WH table in the DB2 Warehouse service instance, and one that will point to the FHT_MONGO database in the Mongo DB 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 Cloud Pak for Data, for use during data preparation and within models, dashboards and notebooks as we will see in our lab exercises.

Add Local Files

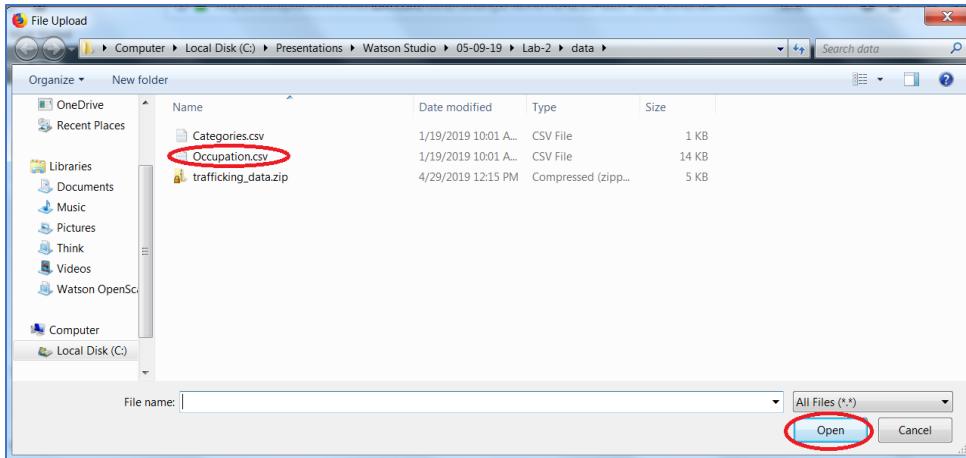
1. Click on **Add to Catalog** and click on **Local Files**.



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

A screenshot of a web page titled 'Add data assets from local files'. At the top left, it says 'Catalogs / Watson Studio Labs'. The main heading is 'Add data assets from local files'. Below that, there's a section with the heading 'Select file(s)' and a sub-instruction 'Drop your files here or browse your files to add new files (up to 5 GB each)'. A red circle highlights the word 'browse'. Below this is a dashed rectangular area with the instruction 'Drag and drop file(s) here to upload'.

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



4. Change the MIME type of the file if it is listed as Microsoft Excel. Click on **Edit name and format**.

A screenshot of a file upload interface. It says 'Select file(s)' and 'Drop your files here or browse your files to add new files (up to 5 GB each)'. Below is a dashed box labeled 'Drag and drop file(s) here to upload'. Underneath, it says 'Selected files (1)*' and shows a table with one item: 'Occupation.csv' under 'Asset name' and 'Microsoft Excel' under 'Format'. To the right of the table is a blue link 'Edit name and format' which is circled in red.

5. Replace application/vnd.ms-excel with text/csv and click **Apply**.

Add data assets from local files

Drag and drop file(s) here to upload

Selected files (1)*

Asset name	Format
Occupation.csv	text/csv

Cancel

Apply

6. The data asset is changed to CSV Format.

Add data assets from local files

Drag and drop file(s) here to upload

Selected files (1)*

Edit name and format 

Asset name	Format	X
Occupation.csv	CSV	X

- We can add tags to the newly added data asset to search for it more easily. Add two tags, enter “occupations”, click on the icon, and enter “jobs” and click on the icon. Leave the **Classification** as blank. Click on Private for the **Privacy**. Click on **Add**.

Search business terms

Tags

occupations × jobs ×

Start typing to add values

Classification

Search classifications

Privacy

Public Private

Only the asset owner and asset members can find or use the asset.

Cancel Add

- The Occupation.csv asset metadata is displayed. Click on **Expand**

Watson Studio Labs

Browse assets Access control Settings

Q What assets are you looking for?

Filter by: Any type Any tag Clear all

Showing 1 of 1 item

Name	Owner	Tags	Business terms	Type	Date added
Occupation.csv	Avril Doe	occupations jobs		Data asset	Aug 21, 2021

Expand

- Note several 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 (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 by clicking **Collapse**. At the bottom is the list of assets (shown in green).

The screenshot shows the Watson Studio Labs Catalogs interface. At the top, there is a search bar with the placeholder "What assets are you looking for?". Below the search bar are two dropdown menus: "Any type" and "Any tag", both with dropdown arrows. A red box highlights this search/filter area. Below this is a section titled "Watson recommends" with tabs for "Highly rated" and "Recently added". Under "Highly rated", there is a card for a "Data asset" named "Occupation.csv". The card includes details: Owner: George Doe, Added: Apr 26, 2020 9:37 AM, Tags: occupations, jobs, and a rating of 0 reviews. A blue box highlights this card. At the bottom, a table shows the results for "Showing 1 of 1 item". The table has columns: Name, Owner, Tags, Business terms, Type, and Date added. The single row shows "Occupation.csv" as the name, George Doe as the owner, and tags "occupations" and "jobs". The "Type" column indicates it is a "Data asset" and was added on "Apr 26, 2020". A green box highlights this table.

- Follow the same procedure (1-7) above to add the local file “Categories.csv” file to the catalog. For step 7, 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** empty, leave the **Privacy** as Public, and then click on **Add**.

The screenshot shows the asset creation dialog for "Categories.csv". The "Tags" section at the top has two tags: "job types" and "job categories", with the "job types" tag circled in red. Below this is a text input field "Start typing to add values" and a plus sign icon **⊕**. The "Classification" section below it has a search bar "Search classifications". The "Privacy" section shows "Public" selected (radio button circled in red) and a note: "All catalog members can find and use the asset.". The "Members" section at the bottom has a plus sign icon **+**. At the very bottom, there are "Cancel" and "Add" buttons, with the "Add" button circled in red.

- The Categories.csv asset is displayed

The screenshot shows the IBM Watson Studio Catalogs interface. At the top, there are several tabs and a search bar. Below the search bar, there are filters for 'Any type' and 'Any tag', and a 'Clear all' button. A 'Watson recommends' section highlights 'Categories.csv' and 'Occupation.csv'. 'Categories.csv' has an owner of George Doe, added on Apr 26, 2020 at 9:56 AM, with tags 'job types' and 'categories'. It has 0 reviews. 'Occupation.csv' also has an owner of George Doe, added on Apr 26, 2020 at 9:37 AM, with tags 'occupations' and 'jobs'. It also has 0 reviews. Below this, a table lists 'Showing 2 of 2 items' with columns for Name, Owner, Tags, Business terms, Type, and Date added. Both assets are listed under the 'Data asset' type. The bottom of the screen shows a Windows taskbar with various icons.

12. Change the name of the Occupation.csv asset to be just Occupation by clicking on the **Occupation.csv** asset.

This screenshot shows the same asset list table from the previous step. The 'Occupation.csv' row is circled in red, indicating it is the target for renaming.

13. Click the “pencil” icon adjacent to **Occupation.csv**.

This screenshot shows the details page for the 'Occupation.csv' asset. The asset name 'Occupation.csv' is displayed with a pencil icon to its right, which is circled in red. Below the name are buttons for 'Edit name', 'Remove', 'Download', 'Add to project', and a '+' icon. At the bottom of the page are tabs for Overview, Asset, Access, Review, Profile, and Activities.

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

This screenshot shows the 'Edit name' dialog for the 'Occupation.csv' asset. The input field contains the text 'Occupation'. At the bottom right of the dialog are 'Cancel' and 'Apply' buttons, with 'Apply' being circled in red. The background shows the standard asset details page with tabs for Overview, Access, Review, Profile, and Lineage.

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

The screenshot shows the Watson Studio Catalogs interface. At the top, there is a breadcrumb navigation: Catalogs / Watson Studio Labs / Occupation.csv. A red circle highlights the 'Watson Studio Labs' link. Below the navigation, there is a section titled 'DATA ASSET' with a text input field containing 'Occupation'. To the right of the input field are 'Cancel' and 'Apply' buttons. At the bottom of the screen, there are tabs for Overview, Access, Review, Profile, and Lineage. The 'Overview' tab is currently selected.

16. Follow steps 12-15 to change the Categories.csv asset name to be Categories.

Add MongoDB Connection

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

The screenshot shows the Watson Studio Catalogs interface. At the top, there is a breadcrumb navigation: Catalogs / Watson Studio Labs. To the right of the navigation, there is a 'Add to Catalog' button with a '+' icon and a dropdown menu. The dropdown menu contains options: 'Local files', 'Connected asset', and 'Connection'. A red circle highlights the 'Connection' option.

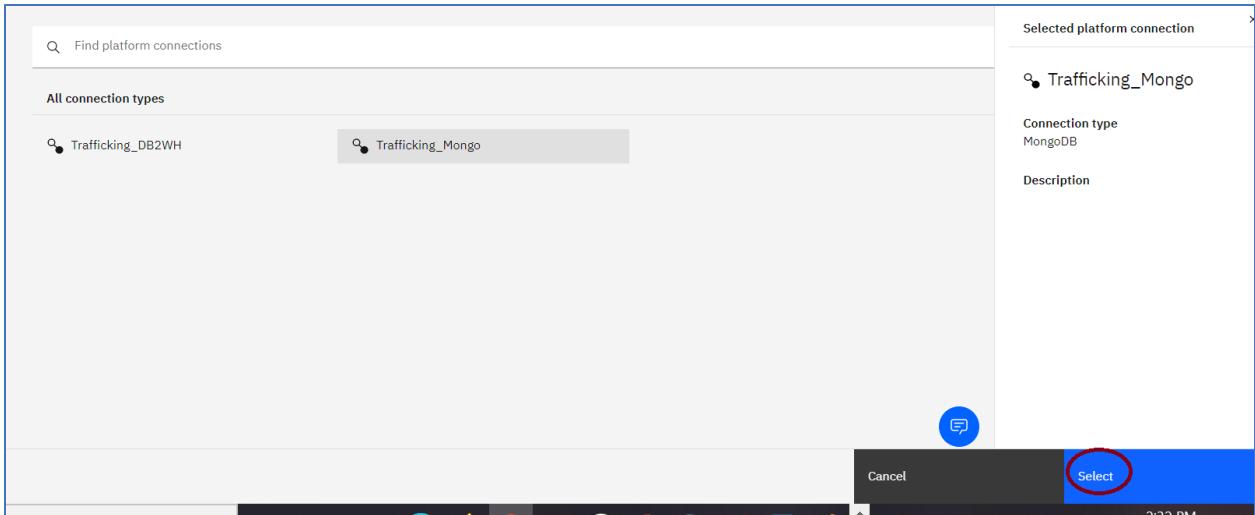
2. The Connection page is displayed with the various connectivity options. Click on **From platform**.

The screenshot shows the 'Add connection' page. At the top, there is a title 'Add connection' and a note: 'Create a new connection, select an existing connection from the list of platform connections, or connect to a service.' Below the title, there are three tabs: 'New' (highlighted), 'From platform' (circled in red), and 'To service'. On the left, there is a 'Provider' dropdown and a search bar with the placeholder 'Find connection types'. On the right, there is a 'Supported connection types' link.

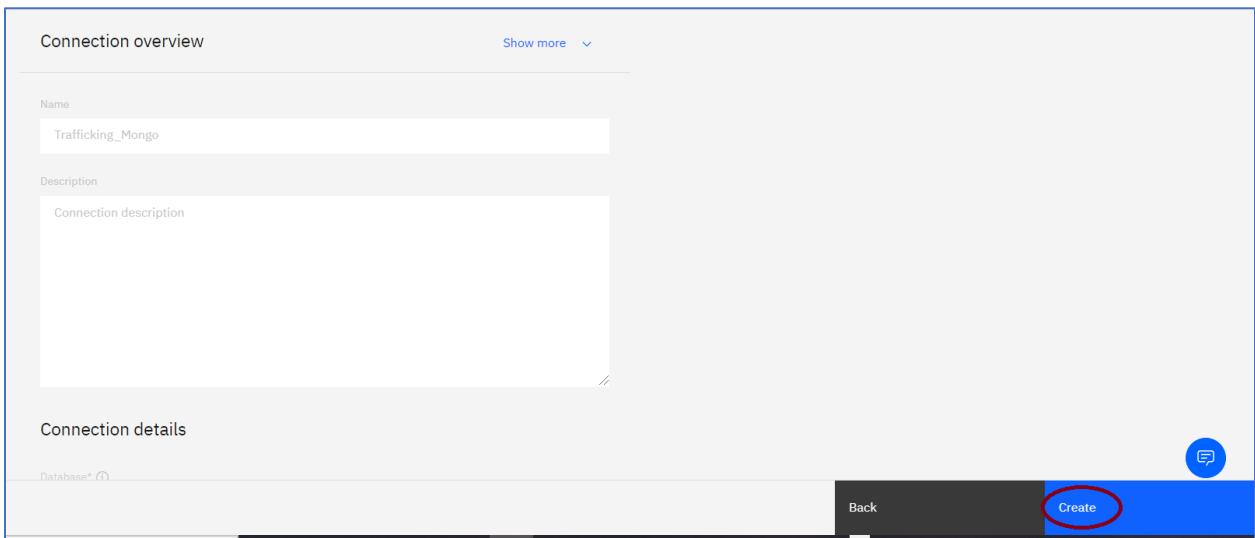
3. Click on **Trafficking_Mongo**.

The screenshot shows the 'Add connection' page with the 'From platform' tab selected. On the left, there is a 'Provider' dropdown with 'IBM' and 'Third-party' options. Below the provider dropdown is a 'Compatible services' dropdown. On the right, there is a search bar with the placeholder 'Find platform connections' and a list of connection types. The 'Trafficking_Mongo' connection type is highlighted with a red circle.

4. Click **Select**.



5. Click **Create** to add the connection.



6. The connection asset is displayed.

Watson Studio Labs

Browse assets Access control Settings

Filter by: Any type Any tag Clear all

Recently added Watson recommends Highly rated Collapse

Connection Trafficking_Mongo Owner: Avril Doe Added: Aug 21, 2021 2:36 PM ☆☆☆☆☆ 0 reviews	Data asset Categories.csv Owner: Avril Doe Added: Aug 21, 2021 2:26 PM Tags: job ty... job c... ☆☆☆☆☆ 0 reviews	Data asset Occupation Owner: Avril Doe Added: Aug 21, 2021 2:17 PM Tags: occu... jobs ☆☆☆☆☆ 0 reviews
--	---	---

Showing 3 of 3 items

Name	Owner	Tags	Business terms	Type	Date added
Categories.csv	Avril Doe	job types job cat...		Data asset	Aug 21, 2021
Occupation	Avril Doe	occupa... jobs		Data asset	Aug 21, 2021
Trafficking_Mongo	Avril Doe			Connection	Aug 21, 2021

Add DB2 Warehouse Connection

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

IBM Watson Studio

Catalogs / Watson Studio Labs

Add to Catalog + ⚙️

Watson Studio Labs

Browse assets Access control Settings

Local files
Connected asset
Connection

- The Connection page is displayed with the various connectivity options. Click on **From platform**.

Add connection

Create a new connection, select an existing connection from the list of platform connections, or connect to a service.

New **From platform** To service

Provider ▾ Find connection types

Supported connection types ↗

- Click **Trafficking_DB2WH**.

Add connection

Create a new connection, select an existing connection from the list of platform connections, or connect to a service.

New **From platform** To service

Provider ^

IBM
 Third-party

Compatible services ^

Catalogs
 Data Virtualization

Find platform connections

All connection types

Trafficking_DB2WH (circled)

Trafficking_Mongo

4. Click **Select**.

Create a new connection, select an existing connection from the list of platform connections, or connect to a service.

New **From platform** To service

Provider ^

IBM
 Third-party

Compatible services ^

Catalogs
 Data Virtualization
 DataStage
 Metadata Import
 Watson Studio

Find platform connections

All connection types

Trafficking_DB2WH (highlighted)

Trafficking_Mongo

Selected platform connection

Trafficking_DB2WH

Connection type
Db2 Warehouse

Description

Cancel **Select** (circled)

5. Click **Create** to add the connection.

Connection overview

Show more ▾

Name
Trafficking_DB2WH

Description
Connection description

Connection details

Database* (circled)

Back **Create** (circled)

6. The connection is added.

Watson Studio Labs

Browse assets Access control Settings

Recently added Watson recommends Highly rated

Showing 4 of 4 items

Name	Owner	Tags	Type	Date added
Categories.csv	Avril Doe	job types, job cat...	Data asset	Aug 21, 2021
Occupation	Avril Doe	occupa..., jobs	Data asset	Aug 21, 2021
Trafficking_DB2WH	Avril Doe		Connection	Aug 21, 2021
Trafficking_Mongo	Avril Doe		Connection	Aug 21, 2021

Add Connected Data

1. We have added a connection to the DB2 Warehouse service. We will now add a Connected Data Asset to point to the FHT_DB2WH table in the warehouse. Click on **Add to Catalog** and click on **Connected asset**.

Catalogs / Watson Studio Labs

Add to Catalog + ^

- Local files
- Connected asset** (circled in red)
- Connection

Any type Any tag Clear all

Watson recommends Highly rated Recently added

2. Click on **Select Source**.

Add asset from connection

Source* Select source

Name* Asset name

Description

3. Click on the **Trafficking_DB2WH** connection, click on the **TRAFFICKING** Schema name click on **FHT_DB2WH**, and then click **Select**.

4. Enter female_human_trafficking for **Name**, enter “trafficking” for **Tags** , click the **+** icon, enter “female human trafficking for **Tags**, click the **+** icon.

Name*
female human trafficking

Description

Business terms

Tags
trafficking X female human traffi... X

Start typing to add values +

5. Leave **Classification** unfilled. 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.

Classification ⓘ

Search classifications

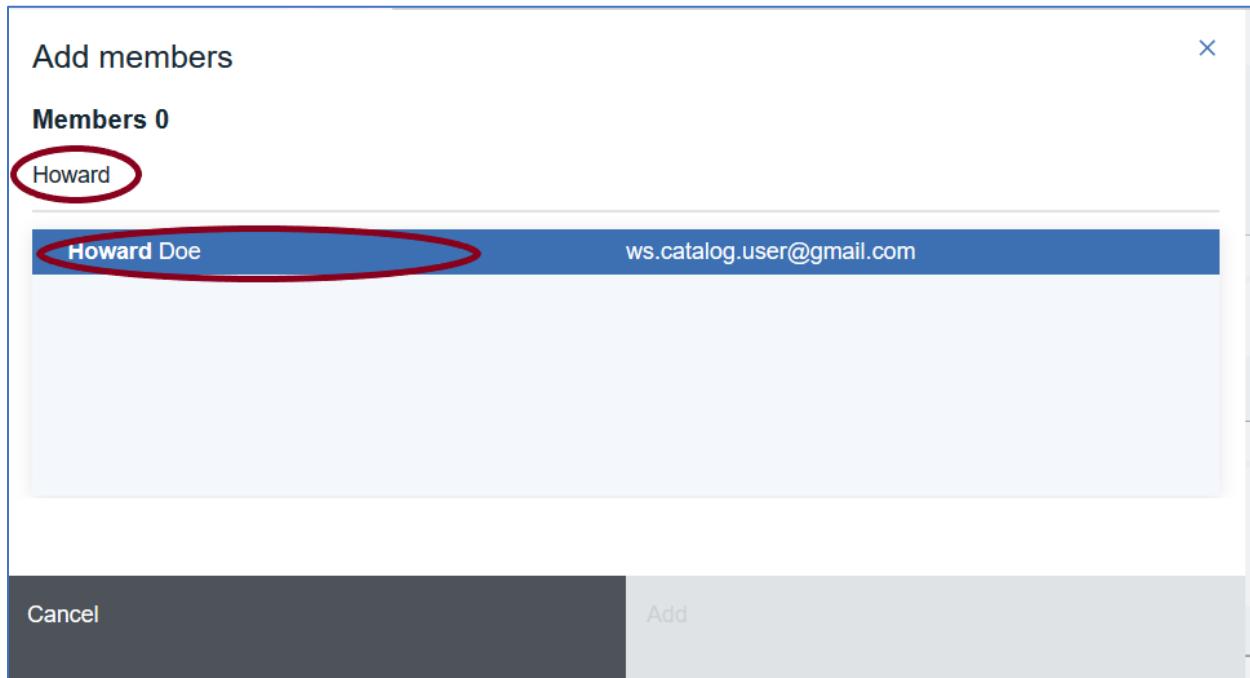
Privacy

Public Private

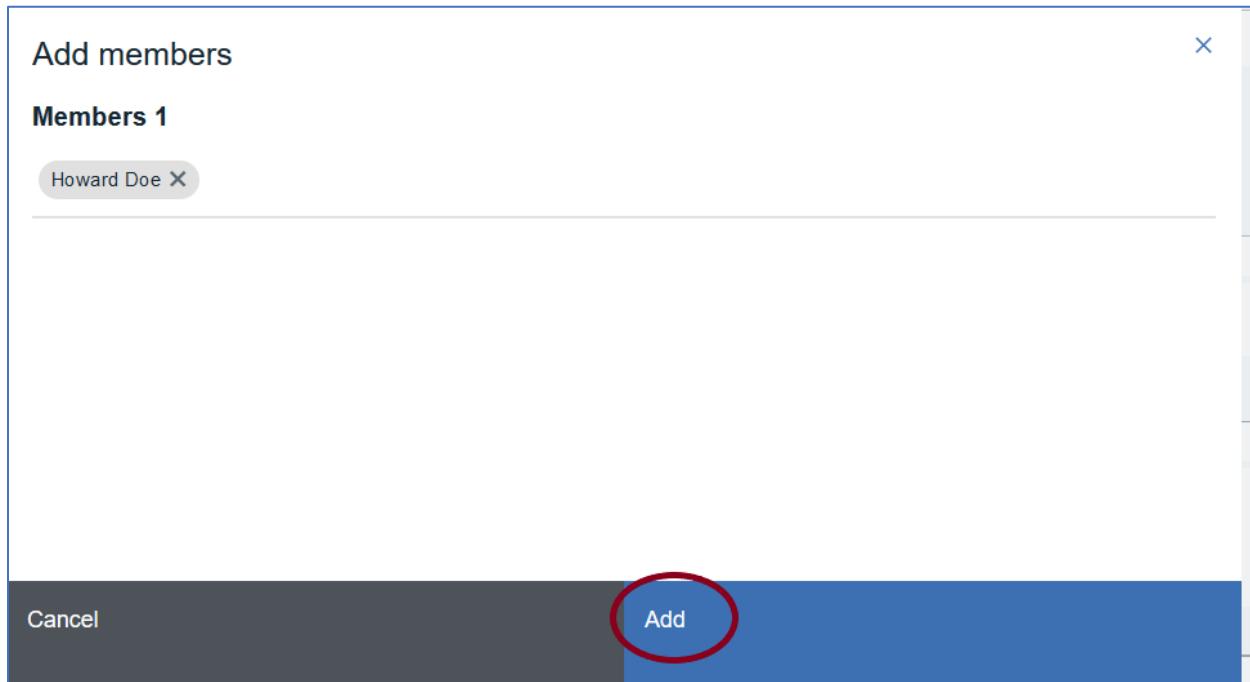
Only the asset owner and asset members can find or use the asset.

Members +

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



7. Click on **Add** to add Howard Doe to the access control list for the female_human_trafficking asset.



8. Click on **Add**.

Name*

Description

Business terms

Search business terms

Tags

trafficking X female human traffi... X

Start typing to add values +

Classification* ⓘ

None

Cancel Add

9. The female_human_trafficking Connected Data asset has been added to the catalog.

Watson Studio Labs

Browse assets Access control Settings

Data asset	Connection	Connection	Data asset	Data asset
female human trafficking	Trafficking_DB2WH	Trafficking_Mongo	Categories.csv	Occupation
Owner: Avril Doe Added: Aug 21, 2021 5:42 PM Tags: traff... fema...	Owner: Avril Doe Added: Aug 21, 2021 4:36 PM	Owner: Avril Doe Added: Aug 21, 2021 2:36 PM	Owner: Avril Doe Added: Aug 21, 2021 2:26 PM Tags: job ty... job c...	Owner: Avril Doe Added: Aug 21, 2021 2:17 PM Tags: occup... jobs
☆☆☆☆☆ 0 reviews	☆☆☆☆☆ 0 reviews	☆☆☆☆☆ 0 reviews	☆☆☆☆☆ 0 reviews	☆☆☆☆☆ 0 reviews

Showing 5 of 5 items

Name	Owner	Tags	Business terms	Type	Date added
Categories.csv	Avril Doe	job types job categor...		Data asset	Aug 21, 2021
Occupation	Avril Doe	occupations jobs		Data asset	Aug 21, 2021
Trafficking_DB2WH	Avril Doe			Connection	Aug 21, 2021
Trafficking_Mongo	Avril Doe			Connection	Aug 21, 2021
female human trafficking	Avril Doe	trafficking female hu...		Data asset	Aug 21, 2021

10. The next step is to add Connected data from the Trafficking_Mongo connection. Click **Add to catalog** and then **Connected asset**.

The screenshot shows the Watson Studio Labs Catalogs interface. At the top, there are filters for 'Any type' and 'Any tag', and buttons for 'Watson recommends', 'Highly rated', and 'Recently added'. On the right, there's a toolbar with 'Add to Catalog' (with a plus sign), a refresh icon, and three other icons. A dropdown menu is open, showing 'Local files', 'Connected asset' (which is circled in red), and 'Connection'.

11. Click Select source.

The screenshot shows the 'Add asset from connection' dialog. It has two main sections: 'Source*' on the left and 'Name*' on the right. The 'Source*' section contains a button labeled 'Select source' which is circled in red. The 'Name*' section has fields for 'Asset name' and 'Description'.

12. Click on Trafficking_Mongo, FHT_MONGO, Activity, and then click Select.

The screenshot shows the 'Select connection source' dialog. It displays a tree structure of connections. In the first column, 'Trafficking_Mongo' is selected and highlighted with a red oval. In the second column, 'FHT_MONGO' is selected and highlighted with a red oval. In the third column, 'ACTIVITY' is selected and highlighted with a red oval. On the right, a 'Selected assets' panel shows a single asset named 'ACTIVITY' with details like 'Asset name: ACTIVITY' and 'Asset type: Table'. At the bottom right of the dialog, there are 'Cancel' and 'Select' buttons, with 'Select' being circled in red.

13. Enter fht_arrival_departure for the Name. Enter “trafficking” for Tags , click the + icon, enter “female human trafficking” for Tags, click the + icon, enter “arrival” for Tags, click the + icon, and enter “departure” for Tags, and click the + icon.

Name*
fht_arrival_departure

Description

Business terms

Search business terms

Tags

trafficking X female human traffi... X arrival X
departure X

14. Scroll down and leave **Classification** blank, click on **Private** for **Privacy**, and click the “+” icon to the right of **Members**.

Classification ⓘ

Search classifications

Privacy

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

Members

15. Enter Howard, and the search should locate Howard Doe.

Add members

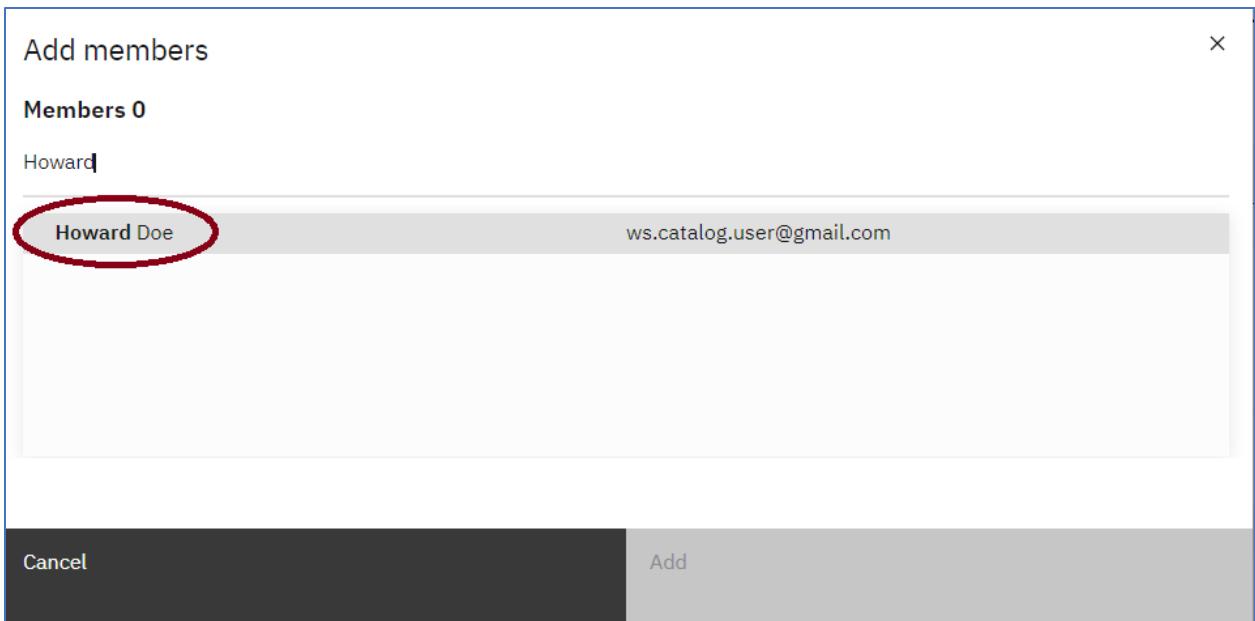
Members 0

Howard

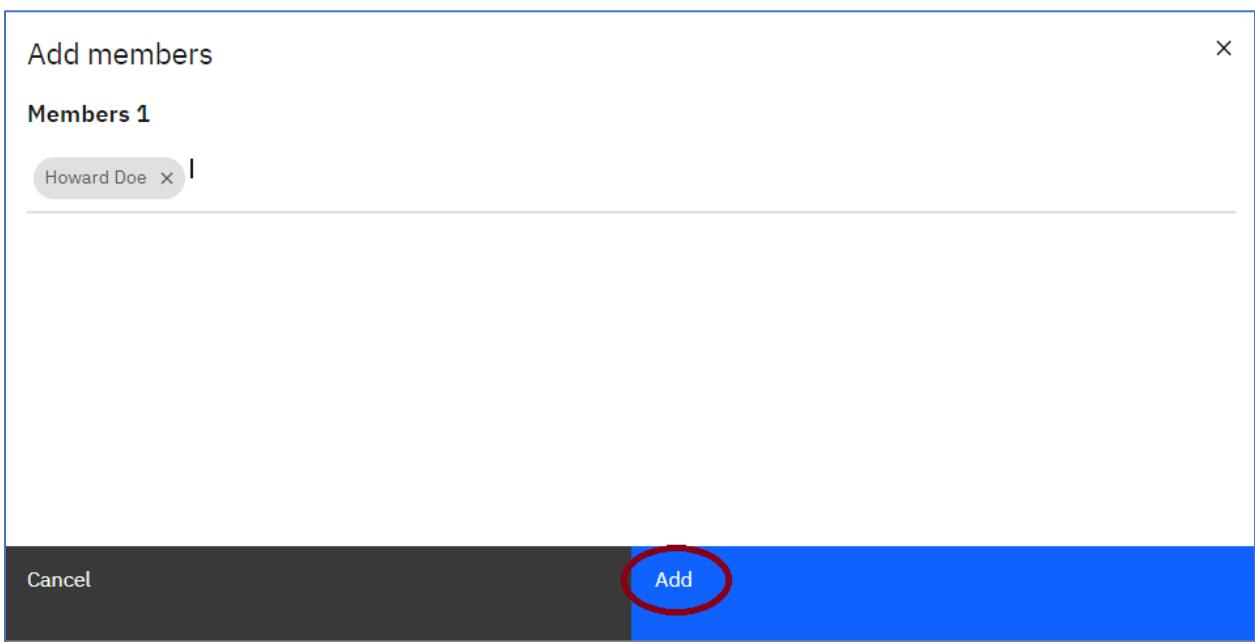
Howard Doe ws.catalog.user@gmail.com

Cancel Add

16. Click on the **Howard Doe** entry in the list.



17. Click **Add** to add Howard Doe to the access control list for this asset.



18. Click **Add**.

The screenshot shows the Watson Studio Catalog interface. At the top left, there's a 'Classification' section with a search bar. Below it is a 'Privacy' section with 'Public' and 'Private' radio buttons, where 'Private' is selected. Under 'Members', there's a 'Name' input field and a 'Cancel' button. To the right of the input field is a blue 'Add' button, which is circled in red.

19. fht_arrive_depart is added to the catalog.

The screenshot shows the Watson Studio Catalog interface with the title 'Watson Studio Labs'. It displays a grid of assets. One asset, 'fht_arrive_depart', is highlighted with a red box. In the bottom panel, there's a table listing assets with columns for Name, Owner, Tags, Business terms, Type, and Date added. The 'fht_arrive_depart' asset is listed at the bottom of the table, also highlighted with a red box.

Name	Owner	Tags	Business terms	Type	Date added
Categories.csv	Avril Doe	job types, job categor...		Data asset	Aug 21, 2021
Occupation	Avril Doe	occupations, jobs		Data asset	Aug 21, 2021
Trafficking_DB2WH	Avril Doe			Connection	Aug 21, 2021
Trafficking_Mongo	Avril Doe			Connection	Aug 21, 2021
female human trafficking	Avril Doe	trafficking, female hu...		Data asset	Aug 21, 2021
fht_arrive_depart	Avril Doe	trafficking, female hu... + 2 more		Data asset	Aug 21, 2021

Step 4: 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.

1. You should be at the Catalog home screen. If not, click on **Watson Studio Labs** to go to the Catalog home screen.

The screenshot shows the IBM Watson Studio navigation bar. The 'Catalogs' and 'Watson Studio Labs' links are highlighted with a red circle. Below the navigation bar is a secondary menu with 'Browse Assets', 'Access Control', and 'Settings'.

2. Click on the checkbox to the left of the **Name** in the bottom panel to select all the assets.

Name	Owner	Tags	Business terms	Type	Date added
Categories.csv	Avril Doe	job types job categor...		Data asset	Aug 21, 2021
Occupation	Avril Doe	occupations jobs		Data asset	Aug 21, 2021
Trafficking_DB2WH	Avril Doe			Connection	Aug 21, 2021
Trafficking_Mongo	Avril Doe			Connection	Aug 21, 2021
female human trafficking	Avril Doe	trafficking female hu...		Data asset	Aug 21, 2021
fht_arrival_departure	Avril Doe	trafficking female hu... + 2 more		Data asset	Aug 21, 2021

3. Click Add to project.

Add to project		Remove	6 of 6 items selected	Cancel		
<input checked="" type="checkbox"/>	Name	Owner	Tags	Business terms	Type	Date added
<input checked="" type="checkbox"/>	Categories.csv	Avril Doe	job types job categor...		Data asset	Aug 21, 2021
<input checked="" type="checkbox"/>	Occupation	Avril Doe	occupations jobs		Data asset	Aug 21, 2021
<input checked="" type="checkbox"/>	Trafficking_DB2WH	Avril Doe			Connection	Aug 21, 2021
<input checked="" type="checkbox"/>	Trafficking_Mongo	Avril Doe			Connection	Aug 21, 2021
<input checked="" type="checkbox"/>	female human trafficking	Avril Doe	trafficking female hu...		Data asset	Aug 21, 2021
<input checked="" type="checkbox"/>	fht_arrival_departure	Avril Doe	trafficking female hu... + 2 more		Data asset	Aug 21, 2021

4. Select the Watson Studio Labs as the Target. Click on Add.

Add to project

Target* Watson Studio Labs

Selected assets (6)

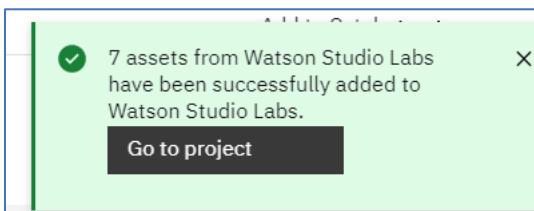
Asset name	Catalog	Connection
Categories.csv	Watson Studio Labs	Watson Studio Labs_DataCatalog
Occupation	Watson Studio Labs	Watson Studio Labs_DataCatalog
Trafficking_DB2WH	Watson Studio Labs	
Trafficking_Mongo	Watson Studio Labs	
female human trafficking	Watson Studio Labs	Trafficking_DB2WH

Connections to be added (2)

- Trafficking_DB2WH Watson Studio Labs
- Trafficking_Mongo Watson Studio Labs

Add

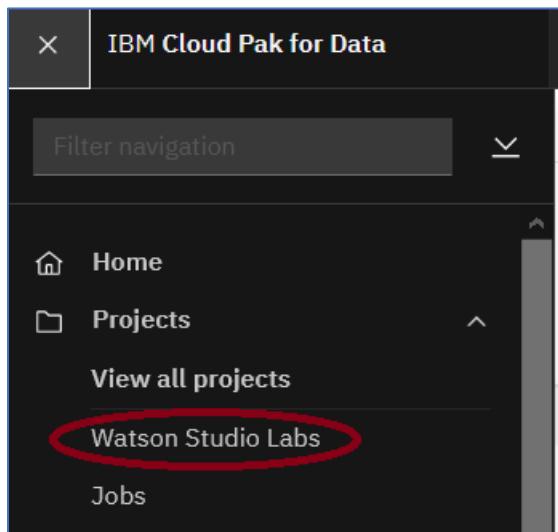
5. Wait for the message that 7 assets have been added to the project in the top right. This corresponds to the 6 assets that are listed plus a connection to object storage where the catalog stores the uploaded local files.



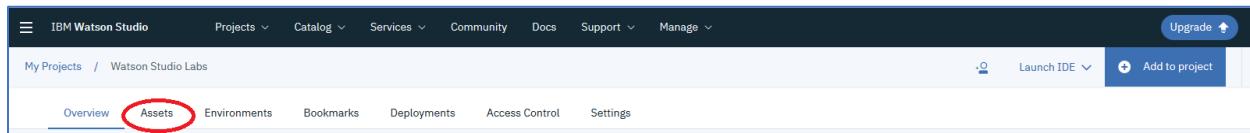
6. Click on the hamburger icon ☰.



7. Click on Watson Studio Labs under Projects.



8. Click on the Assets tab.



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

A screenshot of the IBM Watson Studio interface, specifically the "Data assets" section. It shows a list of assets with columns for Name, Type, Created by, and Last modified. The list includes several CSV files and connections, all of which are highlighted with a red border. The "Assets" tab from the previous screenshot is also highlighted with a red border.

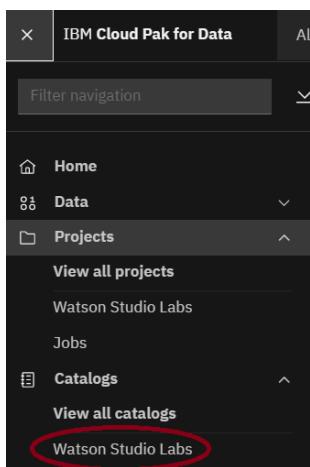
Step 5: Search the Catalog

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 the hamburger icon .

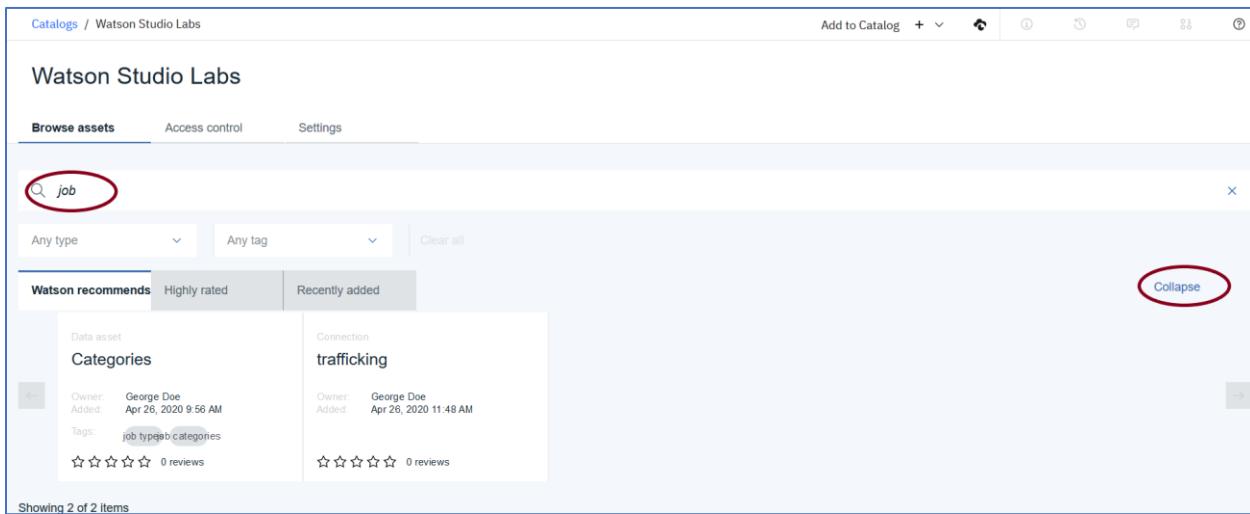


2. Click on **Watson Studio Labs** under the **Catalogs** menu.



The screenshot shows the navigation menu of IBM Cloud Pak for Data. The 'Catalogs' section is expanded, showing 'View all catalogs' and 'Watson Studio Labs'. The 'Watson Studio Labs' option is circled in red.

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



The screenshot shows the Watson Studio Labs catalog search results for the term 'job'. The search bar at the top contains 'job'. A 'Collapse' button is circled in red on the right side of the search results. The results show two items: a 'Data asset' named 'Categories' and a 'Connection' named 'trafficking'. Both items have a star rating of 0 reviews.

Data asset	Connection
Categories Owner: George Doe Added: Apr 26, 2020 9:56 AM Tags: job typeab categories 0 reviews	trafficking Owner: George Doe Added: Apr 26, 2020 11:48 AM 0 reviews

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

Showing 2 of 2 items

Name	Owner	Tags	Business terms	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

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

What assets are you looking for?

Any type

Connection

Data asset

Show Data asset 4 items

6. The search results list the 4 Data Assets. The 2 Connected Data asset, and the 2 uploaded files. Note: Other asset types that can be catalogued include, Connections, Models, Notebooks, and Dashboards.

What assets are you looking for?

Type

Data asset

Showing 3 of 3 items

Name	Owner	Tags	Business terms	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, male human trafficking		Data asset	Apr 26, 2020

7. 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 X

Showing 3 of 3 items

Name Categories Occupation

female human... job categories job types jobs occupations trafficking

8. The Categories Data Asset is listed.

What assets are you looking for?

Type Tag

Clear all

Data asset X job types X

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 6: 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**.

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 filters: "Type" and "Tag", each with a clear button. A red circle highlights the "Clear all" button. Below the filters, there are two buttons: "Data asset" and "job types". The main area displays a table titled "Showing 1 of 1 item". The table has columns: Name, Owner, Tags, Business terms, Type, and Date added. The single item listed is "female_human_trafficking" owned by "George Doe" (represented by a blue profile icon). The "Tags" column shows "job types" and "job categories". The "Type" column indicates it is a "Data asset" added on "Apr 26, 2020".

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

2. Click on the female_human_trafficking asset.

The screenshot shows a list of assets in a table format. The columns are: Name, Owner, Tags, Business terms, Type, and Date added. There are six rows of data:

Name	Owner	Tags	Business terms	Type	Date added
Trafficking_DB2WH	Avril Doe			Connection	Aug 21, 2021
Trafficking_Mongo	Avril Doe			Connection	Aug 21, 2021
Categories.csv	Avril Doe	job types, job cat...		Data asset	Aug 21, 2021
Occupation	Avril Doe	occupa..., jobs		Data asset	Aug 21, 2021
fht_arrival_departure	Avril Doe	traffick..., female ... + 2 more		Data asset	Aug 21, 2021
female human trafficking	Avril Doe	traffick..., female ...		Data asset	Aug 21, 2021

3. The metadata and data contents of the asset are displayed as shown below. Toward the top of the screen, are 6 tabs, highlighted in blue below. The panel shown below the 6 tabs corresponds to the **Overview** tab. The **Asset** tab displays the contents. 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 categorized into one of many classifiers that are built into the platform. This capability will be shown below. Finally, the **Activities** 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.

The screenshot shows the IBM Cloud Pak for Data catalog interface. A specific asset, 'female_human_trafficking', is selected. The top right corner features a green oval highlighting the 'Remove' and 'Download' buttons, and the 'Add to project' button below them. On the left, there's a sidebar with 'Tags' (circled in red), 'Reviews' (with 0 reviews), and a 'Connection' section showing 'Source: trafficking' and 'Source type: Db2 Warehouse'. The main panel displays the asset's details under tabs: Overview, Asset, Access, Review (which is currently selected and circled in red), Profile, and Activities. Under the Overview tab, it shows the asset was added on Jan 09, 2021, at 6:01 PM, has a size of 324 KB, and is in application/octet-stream format. It also lists 'Tags' (female human trafficking, trafficking), 'Business terms' (none), and 'Classifications' (none). The 'Related assets' section is empty.

The actions circled in green will **Remove** the asset from the catalog, or add the asset to a catalog enabled project (**Add to Project**). The **Name** of the data asset, the **Format**, the **Description**, or the **Tags** can be edited by clicking on the icon. Clicking on the icon next to the **Business Terms** or the **Classification** will bring up a list from which you can select entries.

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

This screenshot shows the same asset details as the previous one, but the 'Review' tab is now highlighted with a red circle. The rest of the interface remains the same, including the green oval highlighting the top right buttons.

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

Data asset
female_human_trafficking

Overall rating
0.0
★ ★ ★ ★ ★ 0 reviews

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

Your review

JD Jack Doe Jan 10, 2021
★★★★★
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

Overall rating
4.0
★★★★★ 1 review

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

Your review

JD Jack Doe Jan 10, 2021
★★★★★
Good data set for demonstrating data profiling capabilities

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

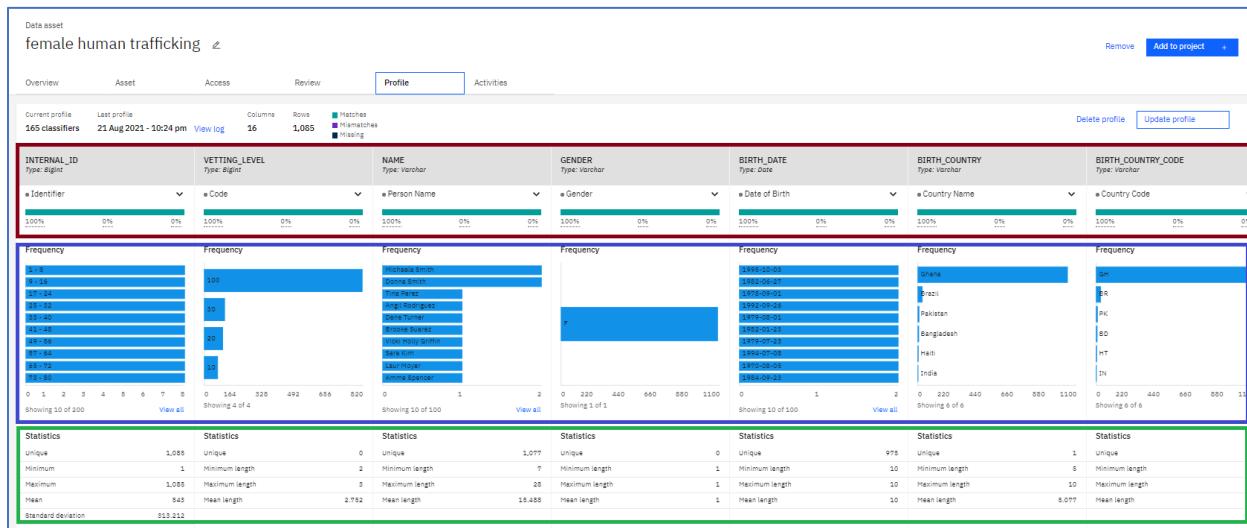
Create Profile

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

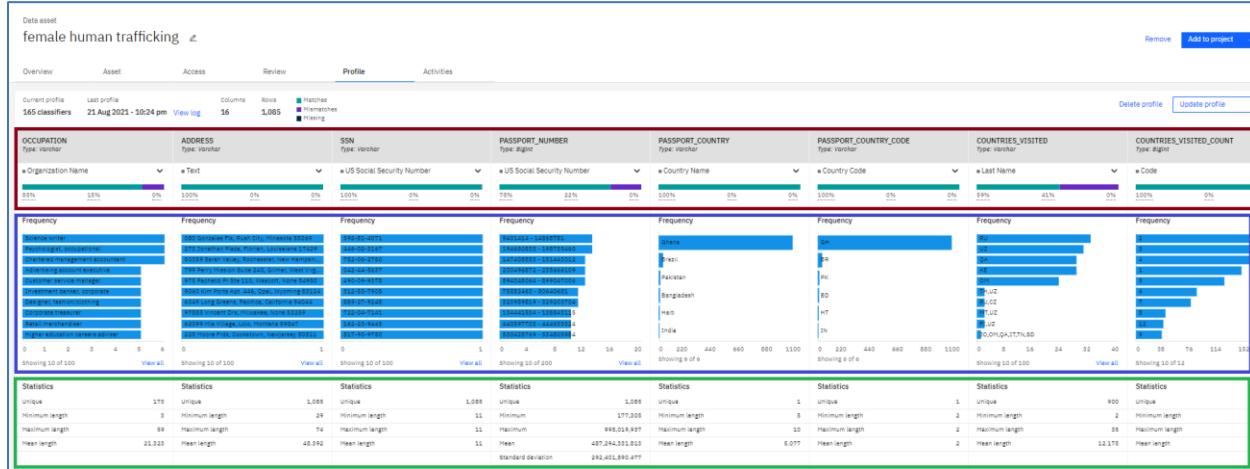
9. The profile information is displayed below using several screenshots, based on horizontal scrolling, to display many of the columns of the data asset. The highlighted red box contains the inferred column classifications. The highlighted blue box contains counts of the number of occurrences of each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



The classifications of the columns are listed below:

- INTERNAL_ID – Identifier
 - VETTING_LEVEL – Code
 - 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** – Organization Name – there is not a classifier for occupation. It incorrectly classifies the Occupation as an Organization Name. We can correct this as shown below.
- **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** – incorrectly classifies as a Last Name. We can correct this as shown below.
- **COUNTRY_VISITED_COUNT** – Code – this is an error. It's not a code. This can be re-classified if necessary.

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_NUMER** 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**.

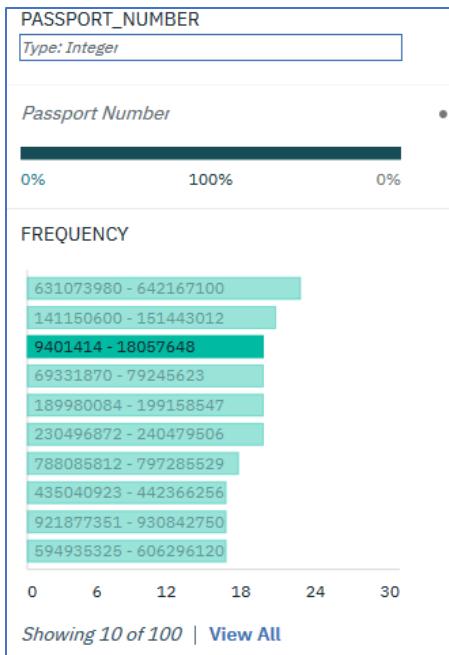
Edit data class

All data classes

- US Social Security Number 78% matches
- Passport Number

Cancel Add

12. The column classification has been changed.



13. Note, follow the above instructions to correct OCCUPATION (make it text), and COUNTRIES_VISITED (make it text).

Step 7: 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:

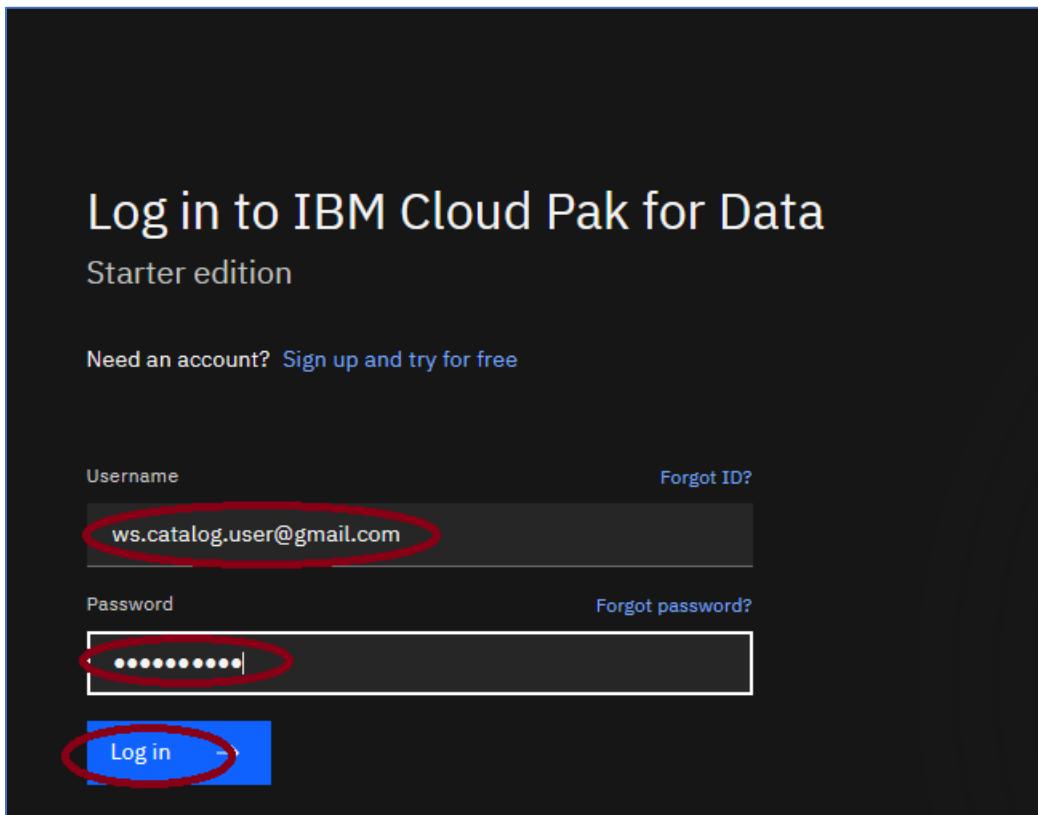
1. Catalog owner (logged on user)
2. ws.catalog.user@gmail.com

Data Asset Privacy:

1. female_human_trafficking (Private) ACL Members:
 - a. asset owner (logged on user)
 - b. ws.catalog.user@gmail.com
2. fht_arrive_depart – (Private) ACL Members:
 - a. asset owner (logged on user)
 - b. ws.catalog.user@gmail.com
3. Occupations (Private) ACL Members:
 - a. asset owner (logged on user)
4. Categories (Public)
5. Trafficking_DB2WH (Connection) (Public)
6. Trafficking_Mongo (Connection)(Public)

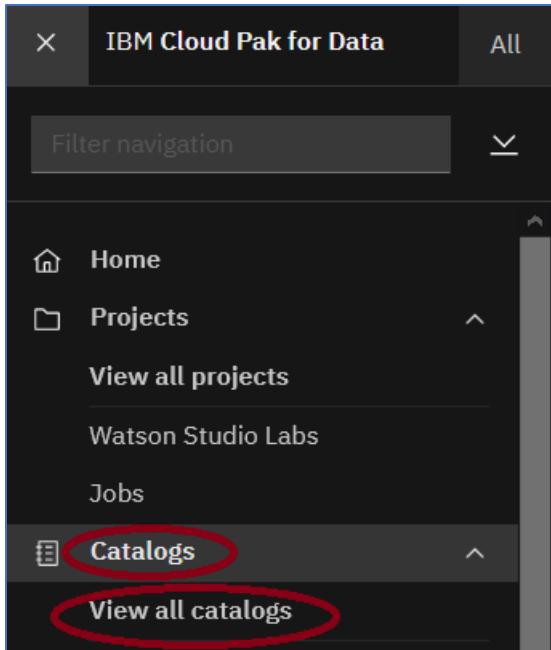
We would expect that if we log in as ws.catalog.user@gmail.com, that only 5 Data Assets would appear. The two public Data Assets, and the Private Data Asset where ws.catalog.user@gmail is on the access control list. This user would not have access to the Occupations asset, since it is a Private asset, 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.cloud.ibm.com** for the URL to navigate to the Cloud Pak for Data sign in.
3. Login using ws.catalog.user@gmail.com for the **Username** and **wsPassw0rd** for the **Password**, and click **Log in**.



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





5. 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 2 catalogs per lab participant (Watson Studio Labs, and Platform access). Click on the **Watson Studio Labs** catalog corresponding to your user name.

The screenshot shows the "Your catalogs" page in IBM Watson Studio. It lists one catalog: "Watson Studio Labs". The catalog details are: Account Name: IBM, Creator: DSX56000 User, Date created: Jan 20, 2019 5:33 PM. A red circle highlights the catalog name "Watson Studio Labs".

6. The Watson Studio Labs catalog is displayed. Five Data Assets are listed as expected.

The screenshot shows the "Watson Studio Labs" catalog page. At the top, there are tabs for "Browse assets", "Access control", and "Settings". Below that, a table lists five data assets. The table has columns for "Name", "Owner", "Tags", "Business terms", "Type", and "Date added". The first asset, "Categories.csv", is highlighted with a red box. The data assets listed are:

Name	Owner	Tags	Type	Date added
Categories.csv	Avril Doe	job types, job categ...	Data asset	Aug 21, 2021
Trafficking_DB2WH	Avril Doe		Connection	Aug 21, 2021
Trafficking_Mongo	Avril Doe		Connection	Aug 21, 2021
female human trafficking	Avril Doe	trafficking, female h...	Data asset	Aug 21, 2021
fht_arrival_departure	Avril Doe	trafficking, female h... + 2 more	Data asset	Aug 21, 2021

7. Leave this browser at this screen for Step 8 below.

Step 8: 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 tries to access the female_human_trafficking data asset.

1. **RETURN TO THE BROWSER WHERE YOUR Cloud Pak for Data id is logged in.**
2. Click on the icon. Click on **Governance** and then **Policies**.

The screenshot shows the left sidebar of the IBM Cloud Pak for Data interface. At the top is a dark header bar with the IBM Cloud Pak for Data logo and a search bar. Below it is a light gray navigation bar with a 'Filter navigation' dropdown and a 'View all projects' link. The main sidebar has several sections:

- A collapsed 'Watson Studio Labs' section.
- A 'Jobs' section.
- A 'Catalogs' section with a 'View all catalogs' link, which also includes 'Watson Studio Labs'.
- A 'Governance' section (circled in red) containing 'Task inbox', 'All drafts', 'Categories', 'Business terms', 'Classifications', 'Data classes', and 'Reference data'.
- A 'Policies' section (circled in red) containing 'Rules'.

3. Click on **Add Policy** and then **New Policy**.

The screenshot shows the 'Policies' page. At the top is a header bar with the IBM Cloud Pak for Data logo, a search bar, and account information for 'Horatio Doe's Account'. Below the header is a table with two rows:

Published	
Policies	

The 'New Policy' button is highlighted with a red circle. A tooltip for the button says 'Add policy'.

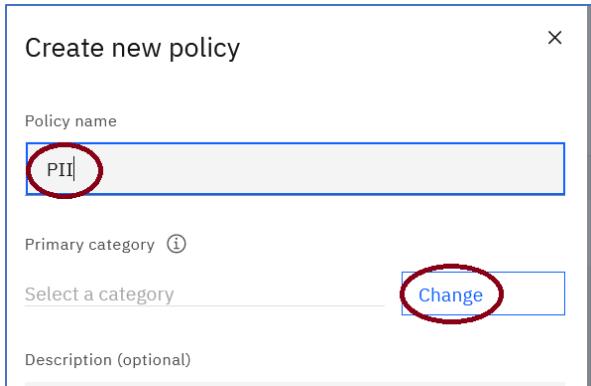
4. Type in PII for the Policy **Name**, click **Change** to select a Category.

Create new policy X

Policy name
 PII

Primary category (1)
Select a category Change Change

Description (optional)

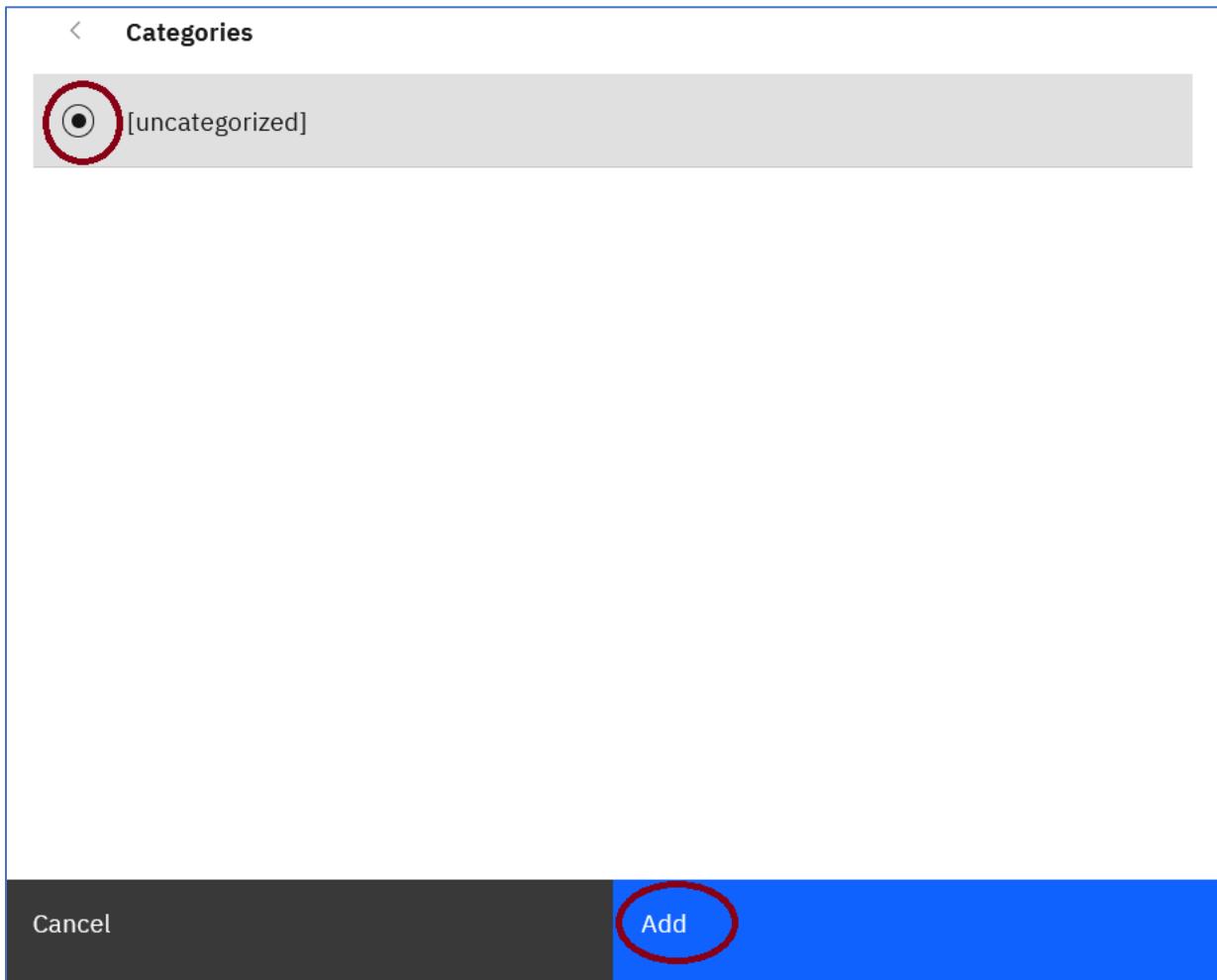


5. Click on **uncategorized**, and then click **Add**

< Categories

[uncategorized]

Cancel Add



6. Type in “Limit access to PII data” for the Policy **Description** and click **Save as draft**.

Create new policy

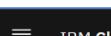
Policy name
PII

Primary category ① [uncategorized] Change

Description (optional)
Limit access to PII Data

Cancel Save as draft

7. Click Publish.

 IBM Cloud Pak for Data All Search Upgrade

8. Click Publish

Publish?

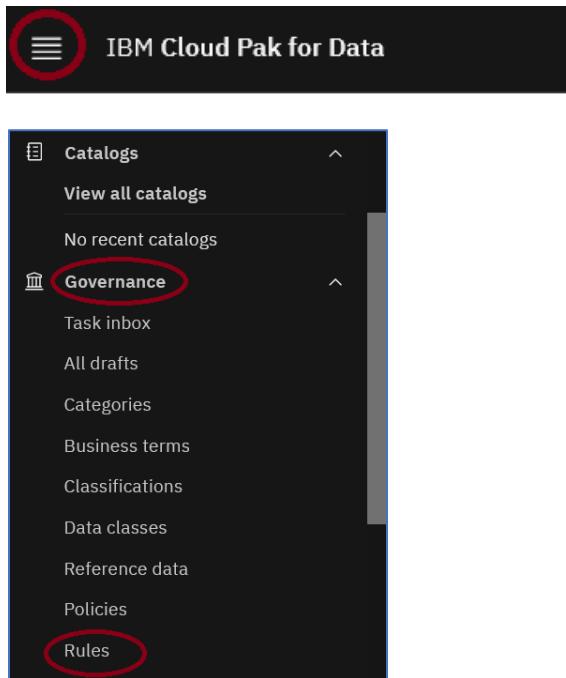
Leave a comment (optional)

Write a comment...

Cancel

Publish

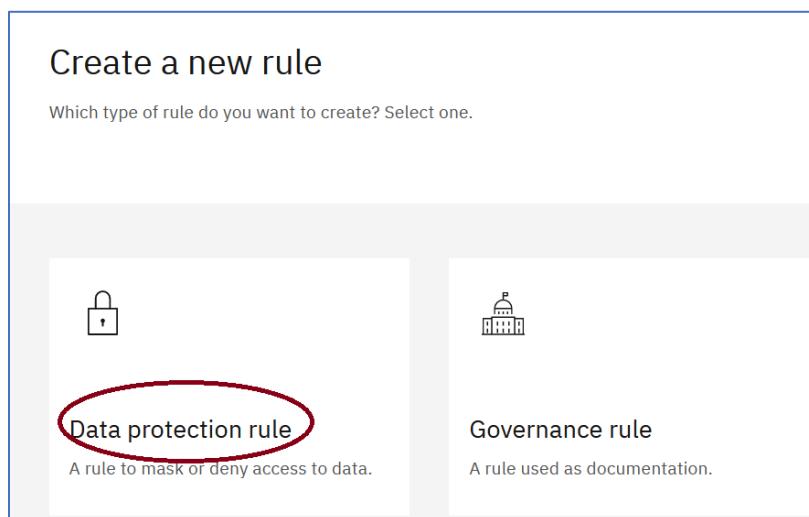
9. Click on the  icon. Click on **Governance** and then **Rules**.



10. Click on **Add rule** and then **New rule**.



11. Click **Data protection rule**.



12. 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 User name contains any
ws.catalog.user@gmail.com

then mask data in columns containing Data class
US Social Security Number Passport Number

Click Redact

Note that when typing the values **Howard Doe**, **US Social Security Number**, and **Passport Number** a search is performed as you type, and you must select the term from the list generated by the search. Once selected the term will be highlighted in black as shown below.

Click Create rule

The screenshot shows the 'Create rule' dialog box. In the 'Condition' section, the 'If' dropdown is set to 'User name' and the operator 'contains any' is selected. The value 'Howard Doe' is entered in the text field. In the 'Action' section, the 'then' dropdown is set to 'mask data' and the 'in columns containing' dropdown is set to 'Data class'. Below these, under 'Select how to mask data:', the 'Redact' option is selected, with the 'After' field containing 'XXXXXXXXXX' and the 'Replace data with Xs' option checked. Other options like 'Substitute' and 'Obfuscate' are also shown. At the bottom right, there is a blue 'Create' button.

13. The rule is displayed with the option to **Edit rule** or **Delete rule**.

Redact Government ID

Edit rule Delete rule

Criteria

Condition 1

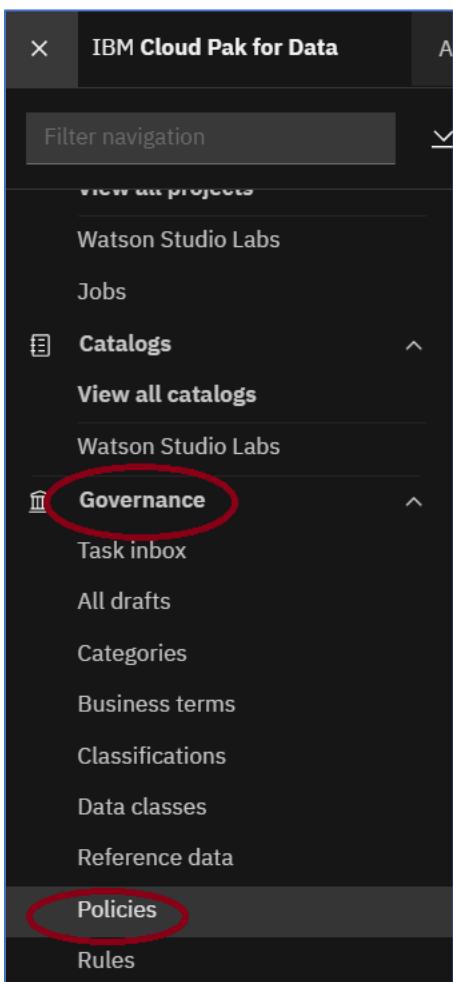
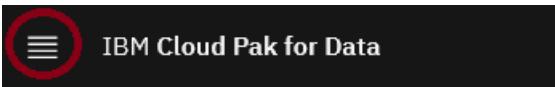
If User name contains any Howard Doe

Action

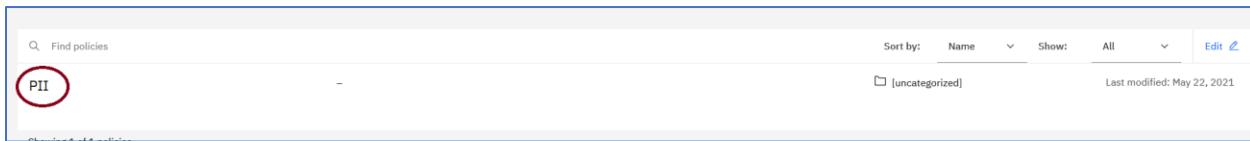
Then Substitute data in columns containing Data class

US Social Security Number, Passport Number

14. Click on the  icon. Click on **Governance** and then **Policies**.

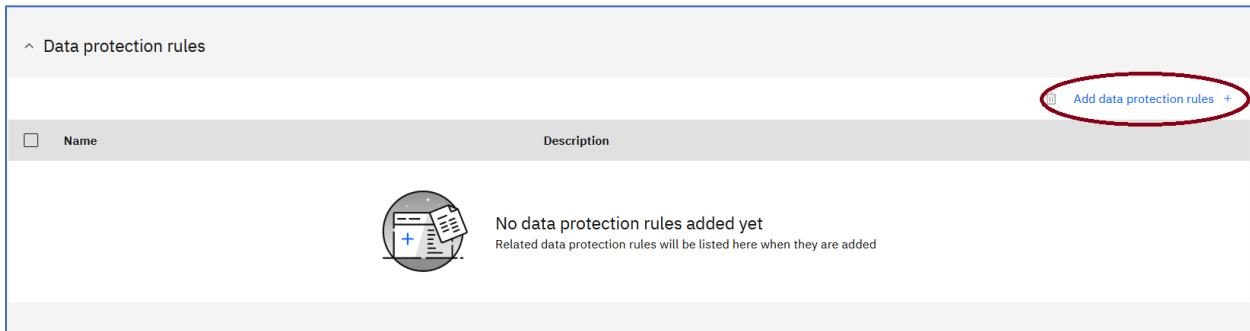


15. Click on **PII**



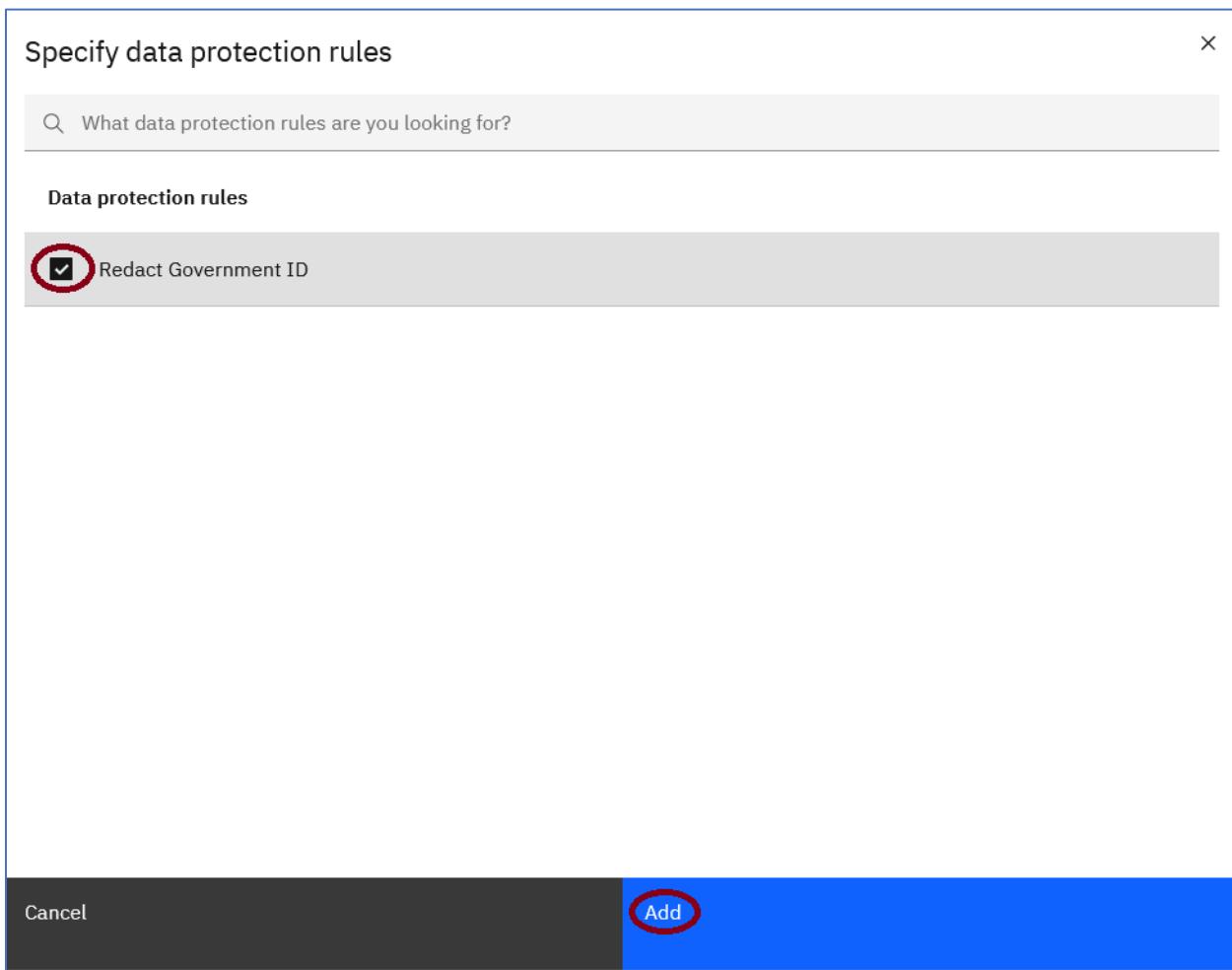
A screenshot of a web-based policy management application. At the top, there's a search bar labeled 'Find policies' and filter options 'Sort by: Name' and 'Show: All'. On the right, it shows 'Last modified: May 22, 2021'. The main area lists categories under 'uncategorized', with 'PII' circled in red.

16. Scroll down and click **Add data protection rule**.



A screenshot of a 'Data protection rules' section. It includes a header with 'Name' and 'Description' columns, a search icon, and an 'Add data protection rules' button circled in red. Below, a message states 'No data protection rules added yet' with a note about related rules appearing when added.

17. Click the checkbox next to **Redact Government ID** and click **Add**.



A screenshot of a modal dialog titled 'Specify data protection rules'. It features a search bar, a 'Data protection rules' section with a checked checkbox for 'Redact Government ID' (circled in red), and a bottom row with 'Cancel' and 'Add' buttons, where 'Add' is also circled in red.

18. Switch back to the browser with the ws.catalog.user@gmail.com user logged in. Click on the **female_human_trafficking** asset.

The screenshot shows the IBM Watson Studio interface. The left sidebar has 'Watson Studio Labs' selected. The main area displays a table of assets:

Name	Owner	Tags	Business terms	Type	Date added
Categories.csv	Avril Doe	job ty... job c...		Data asset	Aug 21, 2021
Trafficking_DB2WH	Avril Doe			Connection	Aug 21, 2021
Trafficking_Mongo	Avril Doe			Connection	Aug 21, 2021
female human trafficking	Avril Doe	traffi... femal...		Data asset	Aug 21, 2021
fht_arrival_departure	Avril Doe	traffi... femal... + 2 more		Data asset	Aug 21, 2021

A modal window titled 'Upload' is open on the right, showing 'In progress' status with a file icon. Below it, a message says 'No loads in progress' and 'Click the "Add to catalog" action to add assets.'

19. Click on Asset.

The screenshot shows the details of the 'female_human_trafficking' asset. The 'Asset' tab is selected. The asset was added on Jan 09, 2021, at 6:01 PM, in application/octet-stream format, and is 324 KB in size. It has two tags: 'female human trafficking' and 'trafficking'. There is one review. The asset is connected to 'trafficking' and is a 'Db2 Warehouse' type. The 'General' section shows no description or path. The 'Governance artifacts' section shows no business terms or classifications. The 'Related assets' section is empty.

20. 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

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
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...	VETTING_L...	DESCRIP...	NAME	GENDER	BIRTH_D...	BIRTH_COU...	BIRTH_COUNTR...	OCCU...
Smallint	Smallint	String	String	String	Date	String	String	String
Identif...	Code	Country...	Person...	Gender	Date o...	Country N...	Country Code	Text

Data masking in progress
This asset is being masked by the data enforcement rule: [PIL](#). You can wait here to see a preview of the asset or we can notify you when the preview is ready.

Notify Me

21. Note that 2 columns are anonymized.

IBM Watson Studio

Catalogs / Watson Studio Labs / female_human_trafficking

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

Connection
Source: Watson Studio Labs_DataCatalog
Source type: Cloud Object Storage

Classification
None

OCCUPAT...	ADDRESS	SSN	PASSPORT_NUMB...	PASSPORT_COU...	PASSPORT_COUNTR...	COUNTRIES_VIS...
String	String	String	String	String	String	String
Text	Text	US So...	Passport Number	Country Name	Country Code	Text
Art therapist	7089 Jack Trail,	fb7d4f9691eba8	306d01cd075d507b4e45	Ghana	GH	RU
Engineering geo	02820 Hart Ville,	2fbccdd8f718a87	86e6233e6a8cfc0162d04	Ghana	GH	ME,PE,UZ,UA
Engineer, civil (c)	7161 James Via,	ddda0cbef571d3	cdb7a80fc4ed910e968	Ghana	GH	TN,QA,BE,VN,PA
Teacher, adult ed	724 Morrow Lodi	a5f5fe28b6e8b2e	e9eac57f67f58e3d3f81fb	Ghana	GH	OM
Ecologist	36678 Henderso	5c884e47ceb68	3cc44ebad035a1dc65b4f	Ghana	GH	LB,AE,SN,PH
Geoscientist	376 Bowers Ran	73dd27ae9d2d5	9f5c9998d97ff12919c09	Ghana	GH	HK,KH,OM,PG,LT
Bookseller	2448 Autumn Co	d0dbd68afe433	a17651e3085bc2179149	Ghana	GH	HR,RU,UY,CA,EG,GB
Medical laborato	1140 Arnold Trail	b41260f74d0c2	2248c250bf21d96a2e55	Ghana	GH	UZ,NZ,LK
Biomedical scienc	38712 Griffith O	fe27fd2e3d87c1	11be85abf2b58bcbfb89c	Ghana	GH	MX,CO,CL,OM,UA,UZ
Dealer	143 Solis Roads,	0ba4ea52625a6	fc2f6c5a7eb557903b2dc1	Ghana	GH	BZ,UZ,IE,LT
Doctor, hospital	30788 Chan Vig	87c98549d6f7a1	9195e4244e18a002add1	Ghana	GH	RU
Architectural tec	3930 Murray Mai	617273dd767d7	e0b0d542c081ea093129	Ghana	GH	MX,IL,RU,IT,BE,CA,EG,U
Glass blower/dec	3206 Singleton T	2eb1e3b53abda	ec27e923ed2e4da79b94	Ghana	GH	BE,IL,PK,OM
Field seismologist	17407 Fitzgerald	86bdd9bc45c84	33e80462082cc9031190	Ghana	GH	AE
Plant breeder/ge	42740 Adam Bro	d5ed650eeea722	94ece1d9368d01027c00	Ghana	GH	GH,BG,RU

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

You have completed Lab-2!

- ✓ Created Platform Connections
- ✓ Created a governed catalog
- ✓ Added a member to the catalog
- ✓ Added Data Assets to the catalog
- ✓ Pushed the Data Assets to the project set up in Lab-1
- ✓ Searched the catalog
- ✓ Edited/Reviewed/Profiled a structured Data Asset
- ✓ Demonstrated access control features