Tech 101: Data Workshop Discovering and Managing Data

Data & Governance Lab
Nigel Jones, Sepi Seifzadeh, David Radley

Data governance & privacy tutorial

Full lab instructions at:

https://dataplatform.cloud.ibm.com/docs/content/wsj/getting-started/df_gov_trust.html

You should already have an IBM cloud account – information was sent in a prior email (ask if you need help)

The lab demonstrates how Watson Knowledge Catalog can be used to catalog assets, create business times, establish relationships between the assets and terms, and finally publish the content in a catalog available to users.

The instructions include explanatory text & video.

If you get stuck ask questions on the web conference, or in the chat

Data governance & privacy tutorial

> Task 0: Preparation

Task 1: Create a catalog

Task 2: Create a category

Task 3: Add business terms

Task 4: Import Data to a project

Task 5: Enrich the imported data

Task 6: View the results of the metadata enrichment

Task 7: Publish Data to a catalog

In this task we:

- Ensure you are signed up for Cloud Pak for Data as a Service & logged in

- Provision the services needed for the lab

- Create a project from a tutorial template

- Import Banking terms

Data governance & privacy tutorial

Task 0: Preparation

> **Task 1: Create a catalog**

Task 2: Create a category

Task 3: Add business terms

Task 4: Import Data to a project

Task 5: Enrich the imported data

Task 6: View the results of the metadata enrichment

Task 7: Publish Data to a catalog

This step creates a catalog in Watson Knowledge Catalog where data will be published to

Data governance & privacy tutorial

Task 0: Preparation

Task 1: Create a catalog

> Task 2: Create a category

Task 3: Add business terms

Task 4: Import Data to a project

Task 5: Enrich the imported data

Task 6: View the results of the metadata enrichment

Task 7: Publish Data to a catalog

We create a category to contain business terms associated with Banking

In a real environment a banking environment will use many categories which act as folders to organize the terms

Data governance & privacy tutorial

Task 0: Preparation

Task 1: Create a catalog

Task 2: Create a category

> Task 3: Add business terms

Task 4: Import Data to a project

Task 5: Enrich the imported data

Task 6: View the results of the metadata enrichment

Task 7: Publish Data to a catalog

We add banking related business terms (definitions of concepts in banking) from a .csv file into the category we just created.

Terms are initially created as draft so that they can be reviewed.

We then publish the business terms to make them available to users

Data governance & privacy tutorial

Task 0: Preparation

Task 1: Create a catalog

Task 2: Create a category

Task 3: Add business terms

> **Task 4: Import Data to a project**

Task 5: Enrich the imported data

Task 6: View the results of the metadata enrichment

Task 7: Publish Data to a catalog

We import technical metadata (ie names of tables, columns etc) relating to mortgage data into a new project

Assets are then created in a project

Data governance & privacy tutorial

Task 0: Preparation

Task 1: Create a catalog

Task 2: Create a category

Task 3: Add business terms

Task 4: Import Data to a project

> Task 5: Enrich the imported data

Task 6: View the results of the metadata enrichment

Task 7: Publish Data to a catalog

We run some algorithms which discover more metadata relating to the banking assets:

- Profile information ie data distribution

- Quality assessment

- Linking business terms to the assets

Data governance & privacy tutorial

Task 0: Preparation

Task 1: Create a catalog

Task 2: Create a category

Task 3: Add business terms

Task 4: Import Data to a project

Task 5: Enrich the imported data

> Task 6: View the results of the metadata enrichment

Task 7: Publish Data to a catalog

This step takes you through looking at the results of the previous step to see what additional metadata has been discovered.

It also shows how some suggestions may be recommended, but must be accepted – depending on confidence

Data governance & privacy tutorial

Task 0: Preparation

Task 1: Create a catalog

Task 2: Create a category

Task 3: Add business terms

Task 4: Import Data to a project

Task 5: Enrich the imported data

Task 6: View the results of the metadata enrichment

> Task 7: Publish Data to a catalog

Enriched data can now be published to a catalog so that others users including data scientists and analysts can use the trusted data

Data governance & privacy tutorial

OPTIONAL - Protect your data

If you have completed the previous steps, you can continue on in the tutorials.

The next section walks through how to apply data protection rules to the assets to control what users can see – perhaps being denied access, or being subject to data masking

You can work through additional exercises in your own time.