

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.

Objectives

The goal of the lab is for the users to gain familiarity with the features of the Watson Knowledge Catalog. Before exercising the catalog functionality, the user will need to set up the environment needed to demonstrate the catalog features. This will include downloading the data assets, creating a database service, uploading two of the data assets to the database, and adding a user to the Watson Studio account. Once this is completed, 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 the prerequisites.

Set Up Environment

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

1. Click on [trafficking_data.zip](#), to download the zipped file.
2. Extract the file contents. You should have three files extracted, (1) `female_human_trafficking.csv`, (2) `Categories.csv`, and (3) `Occupation.csv`.

Step 2: Create a DB2 on Cloud service and Load the `female_human_trafficking.csv` file into the Database.

1. Click on [DB2](#) to create the **DB2 on Cloud** service.
2. Enter `Trafficking_DB` as the **Service Name** and then click on **Create**.

← View all

Db2
Lite • IBM

A fully-managed cloud SQL database. Powered by a turbo-charged Db2 engine.

[View Docs](#) [Terms](#)

AUTHOR IBM
PUBLISHED 12/28/2018
TYPE Service
LOCATION Sydney, Frankfurt, London, Dallas

Service name:
Trafficking_DB

Choose a region/location to deploy in:
Dallas

Choose an organization:
DSXUser56000@gmail.com

Choose a space:
dev

Email:
email@example.com

Please provide an email address where we can reach you for updates

Features

- Fast, Reliable & Robust**
Uses Db2 technology for enterprise-level OLTP performance.
- Fully managed, safe, and secure**
Includes daily backups for 14 days, at-rest database encryption,

Need Help? [Contact IBM Cloud Support](#) [Estimate Monthly Cost](#) [Cost Calculator](#) [Create](#)

3. Click on Trafficking_DB to open the service.

Cloud Foundry Services

Name	Region	CF Org	CF Space	Plan	Service Offering
Trafficking_DB	Dallas	DSXUser56000@g...	dev	Lite	dashDB For Transa...

Services

Name	Location	Resource Group	Plan	Details	Service Offering
KnowledgeCatalog	Dallas	Default	Lite	Provisioned	Knowledge Catalog
WatsonStudio	Dallas	Default	Lite	Provisioned	Watson Studio
cloud-object-storage-cv	global	Default	Lite	Provisioned	Cloud Object Storage

4. We will need the Service Credentials for the DB2 on Cloud service later in the lab. The next few steps will generate the Service Credentials. You will then cut and paste these credentials into a text editor for later use. Click on **Service Credentials**.

Manage

Service credentials

Connections

Dashboard / Trafficking_DB

Location: Dallas Org: DSXUser56000@gmail.com Space: dev

[Open Console](#)

Getting Started

Where can I find my credentials?
Get your username and password by clicking the "Service Credentials" link to the left and selecting "New Credentials".

[Getting Started](#)

Need Help?
Use IBM dW Answers to view recently asked questions or ask your own. Still unable to find an answer? Submit a Bluemix Support Ticket to our team.

[IBM dW Answers](#) [Support Ticket](#)

5. Click on **New credential**.


Dashboard /

Trafficking_DB

Location: Dallas Org: DSXUser56000@gmail.com Space: dev

Service credentials

Credentials are provided in JSON format. The JSON snippet lists credentials, such as the API key and secret, as well as connection information for the service. [Learn more](#)

Service credentials **New credential** 


Click **New credentials** to create a set of credentials for this instance

6. Click on **Add**.

×

Add new credential

Service credentials-1

Add Inline Configuration Parameters (Optional): 

Provide service-specific configuration parameters in a valid JSON object

Choose File...

Cancel **Add**

7. Click on **View Credentials**.

Dashboard / Trafficking_DB

Location: Dallas Org: DSXUser56000@gmail.com Space: dev

Service credentials

Credentials are provided in JSON format. The JSON snippet lists credentials, such as the API key and secret, as well as connection information for the service. [Learn more](#)

Service credentials [New credential +](#)

Items per page 10 | 1-1 of 1 items 1 of 1 pages < 1 >

KEY NAME	DATE CREATED	ACTIONS
<input type="checkbox"/> Service credentials-1	JAN 19, 2019 - 03:00:51 PM	View credentials

8. Cut and paste the credential information into a text editor. We will need the username, password, host, and db fields to set up a connection to the database service.

KEY NAME	DATE CREATED	ACTIONS
<input type="checkbox"/> Service credentials-1	JAN 19, 2019 - 03:00:51 PM	View credentials

```

{
  "password": "215410011818s27",
  "https_url": "https://dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net",
  "port": 50000,
  "ssldsn": "DATABASE=BLUDB;HOSTNAME=dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net;PORT=50000;PROTOCOL=TCP;SSL=TRUE;SECURITY=SSL;",
  "host": "dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net",
  "jdbcurl": "jdbc:db2://dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net:50000/BLUDB",
  "uri": "db2://xc42672:215410011818s27@dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net:50000/BLUDB",
  "db": "BLUDB",
  "dsn": "DATABASE=BLUDB;HOSTNAME=dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net;PORT=50000;PROTOCOL=TCP;SSL=TRUE;SECURITY=SSL;",
  "username": "xc42672",
  "ssljdbcurl": "jdbc:db2://dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net:50000/BLUDB;sslConnection=true;"
}

```

9. We are now going to load the female_human_trafficking file contents into the database service that we just created. The DB2 on Cloud service console provides a built-in load capability. Click on **Manage**.

Manage

Service credentials

Connections

Location: Dallas Org: DSXUser56000@gmail.com Space: dev

Items per page 10 | 1-1 of 1 items 1 of 1 pages < 1 >

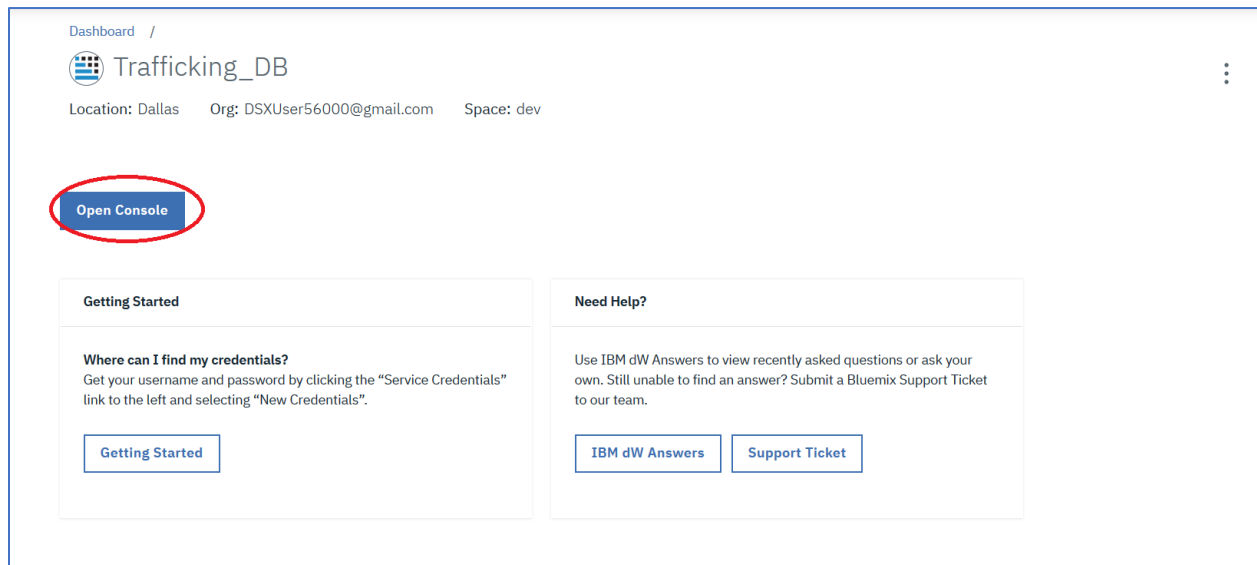
KEY NAME	DATE CREATED	ACTIONS
<input type="checkbox"/> Service credentials-1	JAN 19, 2019 - 03:00:51 PM	View credentials

```

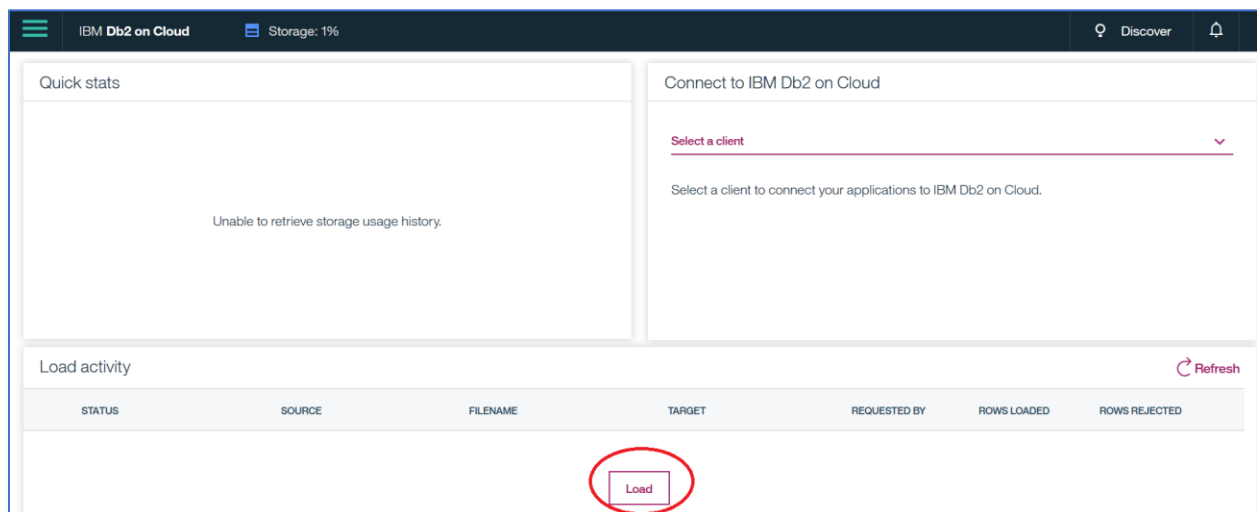
{
  "password": "215410011818s27",
  "https_url": "https://dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net",
  "port": 50000,
  "ssldsn": "DATABASE=BLUDB;HOSTNAME=dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net;PORT=50000;PROTOCOL=TCP;SSL=TRUE;SECURITY=SSL;",
  "host": "dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net",
  "jdbcurl": "jdbc:db2://dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net:50000/BLUDB",
  "uri": "db2://xc42672:215410011818s27@dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net:50000/BLUDB",
  "db": "BLUDB",
  "dsn": "DATABASE=BLUDB;HOSTNAME=dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net;PORT=50000;PROTOCOL=TCP;SSL=TRUE;SECURITY=SSL;",
  "username": "xc42672",
  "ssljdbcurl": "jdbc:db2://dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net:50000/BLUDB;sslConnection=true;"
}

```

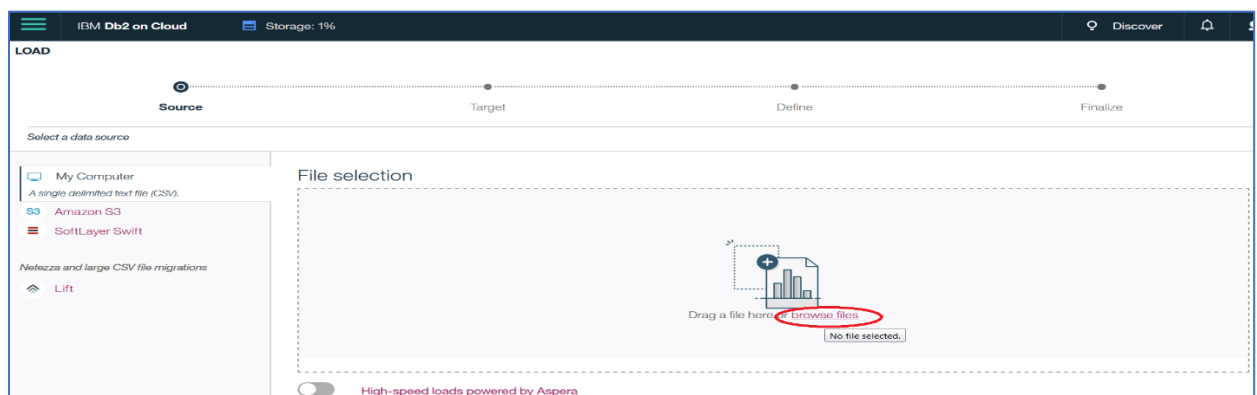
10. Click on **Open Console**



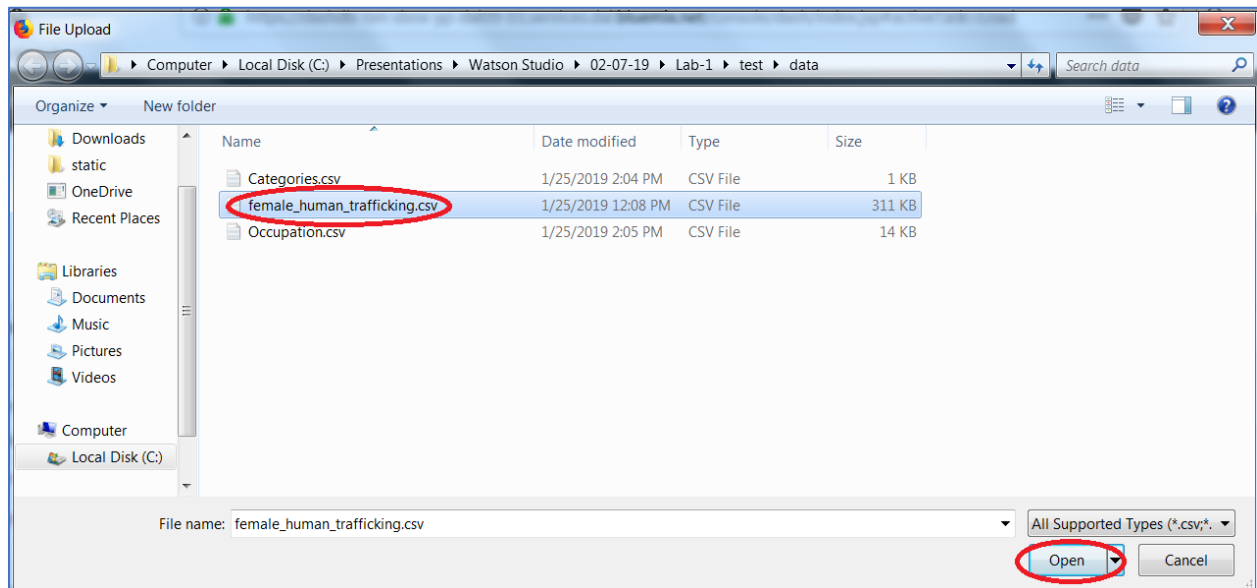
11. Click on **Load**.



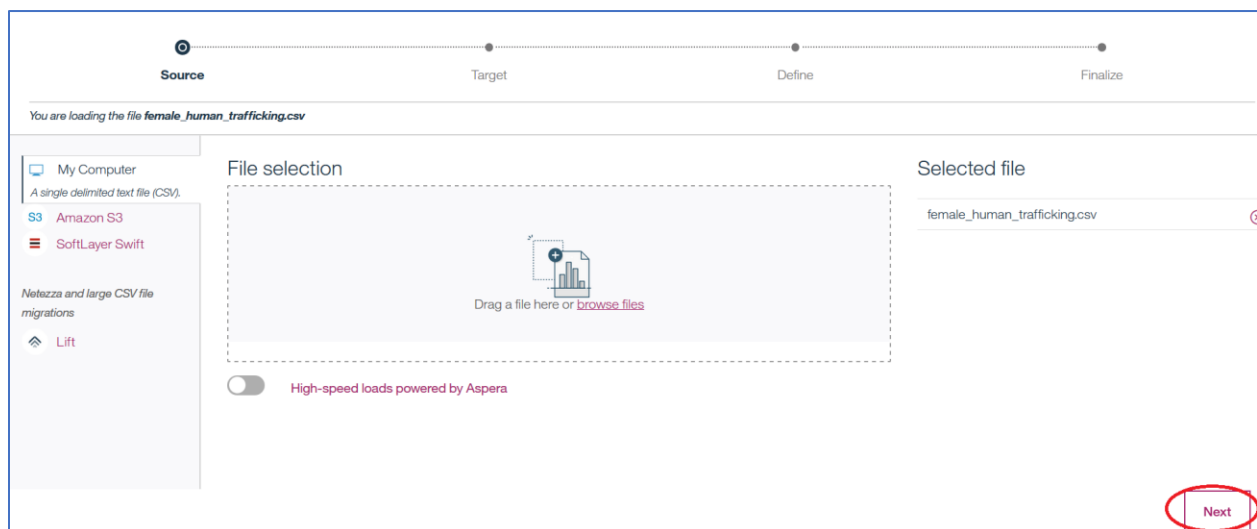
12. Click on **browse files**.



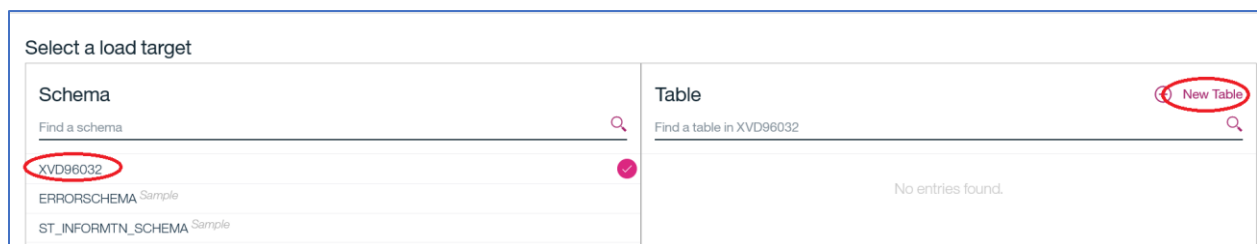
13. Navigate to the folder where you downloaded the trafficking files. Select **Categories.csv**, and then click on **Open**.



14. Click on **Next**.



15. Click on the schema that doesn't have Sample appearing next to it. It should have 3 characters followed by a number. After clicking on the schema, click on **New Table**.



16. Enter **FEMALE_HUMAN_TRAFFICKING** for the name of the database table, and then click **Create** to create the table.

Select a load target

Schema	Table	Create a new Table
Find a schema	Find a table in XVD96032	FEMALE_HUMAN_TRAFFICKING
XVD96032	No entries found.	Create
ERRORSCHEMA <small>Sample</small>		
ST_INFORMTN_SCHEMA <small>Sample</small>		

17. Click **Next**.

You are loading the file *female_human_trafficking.csv* into XVD96032.FEMALE_HUMAN_TRAFFICKING

Select a load target

Schema	Table	Table definition
Find a schema	Find a table in XVD96032	FEMALE_HUMAN_TRAFFICKING No statistics available
XVD96032	FEMALE_HUMAN_TRAFFICKING	Specify the table columns in the next step.
ERRORSCHEMA <small>Sample</small>		
ST_INFORMTN_SCHEMA <small>Sample</small>		

Back **Next**

18. Click **Next**.

You are loading the file *female_human_trafficking.csv* into XVD96032.FEMALE_HUMAN_TRAFFICKING

Code page (character encoding): 1208 (UTF-8) Separator: , Header in first row: ☒ Time & date format: Time Detect data types: ☒

	INTERNAL...	VETTING_L...	DESCRIPT...	NAME	GENDER	BIRTH_D...	BIRTH_COU...	BIRTH_COUNTR...	OCCUPATION
	SMALLINT	SMALLINT	VARCHAR(2)	VARCHAR(27)	VARCHAR(1)	DATE	VARCHAR(10)	VARCHAR(2)	VARCHAR(59)
1	1	100	NA	Kathleen Bailey	F	1985-11-27	Ghana	GH	Chiroprapist
2	2	30	NA	Kelly Robbins	F	1971-01-17	Pakistan	PK	Engineer, structural
3	3	100	NA	Linda Stewart	F	1997-04-15	Ghana	GH	Engineer, land
4	4	30	NA	Stacey Courtney Gonzalez	F	1984-03-27	Ghana	GH	Careers adviser
5	5	30	NA	Erika Patie Fowler	F	1998-06-18	Ghana	GH	Geneticist, molecular
6	6	100	NA	Ashlee Fisher	F	1981-01-05	Ghana	GH	Electronics engineer
7	7	100	NA	Jacqueline Clark	F	1976-01-19	Ghana	GH	Editor, commissioning
8	8	30	NA	Shell Chapman	F	1997-08-24	Ghana	GH	Writer
9	9	10	NA	Sabrina Debra Wilson	F	1977-02-21	Haiti	HT	Recruitment consultant
10	10	100	NA	Ashley Gibson	F	1986-01-05	Ghana	GH	Freight manager

Back **Next**

19. Click **Begin Load**.

You are loading the file `female_human_trafficking.csv` into `XVD96032.FEMALE_HUMAN_TRAFFICKING`

Review settings

Summary

Code page: 1208 (Default)

Separator: , (Default)

Header in first row: Yes (Default)

Time format: HH:MM:SS (Default)

Date format: YYYY-MM-DD (Default)

Timestamp format: YYYY-MM-DD HH:MM:SS (Default)

String delimiter: "" (Default)

Option

Maximum number of warnings

1000

Back **Begin Load**

20. The 1085 records in the `female_human_trafficking.csv` file are successfully loaded. Click on the canvas outside of the Notifications popup.

LOAD

Load details

COMPLETE

My computer
`female_human_trafficking.csv`

Target
`XVD96032.FEMALE_HUMAN_TRAFFICK...`

Status Settings

1,085 Rows read

1,085 Rows loaded

0 Rows rejected

Start time
01/29/2019 4:49:18 PM

End time
01/29/2019 4:49:21 PM

The data load job succeeded.

You can now work with your data.

Errors 0

No errors

Notifications

COMPLETE

My computer
`female_human_trafficking.csv`

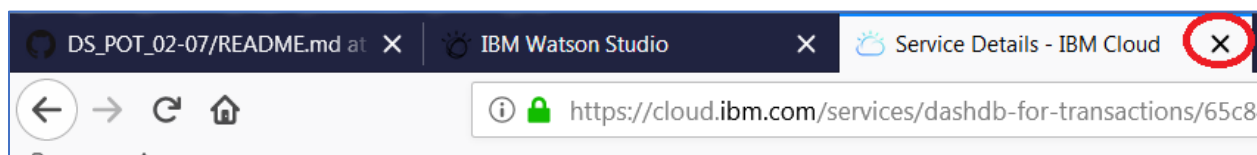
Target
`XVD96032.FEMALE_HUMAN_TRAFFICK...`

Clear All

21. Close the **DB2 on Cloud** browser tab.



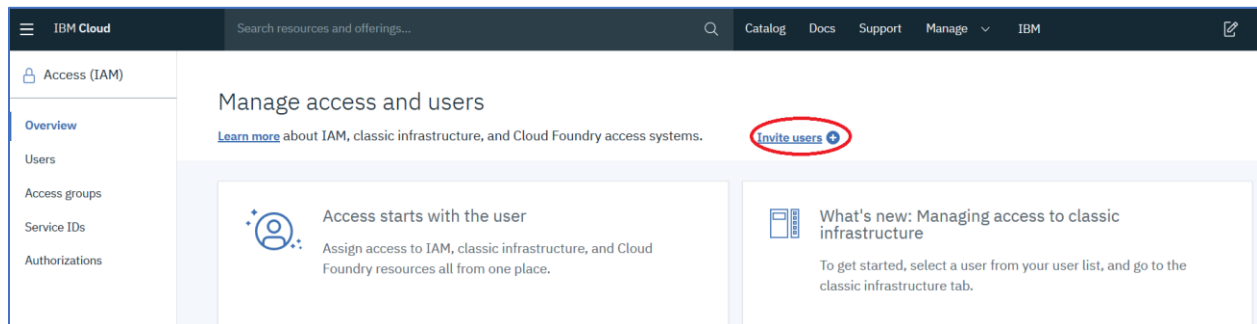
22. Close the **Service Details** browser tab.



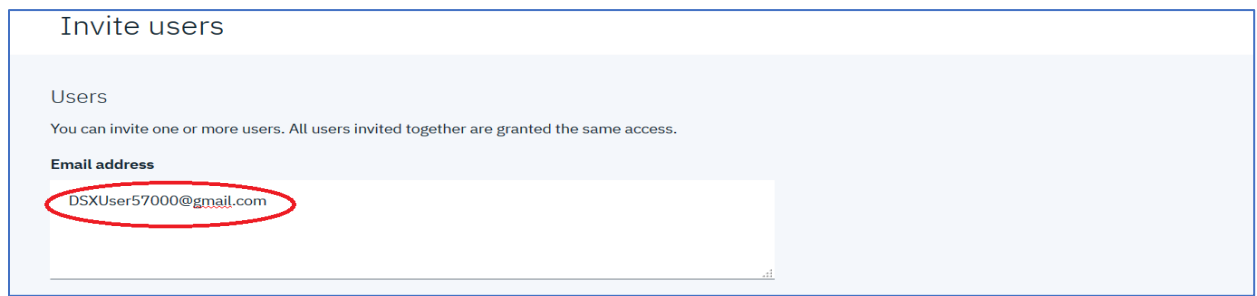
Step 3: Add a User to the Watson Studio account

To demonstrate the access control features of Watson Knowledge Catalog, add a user to your account.

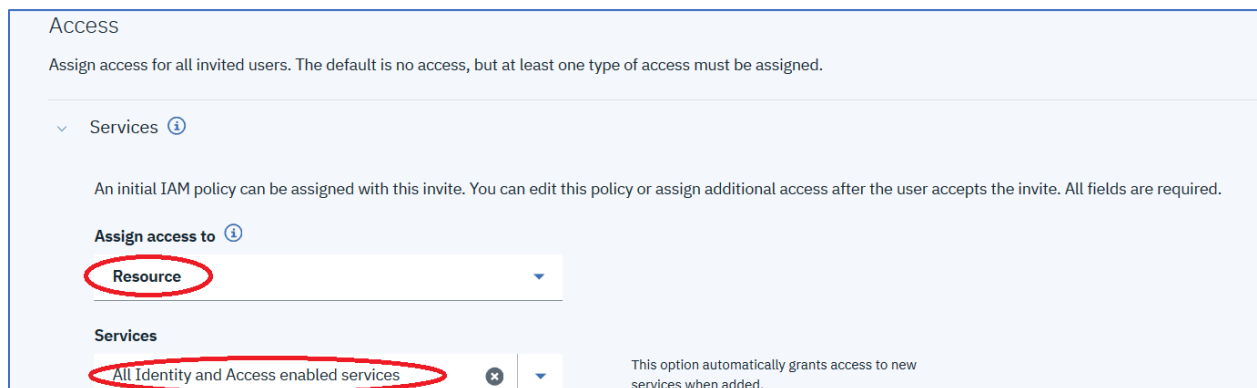
1. Click on [Manage Users](#), and then click on **Invite users**.



2. In the **Users** section, enter DSX57000@gmail.com for the **Email** address.



3. In the **Services** section, select Resources for the **Assign access to**, and select All Identity and Access enabled services for **Services**.



4. Scroll down and assign Reader for the **Assign service access roles**, and then click on **Invite users**.

Assign platform access roles

These roles enable users to complete tasks on platform resources, such as creating instances, connecting instances to apps, and assigning user permissions.

Assign platform access roles

☐ Administrator
As an administrator, you can perform all platform actions based on the resource this role is being assigned, including assigning access policies to other users.

☐ Editor
As an editor, you can perform all platform actions except for managing the account and assigning access policies.

☐ Operator
As an operator, you can perform platform actions required to configure and operate service instances, such as viewing a service's dashboard.

☐ Viewer
As a viewer, you can view service instances, but you can't modify them.

Assign service access roles

These roles allow a user access for using the service and performing service API calls.

Assign service access roles

☐ Manager
As a manager, you have permissions beyond the writer role to complete privileged actions as defined by the service. In addition, you can create and edit service-specific resources.

☐ Writer
As a writer, you have permissions beyond the reader role, including creating and editing service-specific resources.

☒ Reader
As a reader, you can perform read-only actions within a service such as viewing service-specific resources.

Cloud Foundry access ⓘ

Cancel

Invite users

- DSXUser57000@gmail.com is added as a user in the Watson Studio account.

Users

Account: ⓘ IBM

Use the View by option to change your view of the user list by selecting the user grouping. Depending on your access, you might see users grouped at the account level or by Cloud Foundry org or user hierarchy for classic infrastructure access.

View by: Account users ▾

Filter

🔍 Filter by name or email Status ▾

Invite users +

User ▲	Email	Status
DSX56000 User owner self	DSXUser56000@gmail.com	Active
DSX57000 User	DSXUser57000@gmail.com	Active

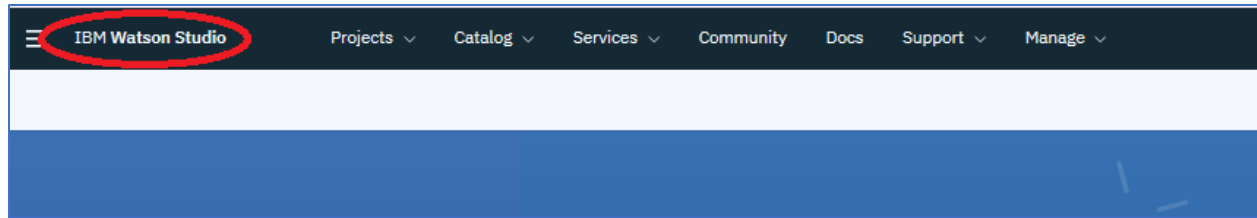
Users per page: 25 ▾ | 1-2 of 2 items

1 of 1 pages < >

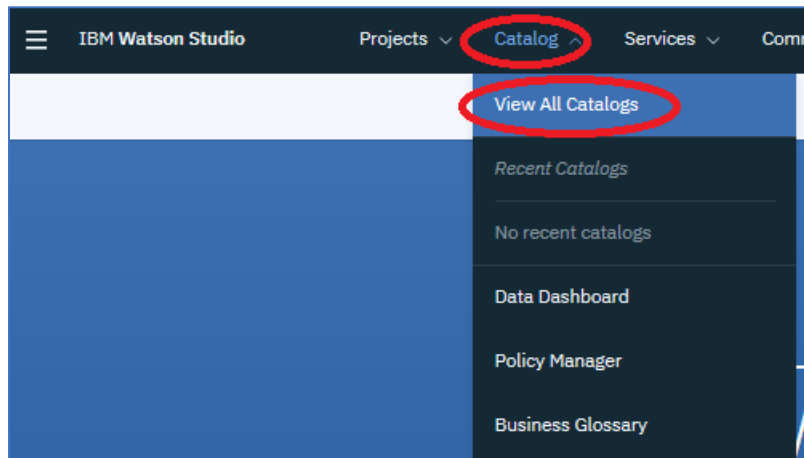
Perform Catalog Tasks

Step 1 – Create a Governed Catalog

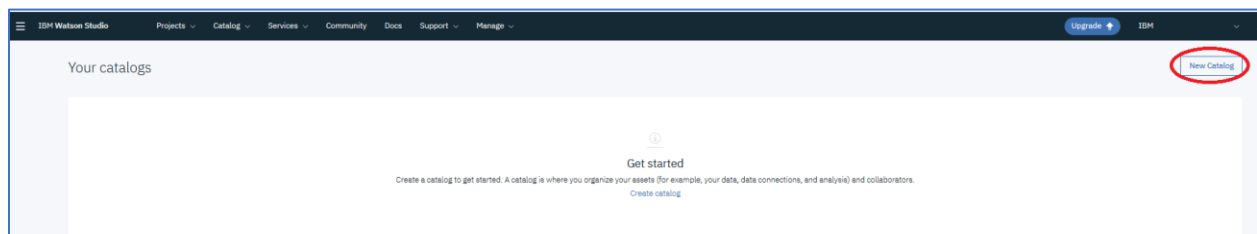
- Click on the browser tab to return to Watson Studio.
- Click on Watson Studio to go to the home screen.



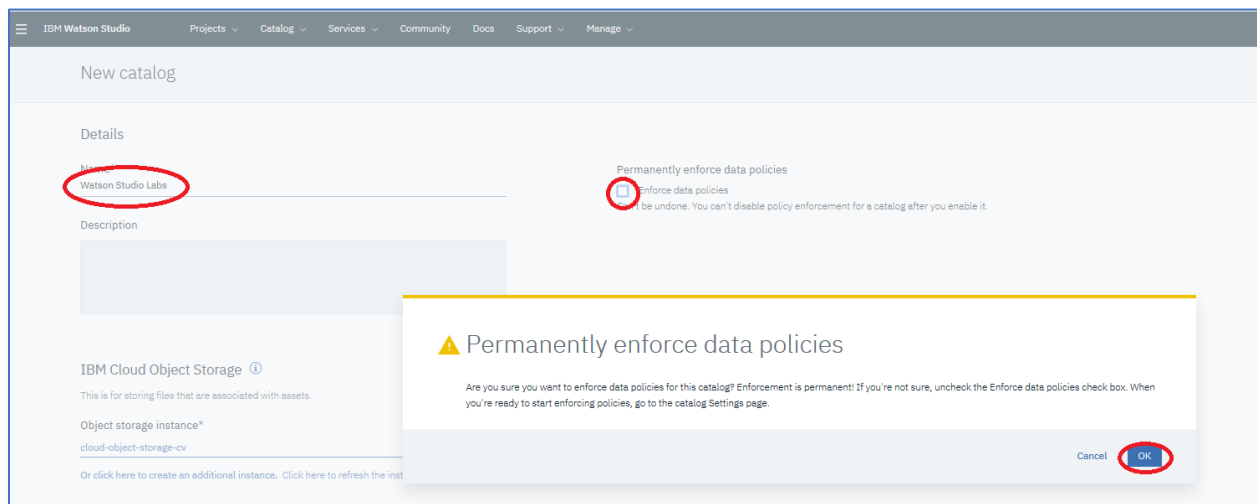
3. Click on **Catalog** and then **View All Catalogs**.



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**.



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
1000

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

Object storage instance*
cloud-object-storage-cv

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

Permanently enforce data policies
☒ Enforce data policies
Can't be undone. You can't disable policy enforcement for a catalog after you enable it.

Cancel Create

Step 2: Add a Member to the Catalog

To demonstrate the Catalog access control capability, we will add the username DSXUser57000@gmail.com as a member of the catalog. Note that the first name of this user is “DSX57000” and the last name of this user is “User”.

1. Click on **Access Control**.

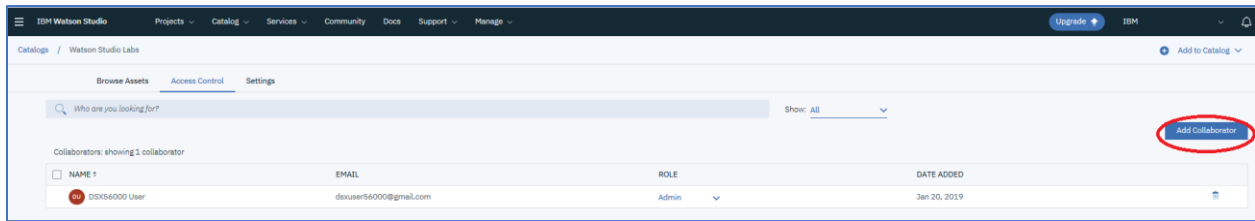
IBM Watson Studio Projects Catalog Services Community Docs Support Manage

Catalogs / Watson Studio Labs

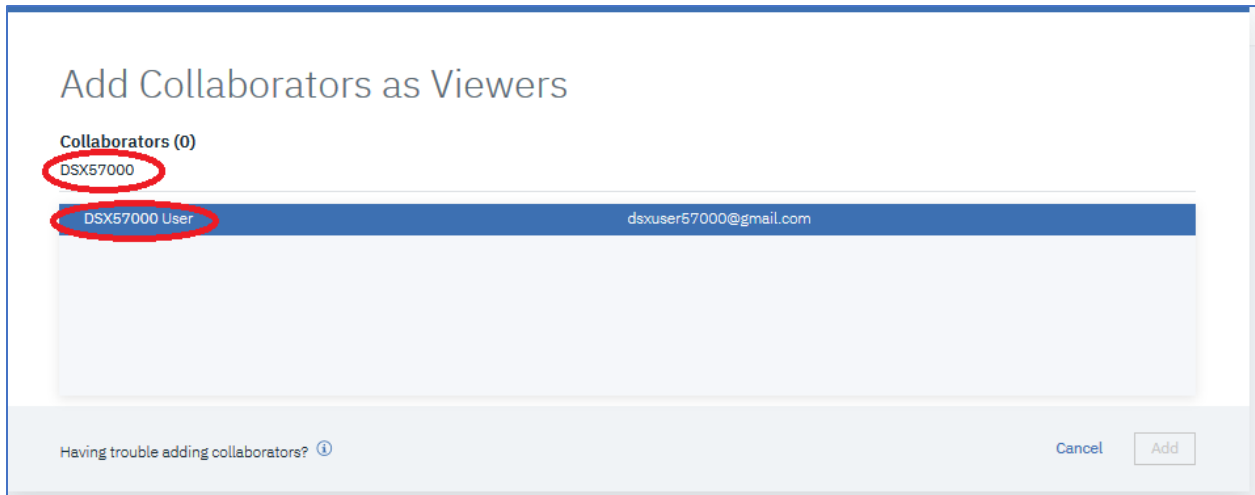
Browse Assets Access Control Settings

Now you can add assets!
Click here to get started.

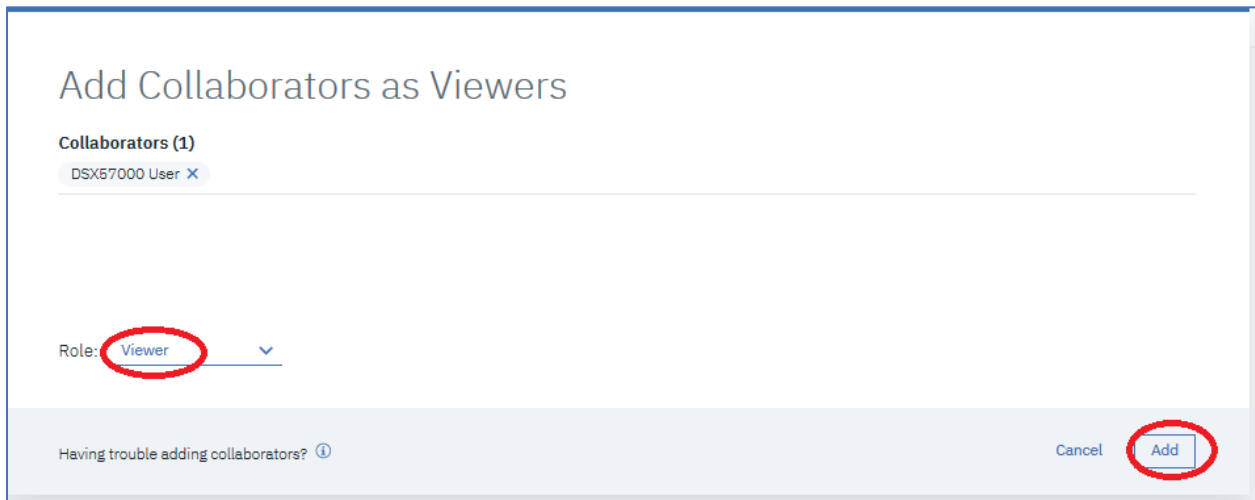
2. Click on **Add Collaborator**.



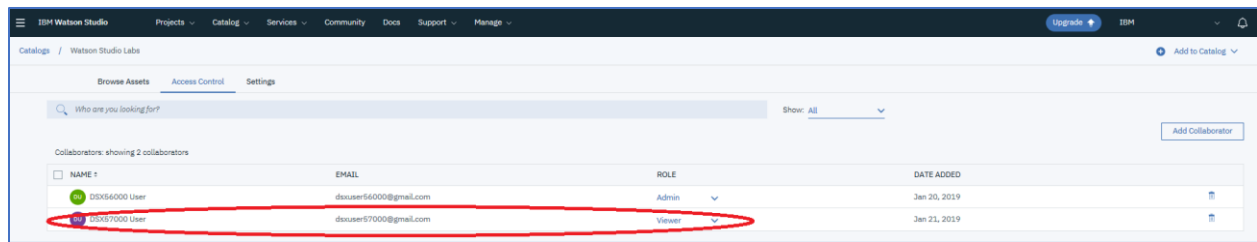
3. Enter DSX57000 (first name of the DSXUser57000@gmail.com user) as the **Collaborator**. DSX57000 should show up in the list box. Click on DSX57000 User.



4. Leave the Viewer as the **Role** and click on **Add**.



5. “DSX57000 User” is added to the catalog.



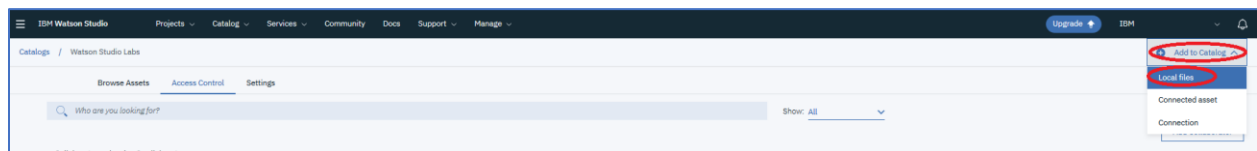
Step 3: Add Data Assets to the Catalog

In this step, we will add 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 the DB2 on Cloud database service that we created earlier (this is where we need the credentials). We will add a connected data asset that will point to the FEMALE_HUMAN_TRAFFICKING table.

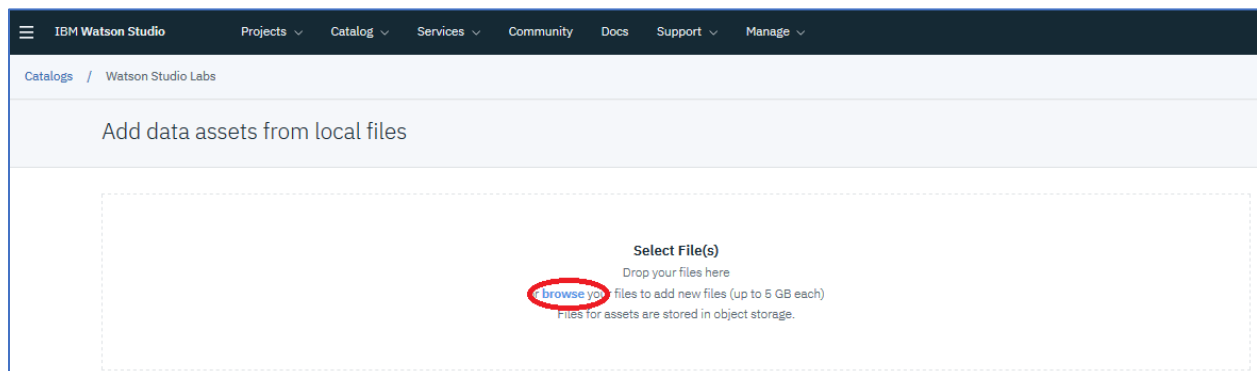
This step will introduce you to the three methods available to discover and catalog data assets; **Local files**, **Connected asset** and **Connection**. You will use these methods to catalog data assets into the newly created Knowledge Catalog and then tag them for users to easily find them, understand 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

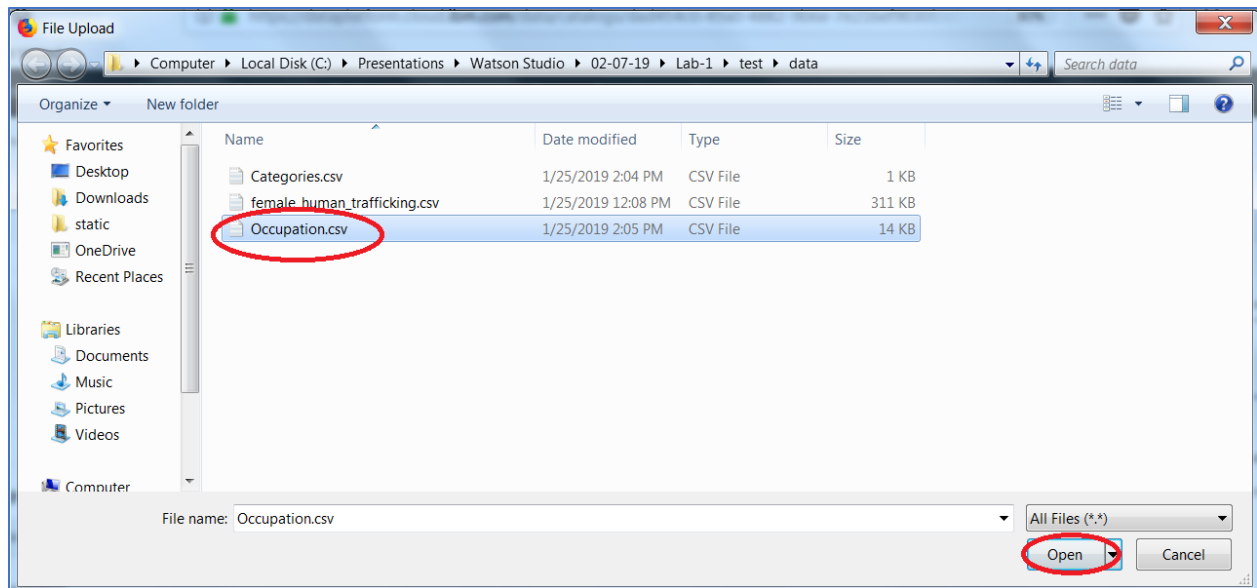
1. Click on **Add to Catalog** and click on **Local Files**.



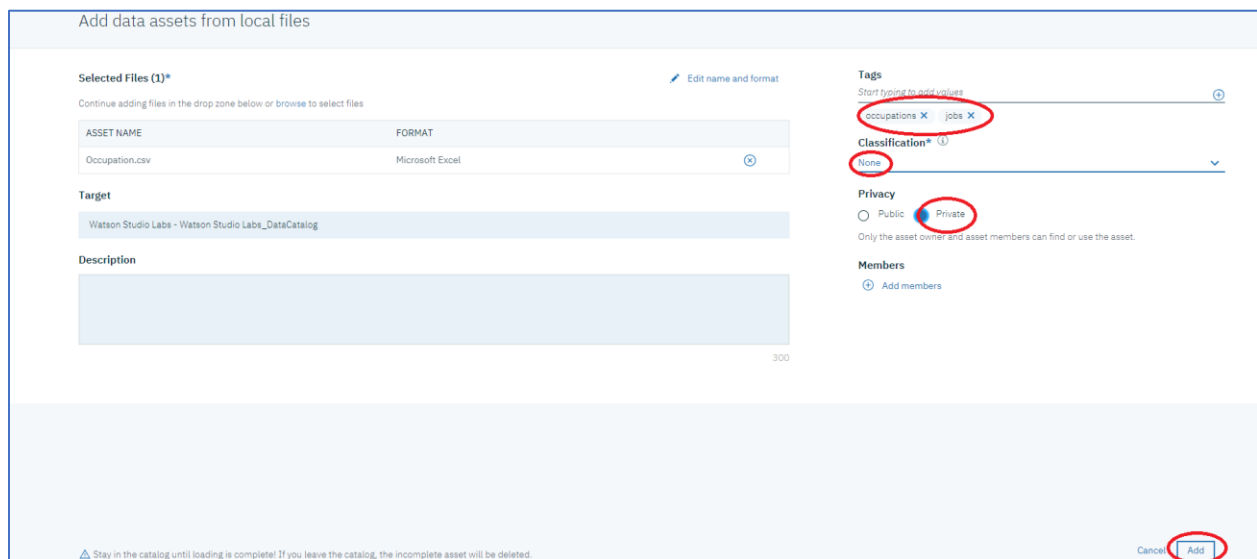
2. Click on **browse** to navigate to the folder where the “Occupation.csv” file is stored.



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



- 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**.



- The Occupation.csv asset metadata is displayed. Note a number of features on the panel. At the top 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. On the left side in the **Filter** area are shown filters by **Asset Type**, and **Tags**. As we add data assets, the tags associated with those assets will be displayed, and an additional asset type will be recorded. The number next to each tag entry indicates how many assets contains the specific tag. The number next to the Asset Type entry indicates how many assets exist for the specific type.

The screenshot shows the 'Browse Assets' page in Watson Studio. A sidebar on the left contains a 'Filter' section with 'Asset types' (Data Asset (1)) and 'Tags' (jobs (1), occupations (1)). The main area displays 'Available Assets' with a table showing one asset: 'Occupation.csv' owned by 'DSX59000 User', with tags 'occupations | jobs', type 'Data Asset', and date added 'Jan 30, 2019'.

NAME	OWNER	TAGS	TYPE	DATE ADDED
Occupation.csv	DSX59000 User	occupations jobs	Data Asset	Jan 30, 2019

- 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**.

The screenshot shows the 'Add data assets from local files' dialog. The 'Selected Files (1)*' section lists 'Categories.csv' as a Microsoft Excel file. The 'Target' is 'Watson Studio Labs - Watson Studio Labs_DataCatalog'. The 'Description' field is empty. On the right, the 'Tags' section shows 'job types' and 'job categories' entered. The 'Classification*' is set to 'None', and 'Privacy' is set to 'Public'. The 'Add' button at the bottom right is circled in red.

- The categories.csv asset is displayed. Note the **filter** area information has been updated to reflect the addition of this asset.

Successfully added data asset Categories.csv to Watson Studio Labs.

Browse Assets Access Control Settings

What assets are you looking for?

Watson Recommends Highly Rated Recently Added Collapse

Data Asset Categories.csv

Owner: DSX59000 User
Added: Jan 30, 2019 11:49 PM
job by... | job cate...
☆☆☆☆ 0 reviews

Data Asset Occupation.csv

Owner: DSX59000 User
Added: Jan 30, 2019 11:38 PM
occup... | jobs
☆☆☆☆ 0 reviews

Filter

Asset types

☐ Data Asset (2)

Tags

☐ job categories (1)
☐ job types (1)
☐ jobs (1)
☐ occupations (1)

Available Assets

Showing 2 of 2 assets

NAME	OWNER	TAGS	TYPE	DATE ADDED
Categories.csv	DSX59000 User	job types job categories	Data Asset	Jan 30, 2019
Occupation.csv	DSX59000 User	occupations jobs	Data Asset	Jan 30, 2019

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

Available Assets

Showing 2 of 2 assets

NAME	OWNER	TAGS	TYPE	DATE ADDED
Categories.csv	DSX59000 User	job types job categories	Data Asset	Jan 30, 2019
Occupation.csv	DSX59000 User	occupations jobs	Data Asset	Jan 30, 2019

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

Catalogs / Watson Studio Labs / Occupation.csv

Overview Access Review Profile Lineage

Data Asset Occupation.csv

Remove Download Add to Project

Description

There is no description available for this asset.

Added: Jan 31, 2019 12:00 AM
Format: Microsoft Excel
Size: 30 KB

Tags

occupations | jobs

Reviews

☆☆☆☆ 0 reviews

Connection

Source: Watson Studio Labs_DataCatalog
Source type: Cloud Object Storage

Classification

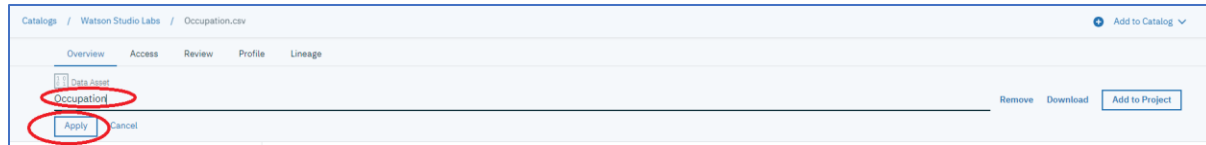
None

Schema: 2 Columns | 492 Rows

Preview: 492 rows | Last refresh: right now | Refresh

OCCUPATION	Code
Text	Type: String
Chiroprapist	6
Engineer, structural	2
Engineer, land	2
Careers adviser	15
Geneticist, molecular	7
Electronics engineer	2
Editor, commissioning	4
Writer	4
Recruitment consultant	15
Theatre manager	8
Statistician	2
Merchant navy officer	15
Advertising account executive	9
Administrator, local government	6

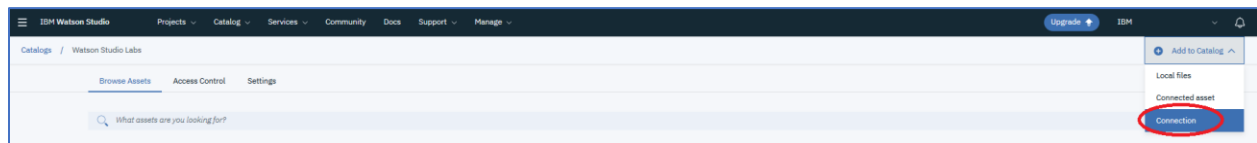
10. Change the name to Occupation and click **Apply**.



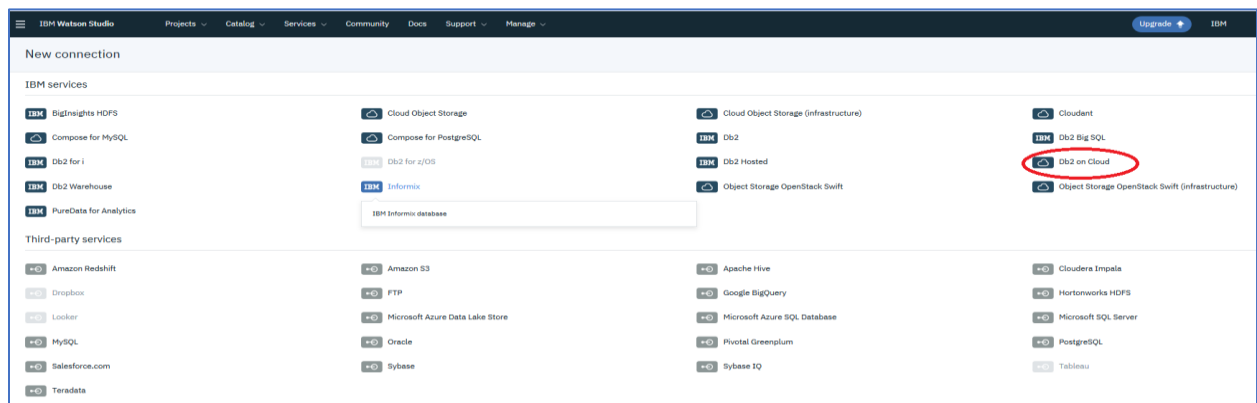
11. Click on **Watson Studio Labs** to return to the catalog display.
12. Follow steps 8-11 to change the Categories.csv asset name to be Categories.

Add Connection

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



2. Click on **DB2 on Cloud**.



3. Enter "trafficking" for the **Name** of the connection. Cut and paste the credentials Username, Host, Password, and Database that you saved earlier into the corresponding fields. Click on **Create**.

Connection overview

Name
trafficking

Description
IBM Db2 fully-managed cloud SQL database

Connection Details

Username *
xxd96032

Hostname or IP Address *
01535-txn-sbox-yp-dal09-03.services.dal.bluemix.net

Secure Gateway
☐ Use a secure gateway

Password *

Database *
BLUDB

Connection discovery
☐ Discover data assets

Enter information for the selected data source

Cancel Create

- The connection asset is displayed. Note that a new Asset Type (Connection) is added in the **Filter-Asset Type** area, as well as in the **Type** description for the data asset. Also, note an untagged entry has been created in the **Filter-Tags** area.

Catalogs / Watson Studio Labs

Browse Assets Access Control Settings

What assets are you looking for?

Watson Recommends Highly Rated Recently Added

Connection
trafficking

Owner: DSXS9000 User
Added: Jan 31, 2019 12:03 AM

0 reviews

Data Asset
Categories

Owner: DSXS9000 User
Added: Jan 31, 2019 12:01 AM

job ty... | job cate...

0 reviews

Data Asset
Occupation

Owner: DSXS9000 User
Added: Jan 31, 2019 12:00 AM

occup... | jobs

0 reviews

Filter

Asset types

☐ Data Asset (2)

☐ Connection (1)

Tags

☐ job categories (1)

☐ job types (1)

☐ jobs (1)

☐ occupations (1)

Available Assets

Showing 3 of 3 assets

NAME	OWNER	TAGS	TYPE	DATE ADDED
Categories	DSXS9000 User	job types job categories	Data Asset	Jan 31, 2019
Occupation	DSXS9000 User	occupations jobs	Data Asset	Jan 31, 2019
trafficking	DSXS9000 User		Connection	Jan 31, 2019

Add Connected Data

- We have added a connection to the DB2 on Cloud database service. We will now add a Connected Data Asset to point to the FEMALE_HUMAN_TRAFFICKING table that was created earlier. Click on **Add to Catalog** and click on **Connected asset**.

IBM Watson Studio

Projects Catalog Services Community Docs Support Manage

Catalogs / Watson Studio Labs

Browse Assets Access Control Settings

What assets are you looking for?

Add to Catalog

Local files

Connected asset

Connection

- Click on **Select Source**.

Catalogs / Watson Studio Labs

Add asset from connection

Source*
Select Source

Name*
Asset name

Description

Tags
Start typing to add values

Classification*
None

Privacy
Public Private

Members
Add members

- Click on the **trafficking** connection, click on the **Schema name**, click on **FEMALE_HUMAN_TRAFFICKING**, and then click **Select**.

Catalogs / Watson Studio Labs

Watson Studio Labs	trafficking	XVD96032
Connections (1)	Schemas (8)	Tables (1)
trafficking	AUDIT	FEMALE_HUMAN_TRAFFICKING
	ERRORSCHEMA	
	IBM_RTMON_DATA	
	SQL08548	
	SQL28627	
	SQL32811	
	ST_INFORMTN_SCHEMA	
	XVD96032	

Cancel Select

- Enter **female_human_trafficking** for **Name**, enter “trafficking” for **Tags**, click the “+” icon, enter “female human trafficking” for **Tags**, click the “+” icon. Select “Personally Identifiable Information” for **Classification**, click on **Private**, for Privacy, and click on **Add 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.

Add asset from connection

Source Db2 on Cloud

trafficking : XVD96032/FEMALE_HUMAN_TRAFFICKING

Name*
female_human_trafficking

Description

Tags
Start typing to add values

trafficking female human trafficking

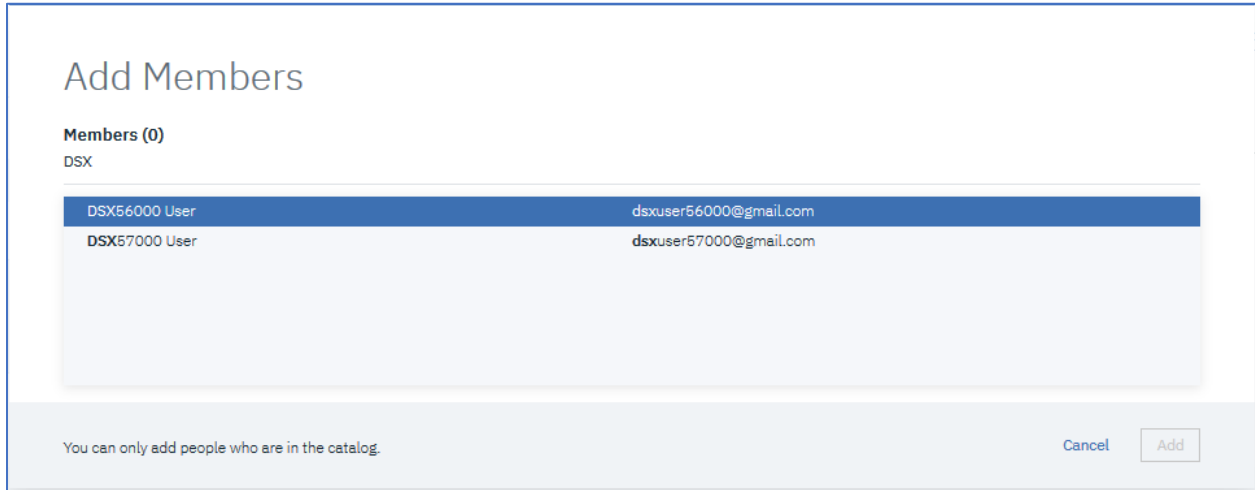
Classification*
Personally Identifiable Information

Privacy
Public Private

Members
Add members

Cancel Add

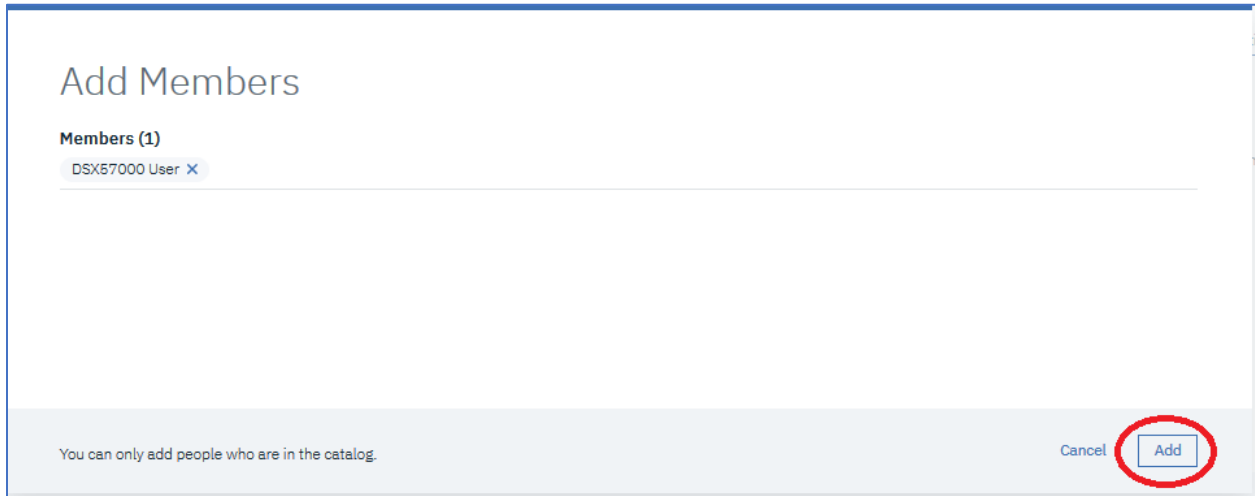
5. Enter DSX in the **Members** field. The list of matching users is displayed in the member list. Select the DSX57000 user.



The screenshot shows a dialog box titled "Add Members". Below the title, it says "Members (0)" and "DSX". A list of users is displayed in a table-like format. The first row is highlighted in blue and contains "DSX56000 User" and "dsxuser56000@gmail.com". The second row contains "DSX57000 User" and "dsxuser57000@gmail.com". At the bottom of the dialog, there is a message "You can only add people who are in the catalog." and two buttons: "Cancel" and "Add".

DSX56000 User	dsxuser56000@gmail.com
DSX57000 User	dsxuser57000@gmail.com

6. Click on **Add** to add DSX5700 User to the access control list for the female_human_trafficking asset.



The screenshot shows the same "Add Members" dialog box, but now it says "Members (1)" and "DSX57000 User X" is listed. The "Add" button at the bottom right is circled in red. The message "You can only add people who are in the catalog." and the "Cancel" button are still present.

7. Click on **Add**.

Add asset from connection

Source Db2 on Cloud Change source

trafficking : /XVD96032/FEMALE_HUMAN_TRAFFICKING

Name* female_human_trafficking 76

Description 300

Tags Start typing to add values
trafficking X female human traffi... X

Classification* ?
Personally Identifiable Information

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

Members ? Add members

NAME
DSX57000 User

Cancel Add

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

IBM Watson Studio Projects Catalog Services Community Docs Support Manage Upgrade IBM

Catalogs / Watson Studio Labs ? Add to Catalog

Browse Assets Access Control Settings

What assets are you looking for?

Watson Recommends Highly Rated Recently Added Collapse

Categories
Owner: DSX56000 User
Added: Jan 21, 2019 3:22 PM
catg... job ty... job cate...
☆☆☆☆ 0 reviews

Occupations
Owner: DSX56000 User
Added: Jan 21, 2019 3:18 PM
occup... jobs
☆☆☆☆ 0 reviews

Connection
Owner: DSX56000 User
Added: Jan 21, 2019 3:05 PM
☆☆☆☆ 0 reviews

female_human_trafficking.csv
Owner: DSX56000 User
Added: Jan 21, 2019 1:13 PM
traff... female...
☆☆☆☆ 0 reviews

Filter ?

Asset types
☒ Data Asset (3)
☐ Connection (1)

Tags
☐ categories (1)
☒ female human trafficking (1)
☐ job categories (1)
☐ job types (1)
☐ jobs (1)
☐ occupations (1)
☐ trafficking (1)
☐ untagged (1)

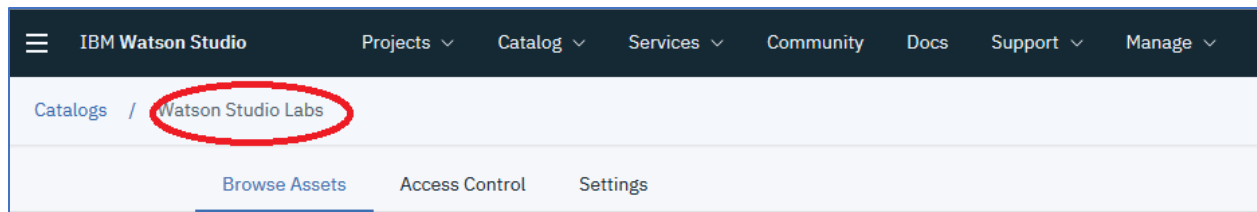
Available Assets
Showing 4 of 4 assets

NAME	OWNER	TAGS	TYPE	DATE ADDED
Categories	DSX56000 User	categories : job types : job categories	Data Asset	Jan 21, 2019
Occupations	DSX56000 User	occupations : jobs	Data Asset	Jan 21, 2019
Occupations	DSX56000 User		Connection	Jan 21, 2019
female_human_trafficking.csv	DSX56000 User	trafficking : female human trafficking	Data Asset	Jan 21, 2019

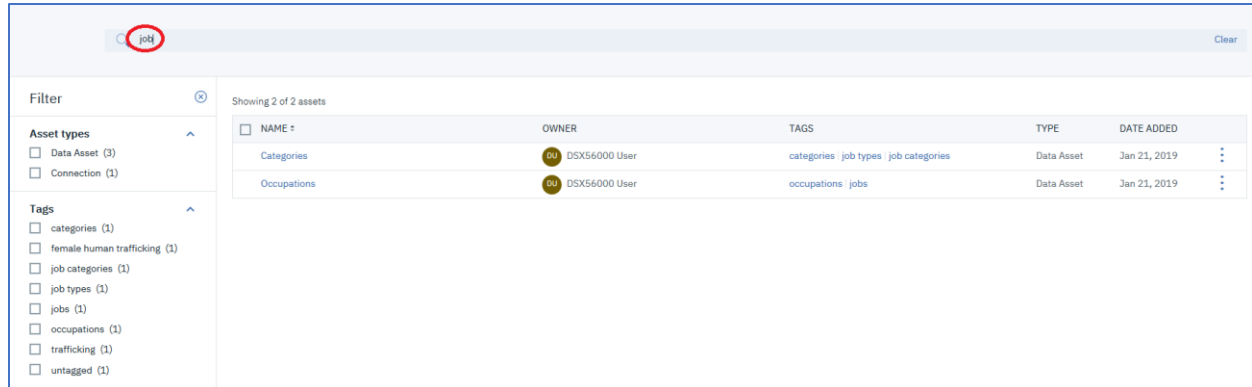
Step 4: Search the Catalog

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.

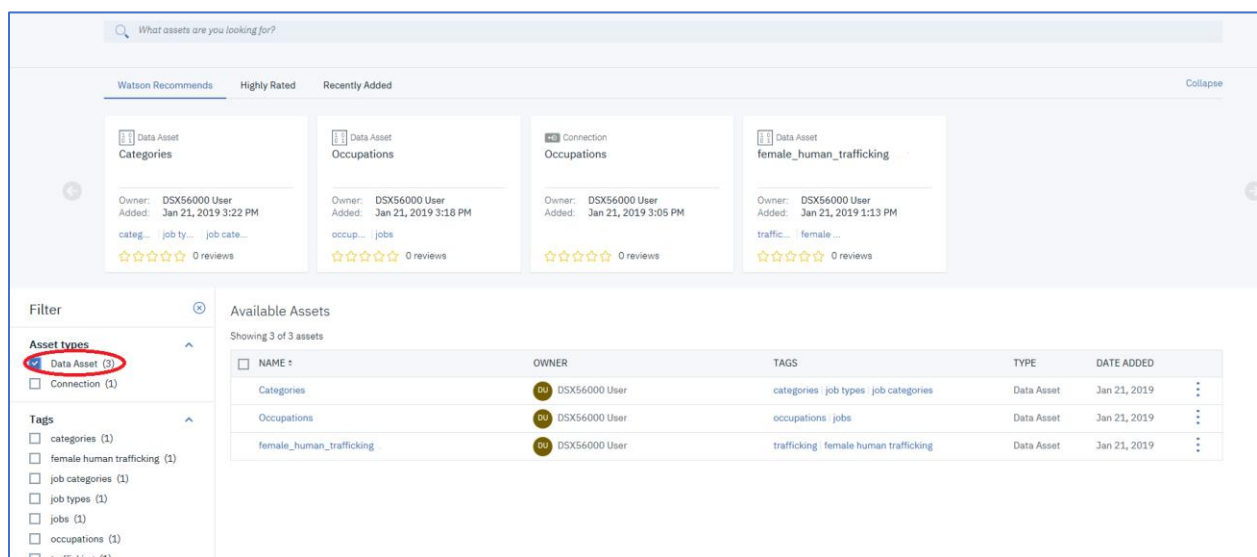
1. Click on **Watson Studio Labs** to return to the Catalog home screen.



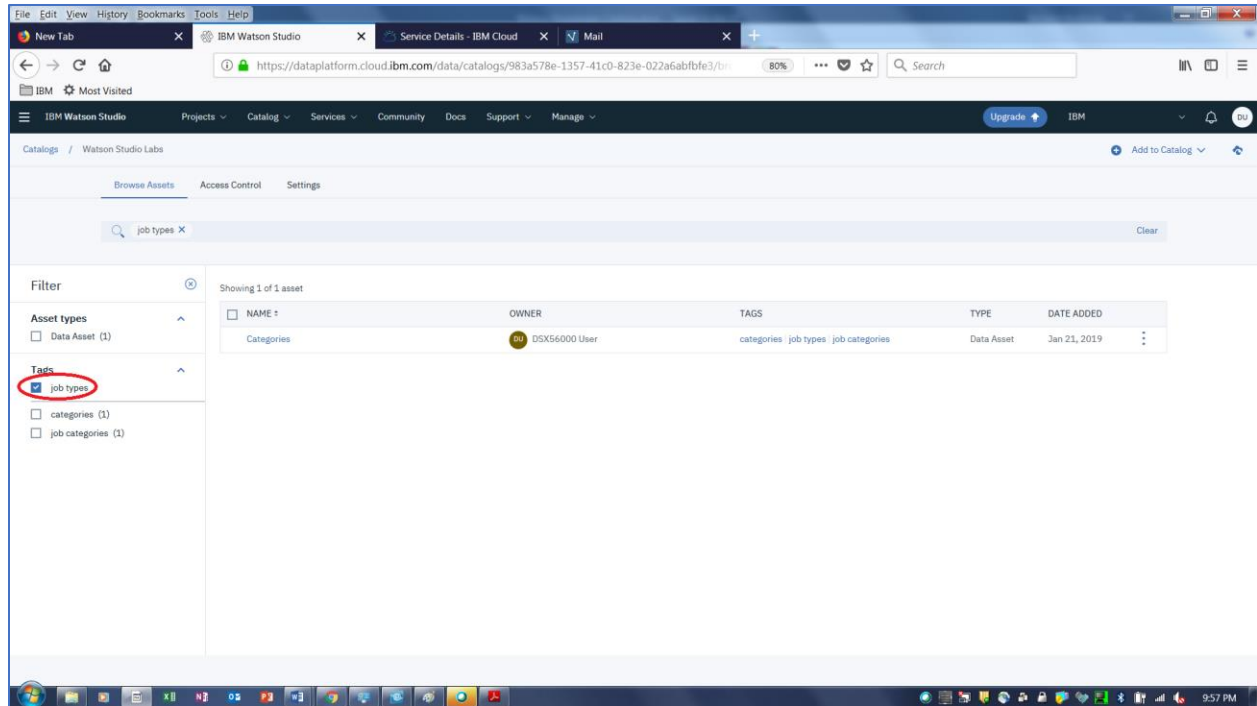
2. Enter “job” in the **search text field**. Data assets are displayed that contain the word or letters **job** in their Name, Description, or tags.



3. Clear the **search text field**. Click on the check box next to **Data Asset** in the Filter area. 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.



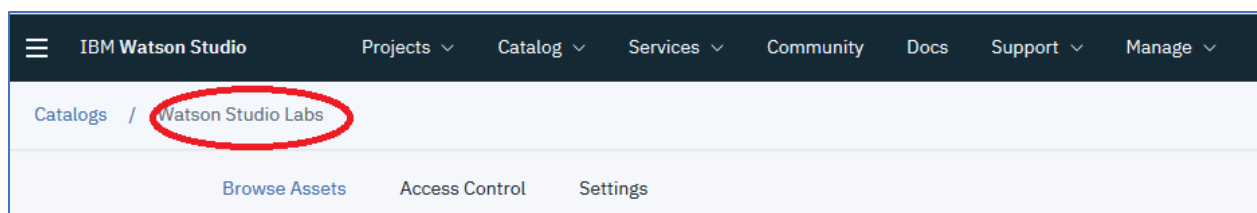
4. Click on the Tag **job types** in the Filter area. The Categories Data Asset is listed. Note that the Filter area has been updated to only list Asset Types, and Tags that correspond to the search result.



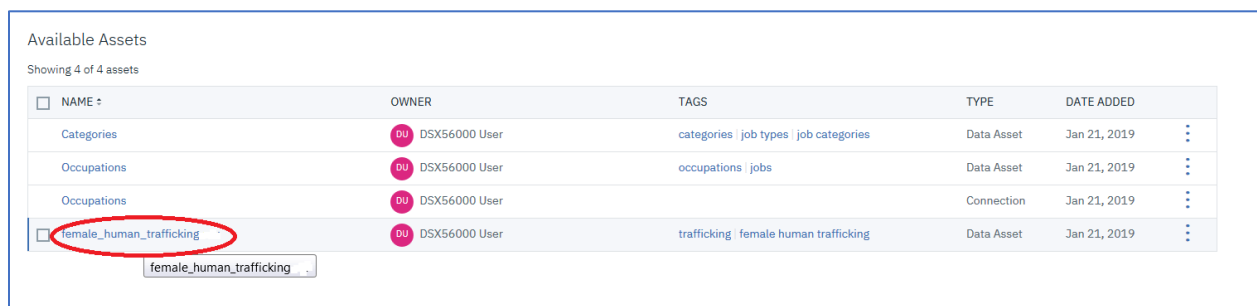
Step 5: Edit/Review/Profile Data Asset

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 Watson Studio Labs to return to the Catalog home screen.



2. Click on the `female_human_trafficking` asset.



- The metadata and data contents of the asset are displayed as shown below. Toward the top of the screen, are 5 tabs, circled in red 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 a number of 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

Overview Access Review Profile Lineage

Data Asset
female_human_trafficking

Remove Download Add to Project

Description
Travel data to be examined for trafficking
Added: Jan 21, 2019 1:13 PM
Format: Microsoft Excel

Tags
Tracking female human trafficking

Reviews
☆☆☆☆ 0 reviews

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

Classification
Personally identifiable information (PII) is defined as any data that could potentially identify a specific individual. Any information that can be used to distinguish one person from another can be considered PII.

Schema: 26 Columns
Preview: 1000 rows Last refresh: 2 hours ago Refresh

INTERNAL_ID	VETTING_LEVEL	DESCRIPTION	NAME	GENDER	BIRTH_DATE	BIRTH_COUNT...	BIRTH_COUNTRY_CODE	OCCUPATION	ADDRESS	SSN	PASSPORT
Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String
1	100	NA	Kathleen Bailey	F	1985-11-27	Ghana	GH	Chiroprapist	79421 Jordan On	885-71-9055	240479504
2	30	NA	Kelly Robbins	F	1971-01-17	Pakistan	PK	Engineer, structu	5808 Jacobs Uni	141-10-9199	449064032
3	100	NA	Linda Stewart	F	1997-04-15	Ghana	GH	Engineer, land	824 Kristin Grv, /	011-46-5304	639217816
4	30	NA	Stacey Courtney	F	1984-03-27	Ghana	GH	Careers adviser	322 Hutchinson	691-25-2647	212838449
5	30	NA	Erika Patie Fowle	F	1998-06-18	Ghana	GH	Geneticist, mole	07695 Michael V	165-33-0802	16514173
6	100	NA	Ashlee Fisher	F	1981-01-05	Ghana	GH	Electronics engin	4838 Cassandra	072-16-8742	513428119
7	100	NA	Jacqueline Clark	F	1976-01-19	Ghana	GH	Editor, commissi	1148 Wang Fall	634-03-1462	742709558
8	30	NA	Shelli Chapman	F	1997-08-24	Ghana	GH	Writer	72346 Wilson Fe	004-07-1446	170002642
9	10	NA	Sabrina Debra W	F	1977-02-21	Haiti	HT	Recruitment com	251 John Sh, Fir	233-53-1156	934214048
10	100	NA	Ashley Gibson	F	1986-04-06	Ghana	GH	Theatre manager	17894 Gary Grev	565-33-0366	3699030
11	100	NA	Tammy Sara Gan	F	1995-08-17	Ghana	GH	Statistician	2902 Rasmusser	069-16-0401	136448546
12	10	NA	Sandry Santiago	F	1971-12-16	Ghana	GH	Merchant navy of	9486 Pratt Wall,	778-07-9880	625553653
13	100	NA	Becky Cynthia St	F	1993-11-05	Ghana	GH	Advertising accou	6362 William Clin	612-01-4772	411247380

The actions circled in blue will **Remove** the asset from the catalog, **Download** the asset to the local file system, or add the asset to a catalog enable project (**Add to Project**).

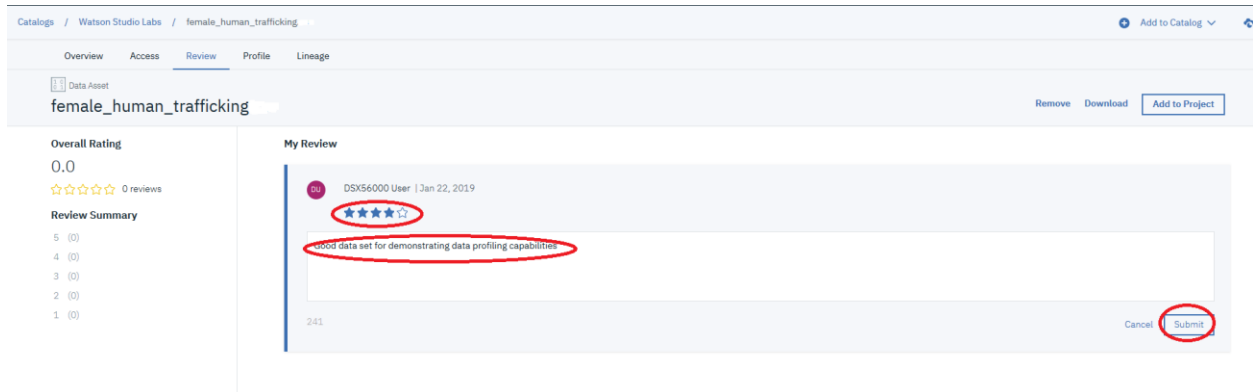
Hovering over the name of the data asset, the **Description**, the **Tags**, or the **Classification** will popup and Edit icon that allows the user to edit these entities.

- Click on the **Review** tab.

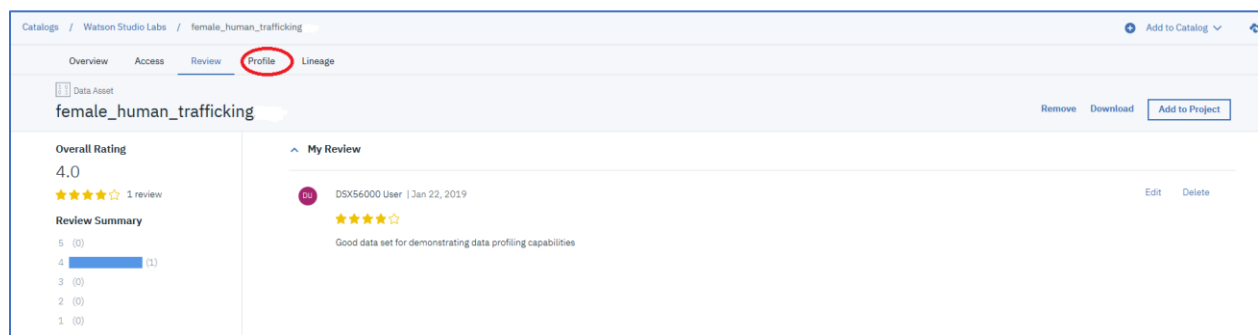
Catalogs / Watson Studio Labs / female_human_trafficking

Overview Access **Review** Profile Lineage

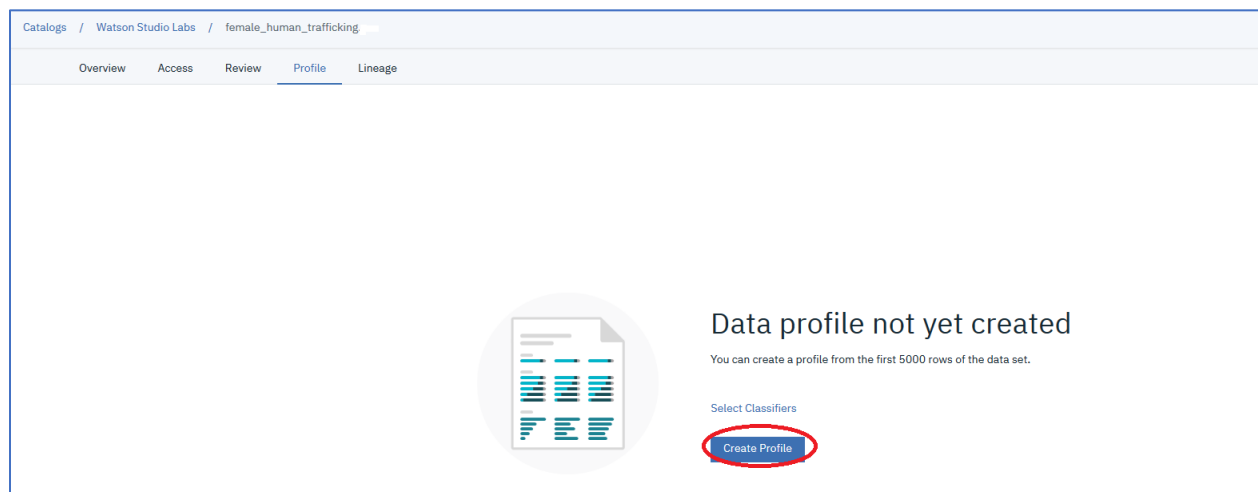
- Click on the fourth star to rate this asset. Enter “Good data set for demonstrating data profiling capabilities” in the **My Review** text field. Click on the **Submit** button.



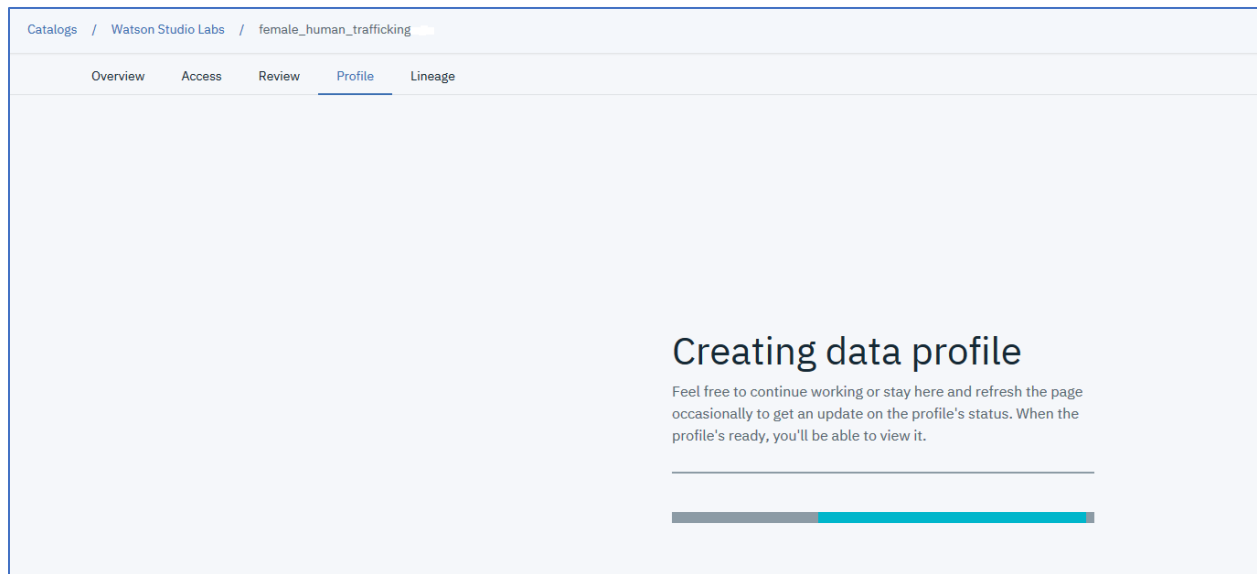
6. The review of the asset is displayed. Click on the Profile tab.



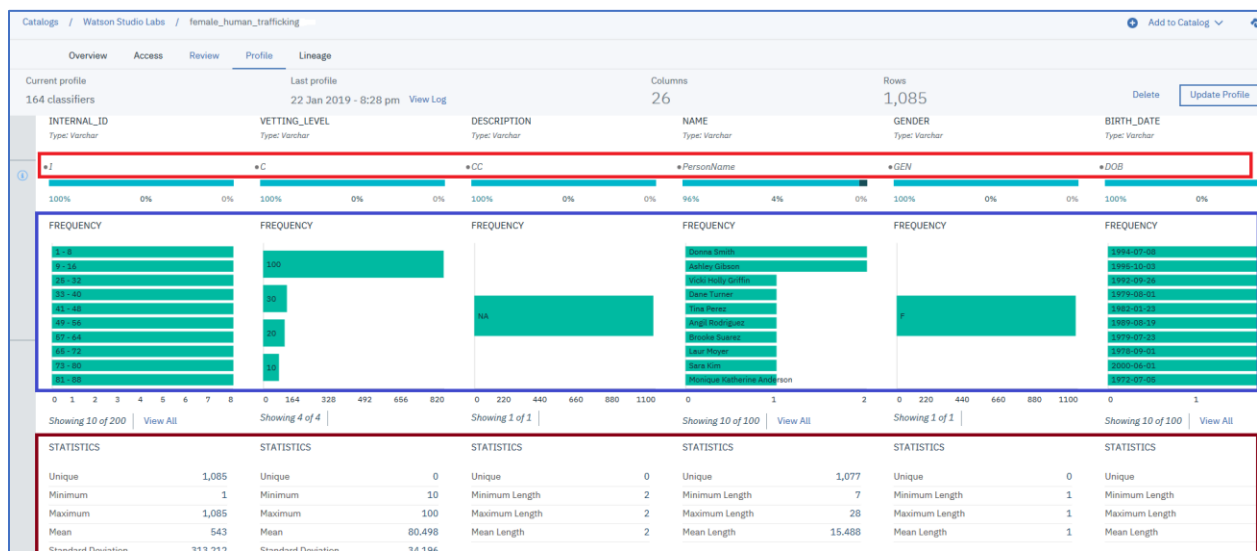
7. Click on **Create Profile**. If a profile was already created, it would be displayed instead of this prompt screen.



8. Wait a minute or so, and then click on browser refresh.



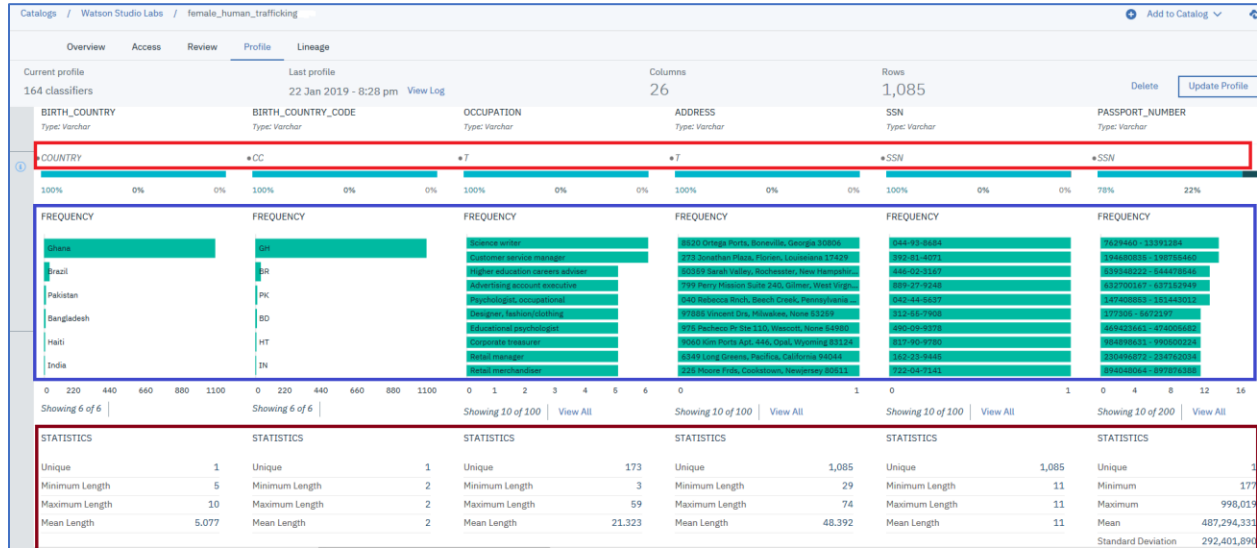
9. The profile information is displayed below in several screenshots based on horizontal scrolling to display many of the columns of the data asset. The highlighted red box contains the column classifications. The highlighted blue box contains counts of the number of occurrences of each value (for discrete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



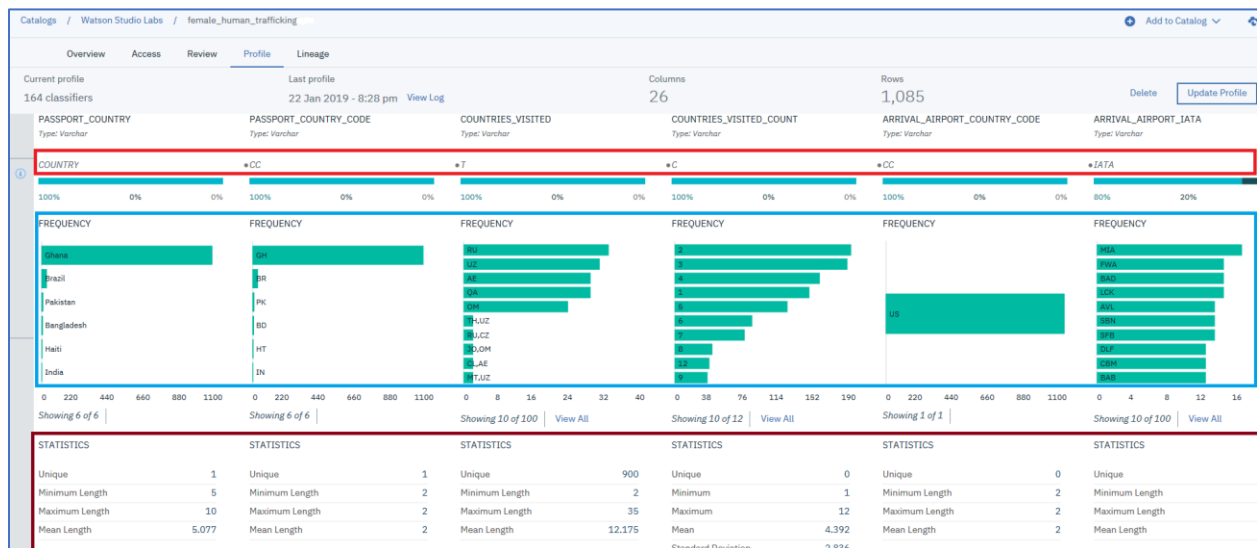
The classifications of the columns are listed below:

- INTERNAL_ID – I (Identifier)
- VETTING_LEVEL – C (Code)
- DESCRIPTION – CC (Country Code) – this is an error due to the description input as NA. We can correct this misclassification.

- NAME: PersonName – the classification accurately determines that this field is the name of a person.
- BIRTH_DATE: DOB – the classification accurately determines that this field represents a date of birth.



- BIRTH_COUNTRY: COUNTRY- correctly classified as a country
- BIRTH_COUNTRY_CODE – CC (Country Code) - correctly classified
- OCCUPATION – T (Text field) – there is not a classifier for occupations so it determines that this is a text string.
- ADDRESS: T (Text Field)
- SSN:SSN – correctly classifies the column as a social security number.
- PASSPORT_NUMBER – incorrectly classifies as a social security number. This can be corrected, although it may be useful to treat as a SSN for privacy reasons.



- PASSPORT_COUNTRY: COUNTRY- correctly classified as a country

- PASSPORT_COUNTRY_CODE: CC(Country Code) - correctly classified
- COUNTRIES_VISITED – T (Text field)
- COUNTRY_VISITED_COUNT – C (Code) – this is an error. It's not a code. This can be re-classified if necessary.
- ARRIVAL_AIRPORT_COUNTRY_CODE: COUNTRY –
- ARRIVAL_AIRPORT_IATA: IATA – the classification accurately determines that this field represents an IATA.

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

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:

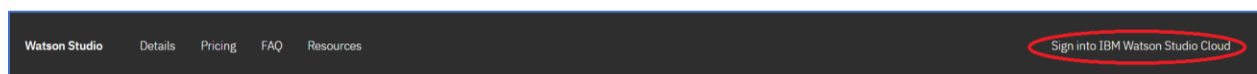
1. Catalog owner (logged on user)
2. DSXUser57000@gmail.com

Data Asset Privacy:

1. female_human_trafficking (Private) ACL Members:
 - a. asset owner (logged on user)
 - b. DSXUser57000@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 DSXUser57000@gmail.com, that only 3 Data Assets would appear. The two public Data Assets, and the Private Data Asset where DSXUser57000@gmail is on the access control list. This user would not have access to the Occupations asseet, 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 datascience.ibm.com for the URL to navigate to.
3. Click on **Sign into IBM Watson Studio Cloud**



4. Login using DSXUser57000@gmail.com for the user name and DSXPassw0rd57000! for the password.

Log in to IBM

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

Log in with your company
credentials (SSO)

Need help? [Contact the IBMid Helpdesk](#)

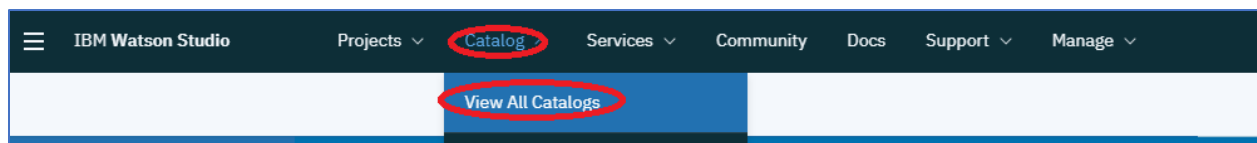
IBMid [Forgot IBMid?](#)

DSXUser57000@gmail.com

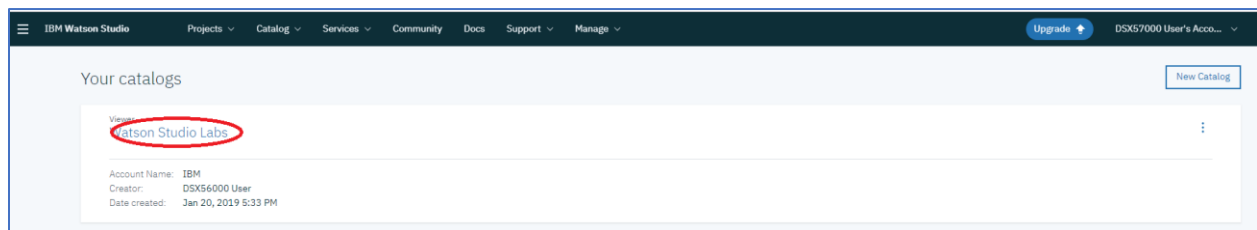
Password [Forgot password?](#)

☐ Remember me ⓘ

5. Click on **Catalog** and then **View All Catalogs**.



6. Click on the **Watson Studio Labs** catalog that corresponds to your username. Note that there will be 1 catalog per lab participant.



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

Catalogs / Watson Studio Labs

Browse Assets Access Control Settings

What assets are you looking for?

Watson Recommends Highly Rated Recently Added Collapse

Data Asset Categories
Owner: DSX56000 User
Added: Jan 21, 2019 3:22 PM
categor... job ty... job cate...
★★★★★ 0 reviews

Connection Occupations
Owner: DSX56000 User
Added: Jan 21, 2019 3:05 PM
★★★★★ 0 reviews

Data Asset female_human_trafficking.csv
Owner: DSX56000 User
Added: Jan 21, 2019 1:13 PM
traffick... female...
★★★★★ 1 review

Filter

Asset types
☐ Data Asset (2)
☐ Connection (1)

Tags
☐ categories (1)
☐ female human trafficking (1)
☐ job categories (1)

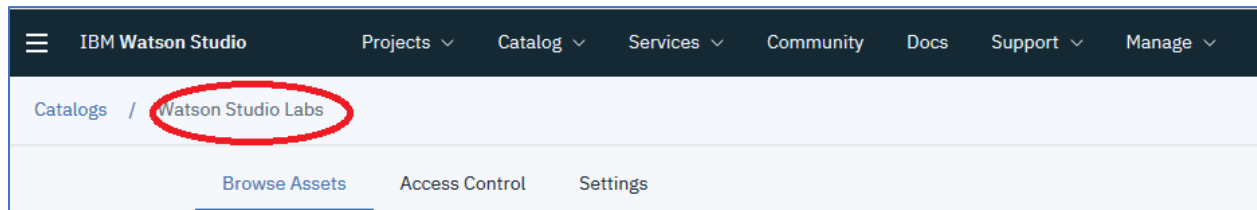
Available Assets
Showing 3 of 3 assets

NAME	OWNER	TAGS	TYPE	DATE ADDED
Categories	DSX56000 User	categories job types job categories	Data Asset	Jan 21, 2019
Occupations	DSX56000 User		Connection	Jan 21, 2019
female_human_trafficking.csv	DSX56000 User	trafficking female human trafficking	Data Asset	Jan 21, 2019

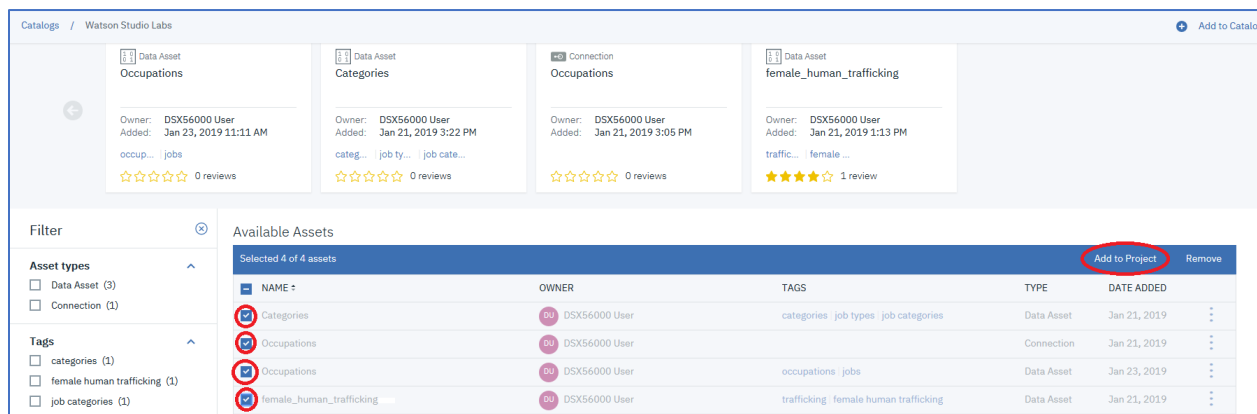
Step 8: Push the data assets to the project

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

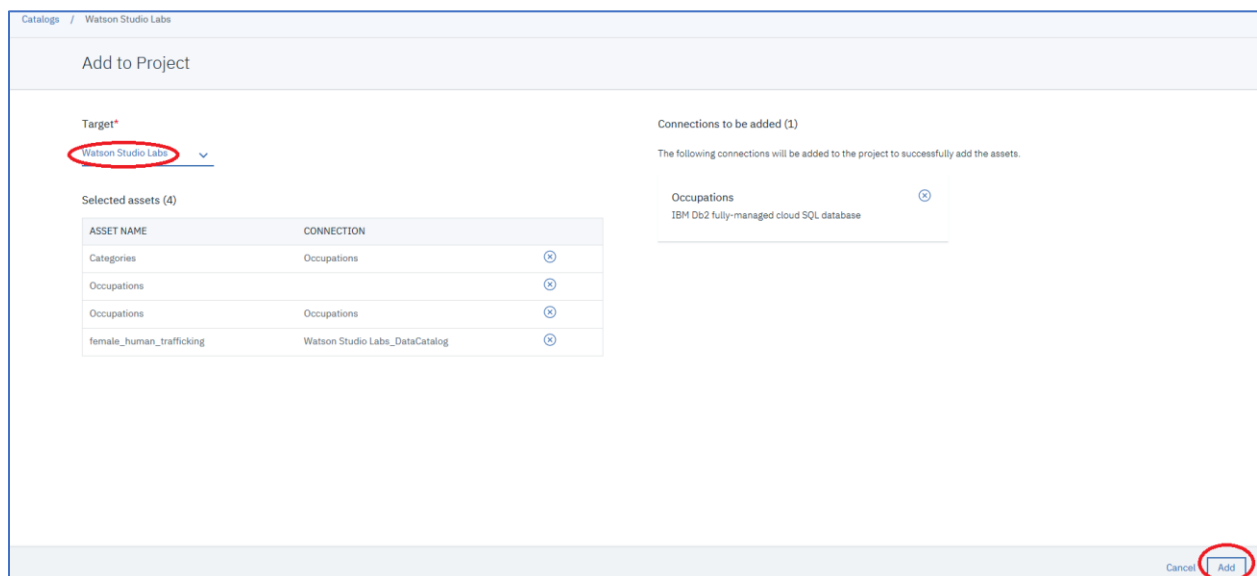
1. Click on Watson Studio Lab to go to the Catalog home screen.



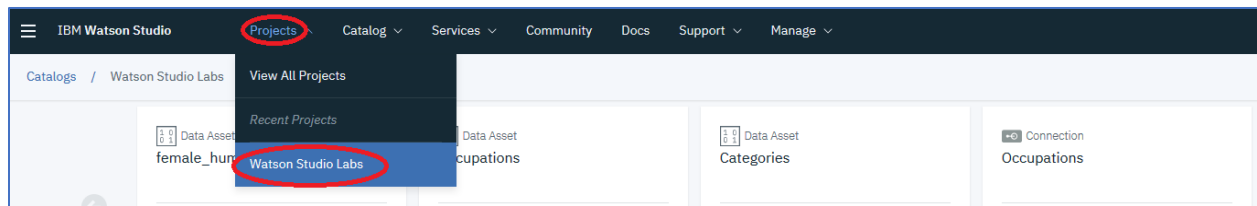
2. Hover the cursor to the left of each data asset and then click on the checkbox. Click on **Add to Project**.



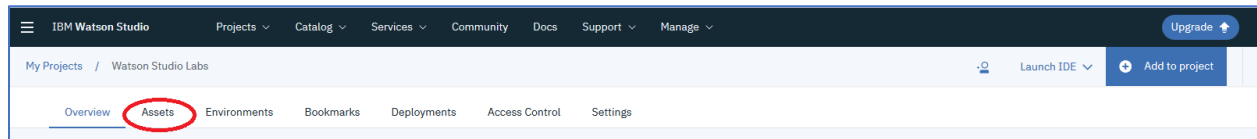
3. Select the Watson Studio Labs as the **Target**. Click on **Add**.



4. Click on **Projects** and then **Watson Studio Labs**.



5. Click on the **Assets** tab.



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

