



Governed MLOps Workshop

Set up data governance

Document version: June 2023

DISCLAIMER

IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. Information regarding potential future products is intended to outline potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract.

The development, release, and timing of any future features or functionality described for our products remains at our sole discretion I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results like those stated here.

Information in these presentations (including information relating to products that have not yet been announced by IBM) has been reviewed for accuracy as of the date of initial publication and could include unintentional technical or typographical errors. IBM shall have no responsibility to update this information. **This document is distributed "as is" without any warranty, either express or implied. In no event, shall IBM be liable for any damage arising from the use of this information, including but not limited to, loss of data, business interruption, loss of profit or loss of opportunity.** IBM products and services are warranted per the terms and conditions of the agreements under which they are provided.

IBM products are manufactured from new parts or new and used parts. In some cases, a product may not be new and may have been previously installed. Regardless, our warranty terms apply."

Any statements regarding IBM's future direction, intent or product plans are subject to change or withdrawal without notice.

Performance data contained herein was generally obtained in controlled, isolated environments. Customer examples are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary. References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business.

Workshops, sessions and associated materials may have been prepared by independent session speakers, and do not necessarily reflect the views of IBM. All materials and discussions are provided for informational purposes only, and are neither intended to, nor shall constitute legal or other guidance or advice to any individual participant or their specific situation.

It is the customer's responsibility to insure its own compliance with legal requirements and to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer's business and any actions the customer may need to take to comply with such laws. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer follows any law.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products about this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. IBM does not warrant the quality of any third-party products, or the ability of any such third-party products to interoperate with IBM's products. **IBM expressly disclaims all warranties, expressed or implied, including but not limited to, the implied warranties of merchantability and fitness for a purpose.**

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents, copyrights, trademarks or other intellectual property right.

IBM, the IBM logo, and ibm.com are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at: www.ibm.com/legal/copytrade.shtml.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

OpenShift is a trademark of Red Hat, Inc.

UNIX is a registered trademark of The Open Group in the United States and other countries.

© 2023 International Business Machines Corporation. No part of this document may be reproduced or transmitted in any form without written permission from IBM.

Table of Contents

Introduction	4
Log into Cloud Pak for Data	5
Categories	6
Workflows.....	13
Business terms	25
Governance Policies.....	34
Governance rules.....	39
Reference data sets and hierarchies	47
Data classes	56
Summary	62

Introduction

Governance is the process of curating, enriching, and controlling your data. The task of setting up a governance foundation for an enterprise is not specific to one use case but rather aligned to the governance and compliance requirements of the enterprise. Certain use cases may require defining additional governance artifacts.

IBM Watson Knowledge Catalog is an intelligent data catalog that powers self-service discovery of data, models and more. With Watson Knowledge Catalog, users can access, curate, and share data, knowledge assets and their relationships wherever they reside.

Watson Knowledge Catalog supports the creation of governance artifacts such as business terms, classifications, data classes, reference data sets, policies, and rules which can be used throughout IBM Cloud Pak for Data. The governance artifacts can be organized using categories. Watson Knowledge Catalog also supports workflows for governance artifacts which enforce task-based processes (which can be automated) to control the creating, updating, deleting, and importing of governance artifacts.

In this tutorial, you will create the categories and the governance artifacts in IBM Cloud Pak for Data that are required for a Telco Churn use case. The governance artifacts being created in this tutorial are derived from the [Telco Churn Industry Accelerator](#).

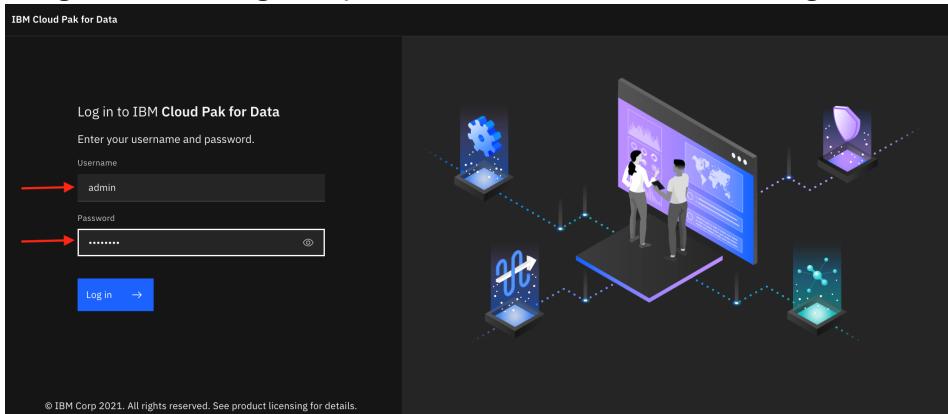
Learning objectives

In this tutorial, you will learn about and create:

- Categories
- Governance artifact workflows
- Business terms
- Governance policies
- Governance rules
- Reference data
- Data classes

Log into Cloud Pak for Data

1. Using a browser, navigate to your Cloud Pak for Data instance and log in as the admin user.



The admin user has the “Administrator” user role in IBM Cloud Pak for Data which includes the following permissions.

Note: the permissions below could be assigned to a role other than the Administrator):

- Manage governance categories
- Manage governance workflows

Categories

Watson Knowledge Catalog supports organizing governance artifacts using categories. A category is like a folder or directory that organizes your governance artifacts and the users who can view and manage those artifacts. Categories provide the logical structure for all types of governance artifacts, *except data protection rules*. You group your governance artifacts in categories to make them easier to find, control their visibility, and to manage them.

Categories can be organized in a hierarchy based on their meaning and relationships to one another. A category can have subcategories, but a subcategory can have only one direct parent category.

You can create categories manually or import them using a CSV file.

Create a category manually

1. Go to the **hamburger (≡)** menu in the upper-left corner, expand **Governance** and click **Categories**.

The screenshot shows the IBM Cloud Pak for Data interface. The top navigation bar includes a logo, a search bar, and user icons. On the left, a sidebar menu is open under the 'Governance' section. The 'Categories' option is highlighted with a red box. The main area displays an 'Overview' section with three cards: 'Recent projects' (0 recent projects), 'Requests' (0 data requests), and 'Notifications' (0 notifications). To the right is a dark-themed dashboard with a central image of a cluster of cubes and a magnifying glass.

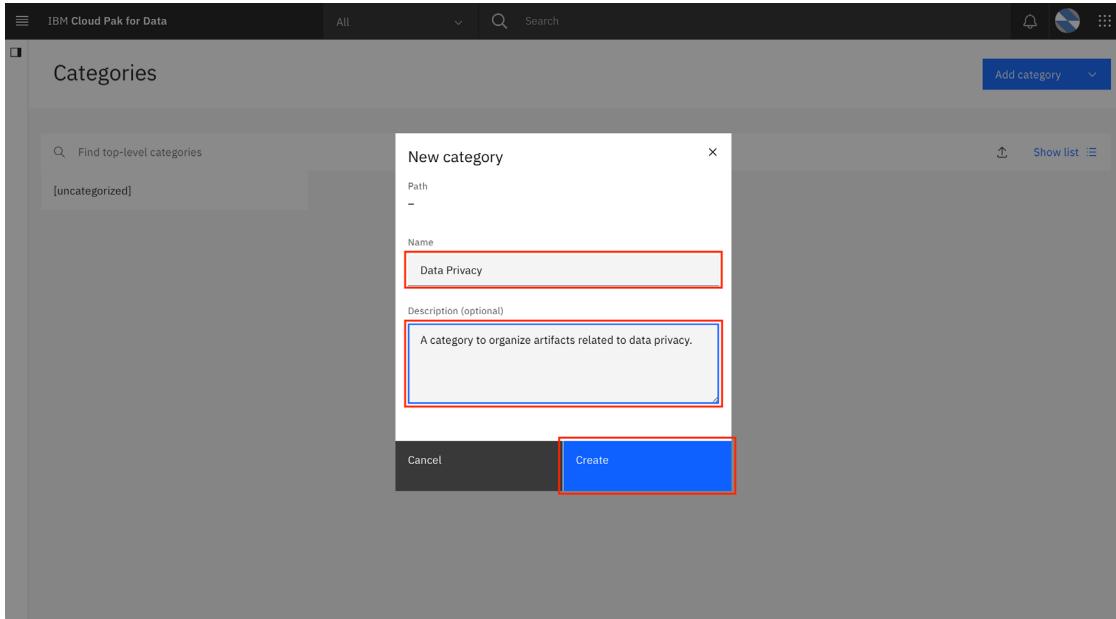
2. Add a new category by clicking on **Add category > New category**.

The screenshot shows the 'Categories' page. The top navigation bar is visible. The main content area shows a list of categories, with '[uncategorized]' listed at the bottom. On the right, a vertical menu titled 'Add category' has a sub-item 'New category' highlighted with a red box. Other options in the menu include 'Import from file' and 'Import from catalog'.

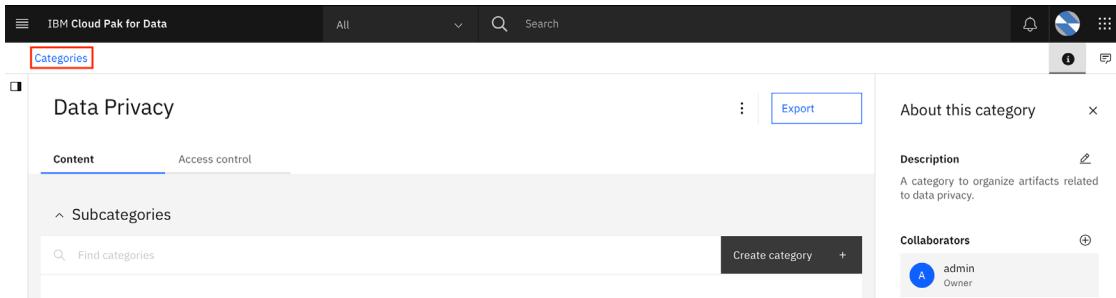
3. Provide the name for the category as **Data Privacy**.

Provide an optional description (**A category to organize artifacts related to data privacy.**) for the category.

Click **Create**.

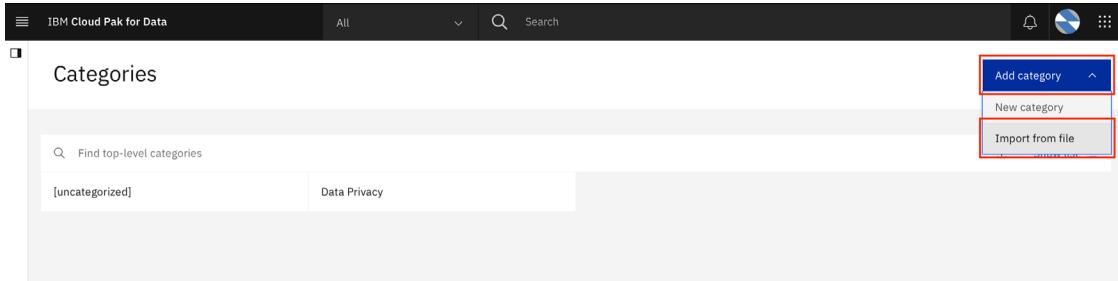


4. The Data Privacy category is created. Go back to the Categories main page by clicking on **Categories** in the breadcrumbs.

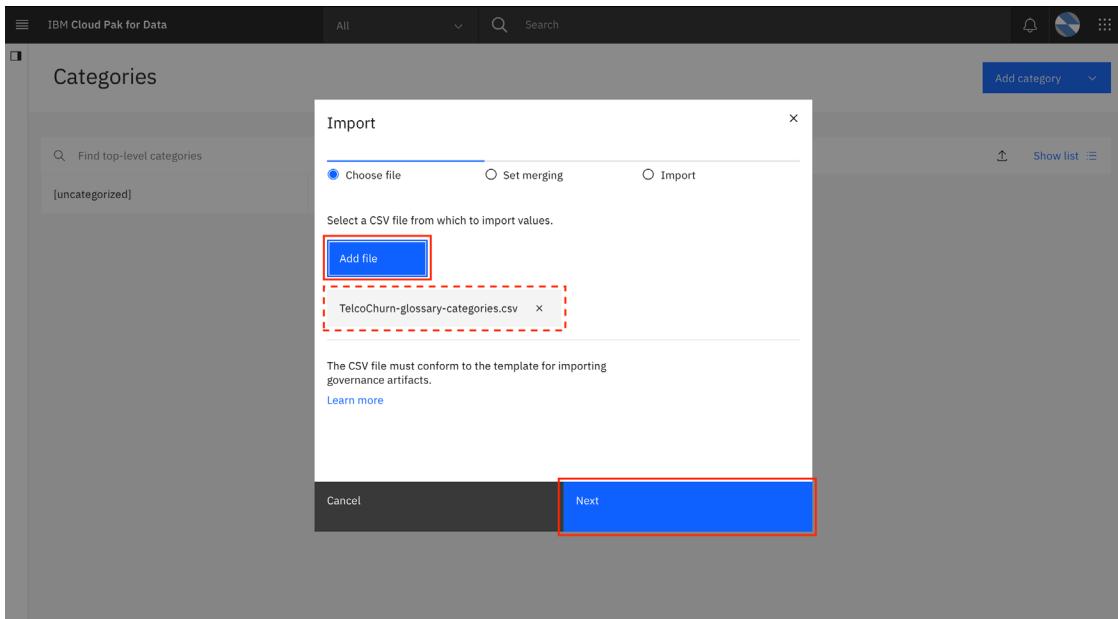


Import categories by importing a file

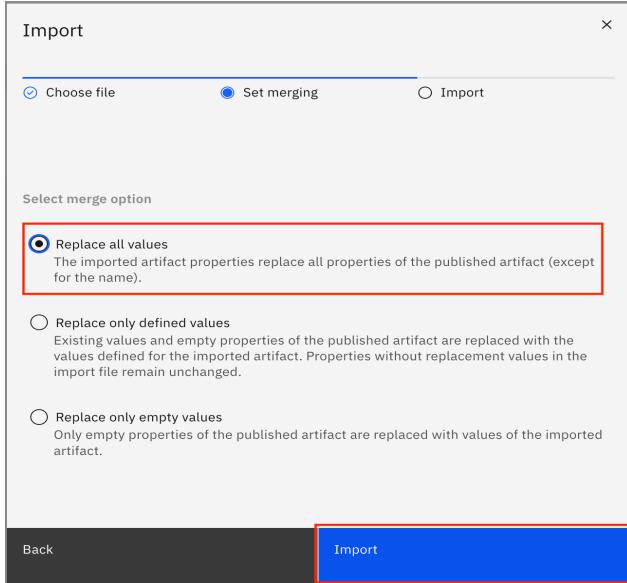
1. Download the [TelcoChurn-glossary-categories.csv](#) file.
2. Click Add category > Import from file.



3. Click Add file and select the TelcoChurn-glossary-categories.csv file you downloaded earlier, then click Next.



- After reviewing the different merge options, select **Replace all values**, then click **Import**.



- The categories will be imported from the file. Once the import is successful, you will see the *Import summary* modal that says, “The import completed successfully.” It will also show you that three new categories were created, and no errors were encountered.

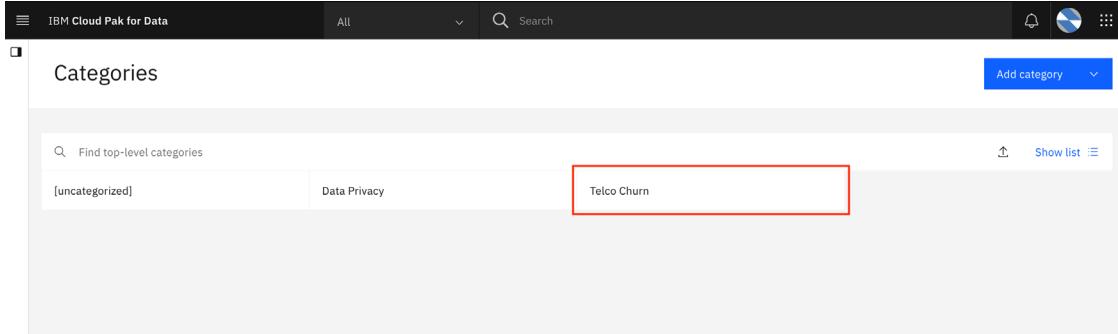
Click **Close** to go back.



- The categories that were created by the import process are the “Telco Churn” category and its subcategories - “Personal Info” and “Customer Data”. Please consult the [Telco Churn Industry Accelerator](#) for more details on the overall solution for telco churn prediction.

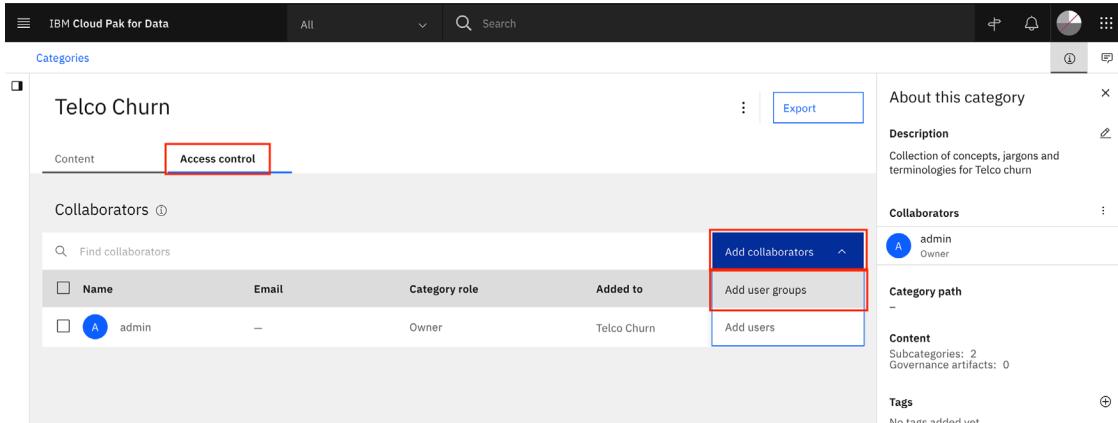
Add collaborators to categories

1. Refresh the webpage to see the newly imported categories. Click on the **Telco Churn** category.



The screenshot shows the 'Categories' section of the IBM Cloud Pak for Data interface. At the top, there's a search bar and a 'Find top-level categories' input field. Below it, a list of categories includes '[uncategorized]', 'Data Privacy', and 'Telco Churn'. The 'Telco Churn' category is highlighted with a red box. On the right side of the screen, there are various navigation and search icons.

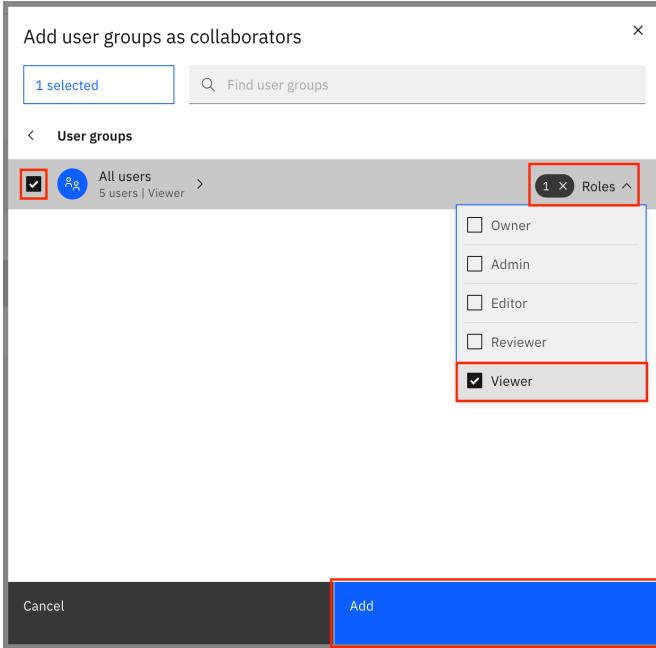
2. When first created, the categories will only have a Name, Description, and a default set of Collaborators. To add collaborators to the Telco Churn category, go to the Access Control tab by clicking on **Access control**. Click **Add collaborators > Add user groups**.



The screenshot shows the detailed view of the 'Telco Churn' category. The 'Content' tab is selected, while the 'Access control' tab is highlighted with a red box. In the 'Collaborators' section, there's a table with one entry: 'admin' (Owner). To the right of the table, there are three buttons: 'Add collaborators' (highlighted with a blue box), 'Add user groups' (highlighted with a red box), and 'Add users'. On the right side of the page, there are sections for 'About this category', 'Description' (Collection of concepts, jargons and terminologies for Telco churn), 'Collaborators' (admin, Owner), 'Category path' (empty), 'Content' (Subcategories: 2, Governance artifacts: 0), and 'Tags' (empty).

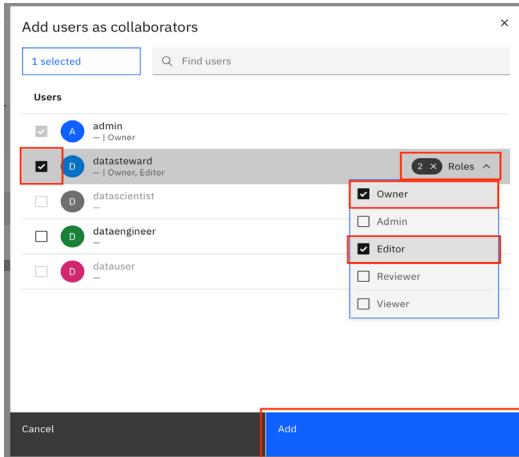
Note: To add a user or user group as a collaborator to a category, that user or user group needs to have the Cloud Pak for Data platform permission to “Access governance artifacts”.

3. Add the “All users” group as Viewers of the Telco Churn category. Select the checkbox next to the user group **All users**. Click on **Roles** and select the role(s) you wish to add – in this case, click on **Viewer**. Click **Add**.



4. Back on the Access Control tab of the Telco Churn category, click **Add collaborators > Add users**.

5. Add the user “datasteward” as the Owner and Editor of the Telco Churn category:
 - Select the checkbox next to the user **datasteward**.
 - Click on **Roles** and select the roles you wish to add – **Owner** and **Editor**. Click **Add**.



6. When you add users to a category, the users are added with the same permissions to all the subcategories of that category.

Back on the **Content** tab, you can click on the name of one of the subcategories of “Telco Churn” category to verify the same.

Finally, click on **Categories** in the breadcrumbs to go to the “Categories” main page.

Collaborator	Role
admin	Owner
All users	Viewer
datasteward	Editor, Owner

7. Repeat the steps above to add the “datasteward” user as the Owner/Editor for the remaining category, “Data Privacy”. The “All users” group has already been provided with the Viewer role for the Data Privacy category.

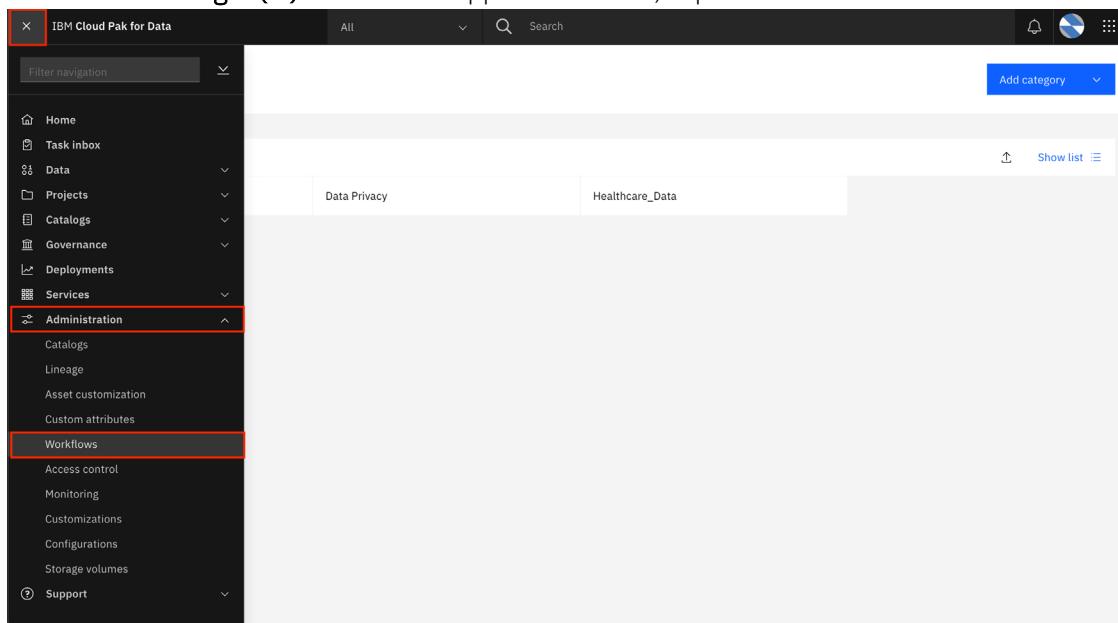
Workflows

Watson Knowledge Catalog allows you to create a custom request process (also called a [global workflow](#)) which delivers more freedom to build completely custom workflows triggered by a request form. Examples of workflows that could be built using this framework include a model request process that notifies data scientists that a new model has been requested by a line of business, and a new user request that notifies admins to set up permissions for a new user in the organization.

You will create a governance artifact request process, so that other people in your organization can request changes or updates to governance artifacts in an existing category.

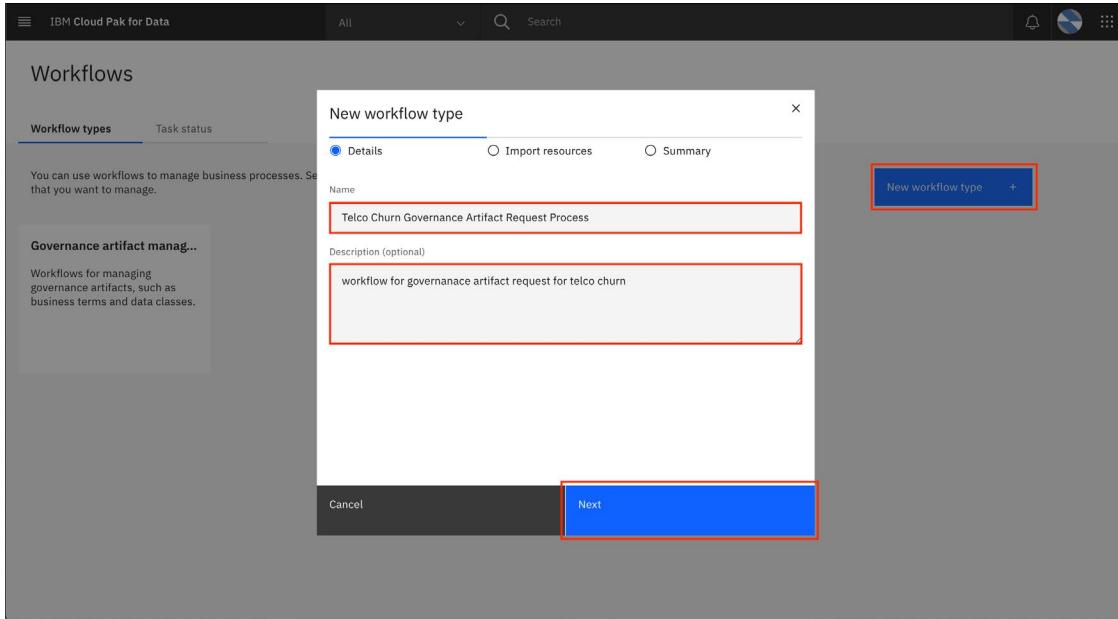
Create the governance artifact request process

1. Download the [Governance_change_request_with_priority.bpmn20.xml](#) file. This is a workflow definition file that was created using the free open source [Flowable](#) web designer. You can look at the [CP4DWorkflows](#) github repo for more information on using the Flowable web designer to create new workflow types in Watson Knowledge Catalog.
2. Go to the **hamburger (≡)** menu in the upper-left corner, expand **Administration** and click **Workflows**.

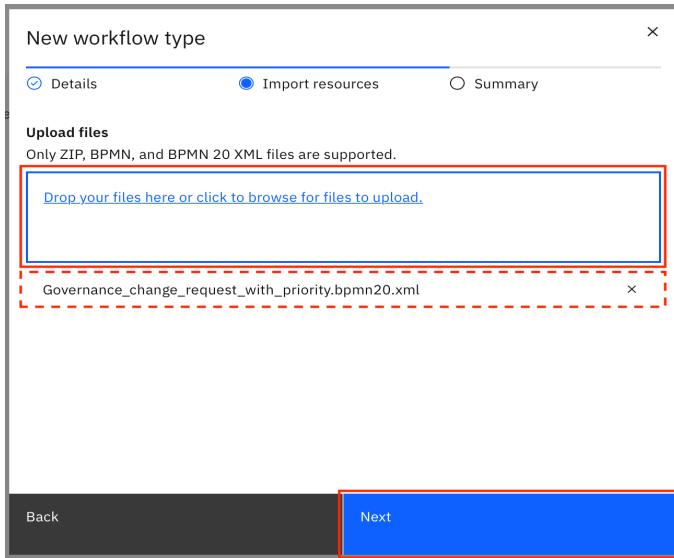


3. Click **New workflow type**.

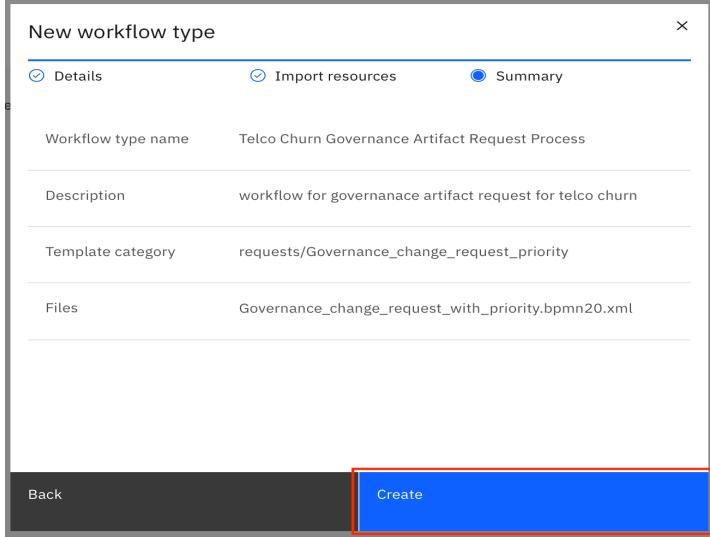
Provide a Name (Telco Churn Governance Artifact Request Process) and an optional Description (Workflow for governance artifact request for telco churn) for the new workflow type and click **Next**.



4. Upload the workflow definition file “*Governance_change_request_with_priority.bpmn20.xml*” that you had downloaded earlier and click **Next**.



5. Review the details of the new workflow type on the Summary page and click **Create**.

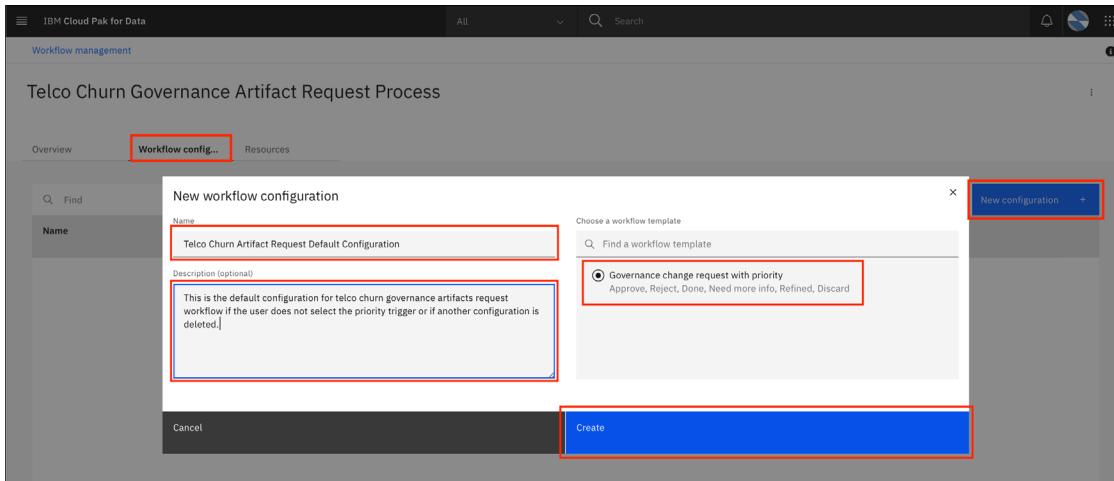


6. Click **Workflow configurations** to go to the “Workflow configurations” tab.

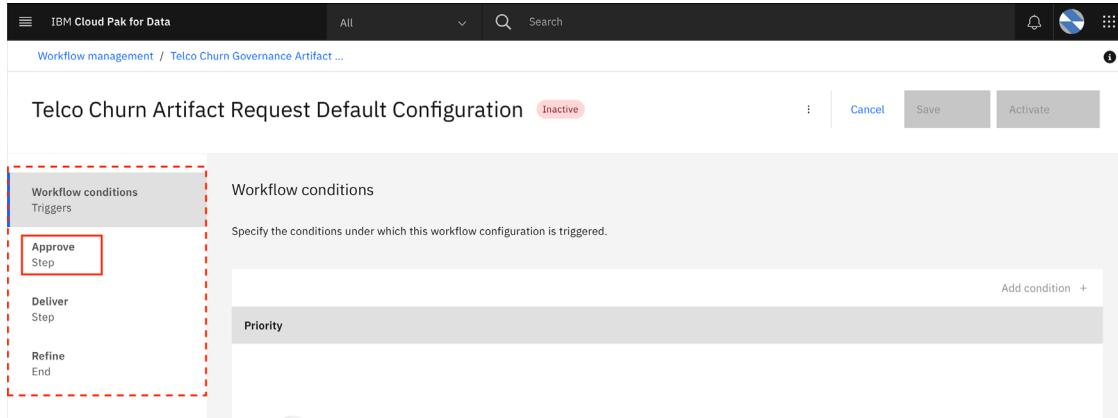
Click the **New configuration +** button to create the default configuration. The default configuration is used if trigger is left blank by the requester or if the configuration is deleted.

In the pop-up window, provide a Name (**Telco Churn Artifact Request Default Configuration**) and an optional Description (**This is the default configuration for telco churn governance artifacts request workflow if the user does not select the priority trigger or if another configuration is deleted.**) for the default configuration for your governance artifact workflow request.

Select the radio button to select the **Governance change request with priority** template which was created earlier and click **Create**.

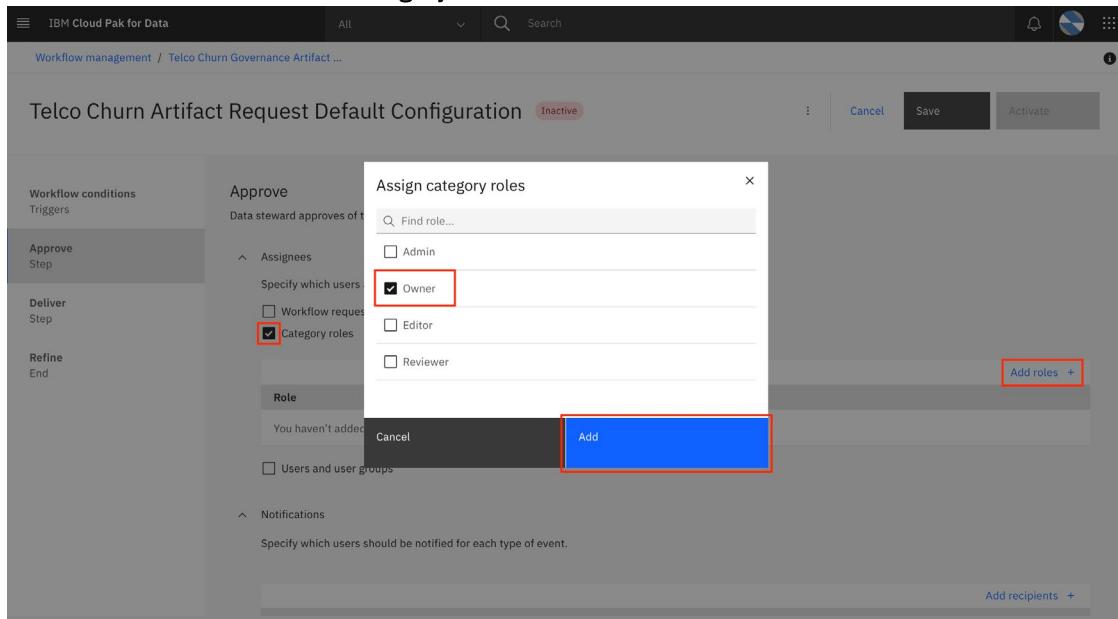


7. On the next page, specify the details of this configuration based on the workflow template. For this template, there are 4 steps as outlined in the left panel. Since this is the default configuration, you can't choose any conditions to trigger the workflow. Click on **Approve** in the left panel to go to the Approve step.



8. You can select who can approve submitted requests for governance artifacts. For now, select the Owner of the Category as the individual responsible for approving requests for governance artifacts associated with that category.

Select the check box next to **Category roles**, then click on **Add roles +**. Select **Owner** and click **Add**.



9. Scroll down to the **Notifications** section. Select the checkbox under **Task assigned** for Assignee to indicate that the Assignee should get a notification when a request is submitted.

The screenshot shows the 'Telco Churn Artifact Request Default Configuration' page in IBM Cloud Pak for Data. The left sidebar lists steps: Approve, Deliver, and Refine. The 'Deliver Step' is selected. The main area has sections for 'Assignees' and 'Notifications'. Under 'Assignees', 'Category roles' is checked. Under 'Notifications', the 'Recipient' table shows 'Assignee' with the 'Task assigned' checkbox checked (highlighted with a red box). Other checkboxes for 'Approve', 'Reject', and 'Overdue' are unselected.

10. Go to the Deliver step by clicking on **Deliver** in the left panel. Here you will need to specify who must take action and/or deliver a response to the submitted request.

Select the check box next to **Category roles**, click **Add roles +** and select the checkbox next to **Owner**. Click **Add**.

The screenshot shows the 'Assign category roles' dialog box. It contains a search bar 'Find role...' and a list of roles: Admin, Owner (which is checked), Editor, and Reviewer. A red box highlights the 'Owner' checkbox. At the bottom right of the dialog is a button labeled 'Add roles +' (also highlighted with a red box). The background shows the 'Deliver Step' configuration from the previous screenshot.

By doing this, you have specified that the category owner is the individual who should take action and indicate when they've completed the action whenever a request is submitted.

11. Click **Refine** in the left panel to go to the Refine step.

Click on the checkbox next to **Workflow requester** to select the Workflow requester as the assignee since this workflow template was defined such that this step applies only when the approver returns the request back to the workflow requester to refine their request.

Click the **Activate** button.

Click **Save and activate** in the pop-up window to confirm that you want to save the changes and activate the workflow configuration.

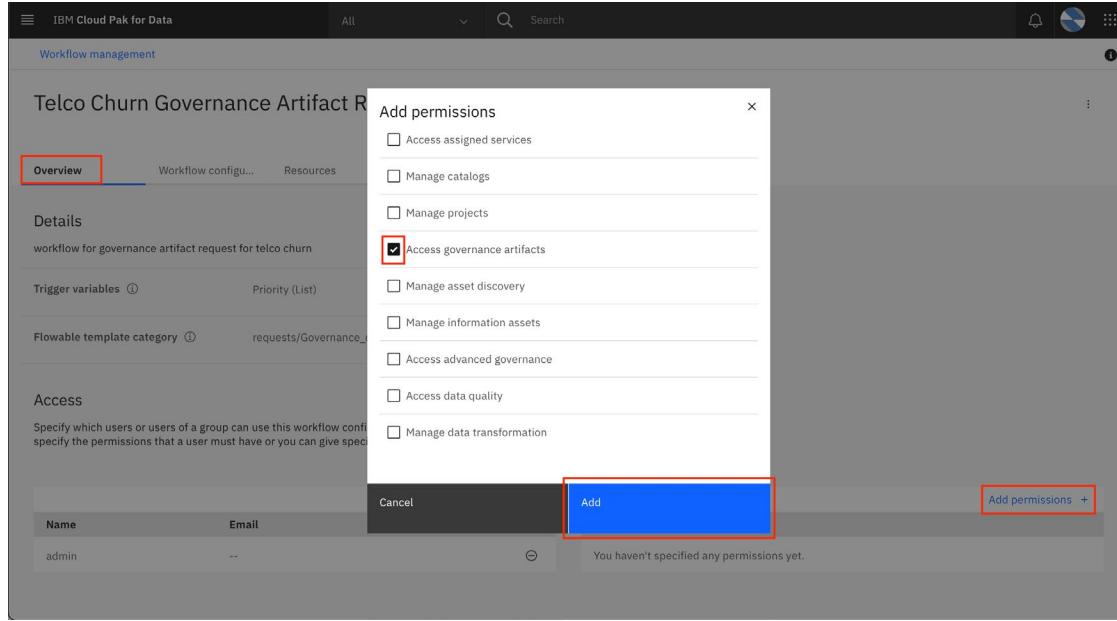
The screenshot shows the 'Telco Churn Artifact Request Default Configuration' page. The left sidebar lists steps: Triggers, Approve Step, Deliver Step, and Refine End. The 'Refine End' step is selected. The main content area has a 'Refine' heading with the sub-instruction 'Requester needs to reiterate on the request'. Under the 'Assignees' section, there is a checkbox for 'Workflow requester' which is checked. There are also options for 'Category roles' and 'Users and user groups', both of which are unchecked. Below this is a 'Notifications' section with a note about specifying users for notifications. At the top right of the page are 'Cancel', 'Save', and 'Activate' buttons, with 'Activate' being highlighted by a red box.

12. Click **Cancel** at the top right of the screen.

This screenshot shows the same configuration page as the previous one, but the 'Default' tab is selected in the top navigation bar instead of 'Active'. The rest of the interface is identical to the previous screenshot, including the checked 'Workflow requester' checkbox in the 'Assignees' section of the 'Refine' step's configuration. The 'Cancel' button is highlighted with a red box.

13. Navigate back to the **Overview** tab and click **Add permissions +**.

On the Add permissions pop-up window, scroll down to find, and select the **Access governance artifacts** check box (or use the search bar to find the role quickly). Click **Add**.



This effectively grants permission to access the request form to any user who has the platform permission of Access governance artifacts. You could have chosen to provide access to a specific list of users or user groups as well.

Configure governance artifact workflows

You can change the workflow for all governance artifacts associated with the categories you have defined earlier for which you are the Owner (Telco Churn, Personal Information, Customer Data, and Data Privacy). The following steps highlight the required steps for the Telco Churn category and the same steps can be followed for the other categories too.

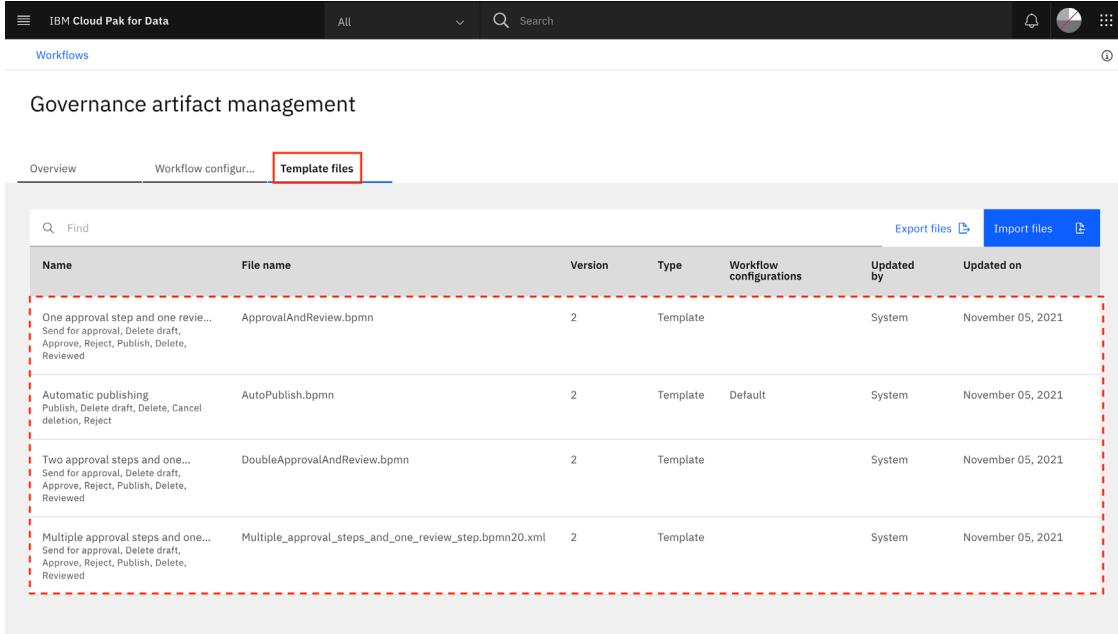
1. Ensure you are logged into Cloud Pak for Data as the admin user.
2. Go to the **hamburger (≡)** menu in the upper-left corner, expand **Administration** and click **Workflows**.

The screenshot shows the left navigation sidebar of the IBM Cloud Pak for Data interface. A red box highlights the 'Administration' section, which is expanded to show 'Workflows'. The main content area displays two tabs: 'Data Privacy' and 'Healthcare_Data'. At the top right of the content area, there is a blue button labeled 'Add category'.

3. Select the **Governance artifact management** tile which is the default workflow type supported by Watson Knowledge Catalog and is available out of the box.

The screenshot shows the 'Workflows' page. A red box highlights the 'Governance artifact manag...' tile. The tile contains the text: 'You can use workflows to manage business processes. Select the process that you want to manage.' Below this, it lists 'Governance artifact management' and 'Workflows for managing governance artifacts, such as business terms and data classes.' To the right of the tile, another card is visible: 'Telco Churn Governance Ar...', 'workflow for governance artifact request for telco churn'. At the top right of the page, there is a blue button labeled 'New workflow type'.

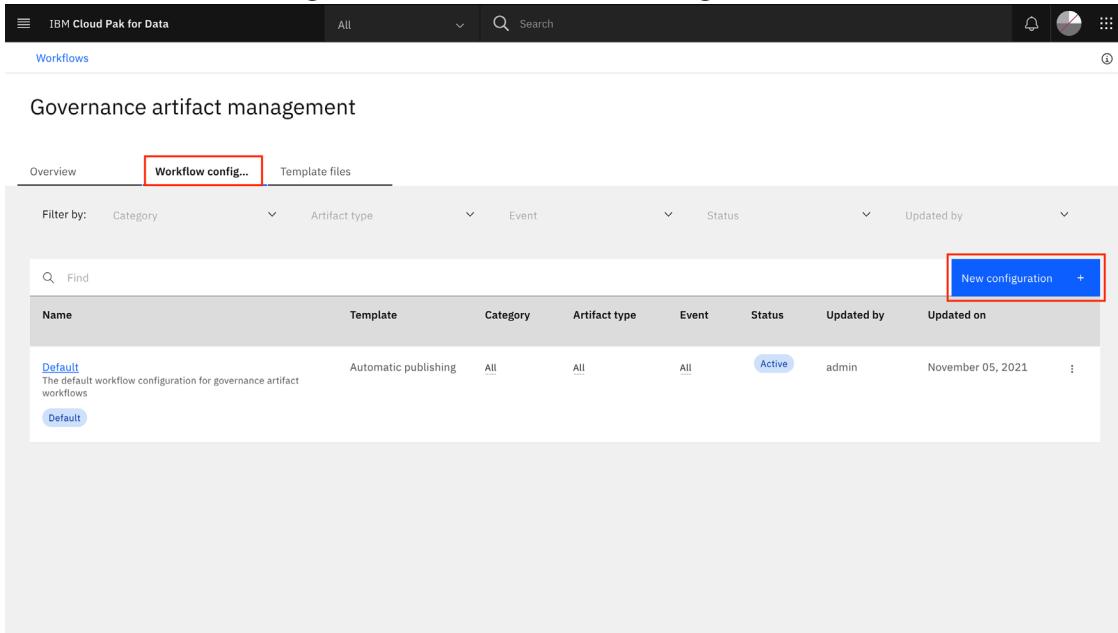
4. Select the **Template files** tab and review the four available workflow templates and associated approval process for each. You can upload additional custom workflow template configurations by importing such templates, if desired. For now, you will use the “Automatic publishing” template which does not go through specific workflow tasks before getting published.



The screenshot shows the 'Governance artifact management' section of the IBM Cloud Pak for Data interface. The 'Template files' tab is selected, highlighted with a red box. The table below lists four workflow templates:

Name	File name	Version	Type	Workflow configurations	Updated by	Updated on
One approval step and one review... Send for approval, Delete draft, Approve, Reject, Publish, Delete, Reviewed	ApprovalAndReview.bpmn	2	Template		System	November 05, 2021
Automatic publishing Publish, Delete draft, Delete, Cancel deletion, Reject	AutoPublish.bpmn	2	Template	Default	System	November 05, 2021
Two approval steps and one... Send for approval, Delete draft, Approve, Reject, Publish, Delete, Reviewed	DoubleApprovalAndReview.bpmn	2	Template		System	November 05, 2021
Multiple approval steps and one... Send for approval, Delete draft, Approve, Reject, Publish, Delete, Reviewed	Multiple_approval_steps_and_one_review_step.bpmn20.xml	2	Template		System	November 05, 2021

5. Select the **Workflow configurations** tab and click **New Configuration +**.

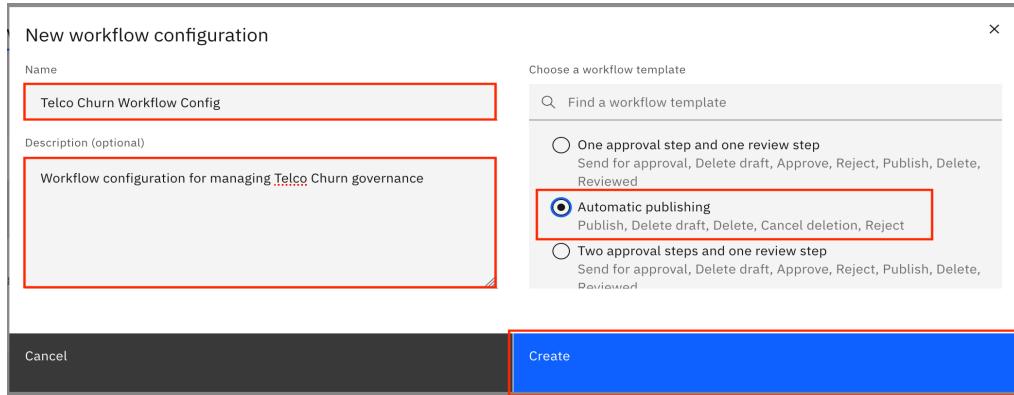


The screenshot shows the 'Governance artifact management' section with the 'Workflow config...' tab selected, highlighted with a red box. A 'New configuration +' button is also highlighted with a red box. The table below lists the current workflow configuration:

Name	Template	Category	Artifact type	Event	Status	Updated by	Updated on
Default The default workflow configuration for governance artifact workflows	Automatic publishing	All	All	All	Active	admin	November 05, 2021

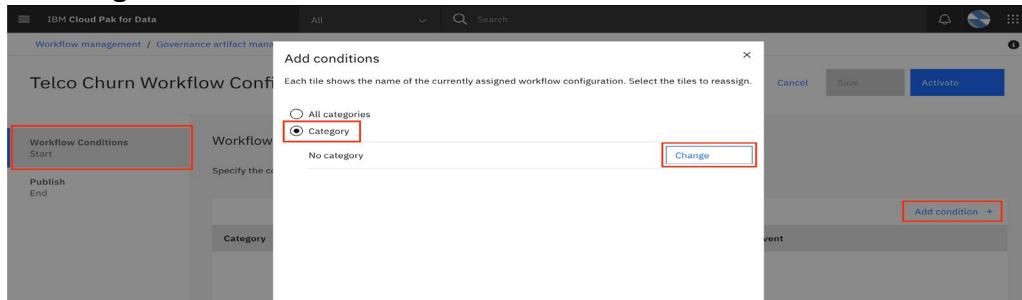
- Provide a Name (**Telco Churn Workflow Config**) and an optional Description (**Workflow configuration for managing Telco Churn governance**) for the new workflow configuration.

Select **Automatic publishing** and click **Create**.

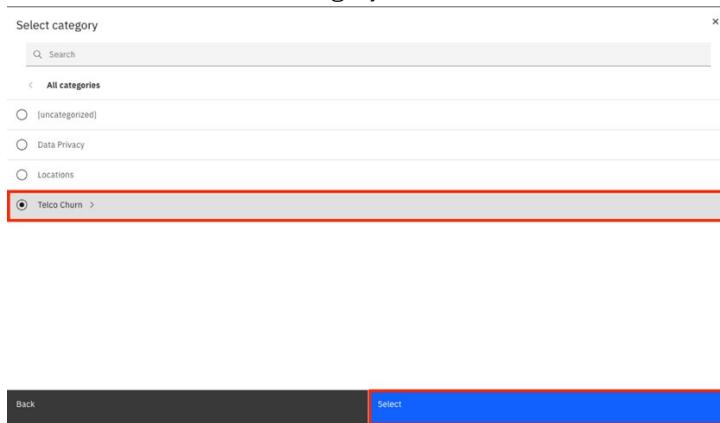


- On the **Workflow Conditions** step, click **Add condition +**.

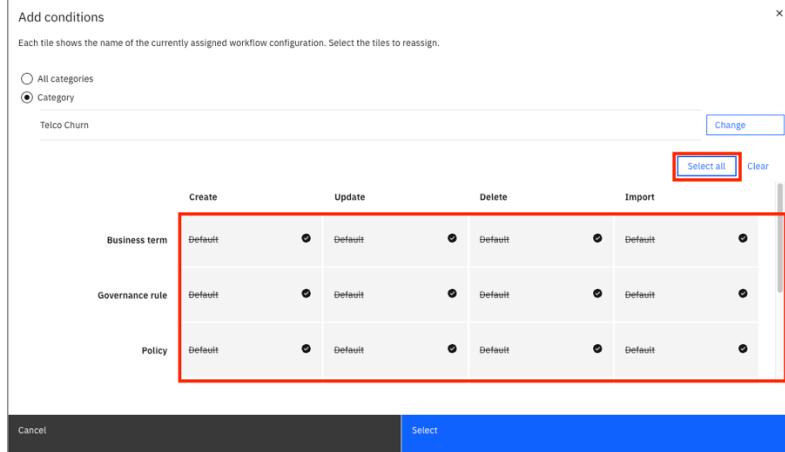
In the pop-up window, click the radio button next to **Category** to select a specific category, then click the **Change** button.



- Select the **Telco Churn** category and click **Select**.



- You will be brought back to the previous screen. Click **Select all** to replace the Default configuration. This will ensure that this workflow configuration will apply for any kind of changes made to all governance artifacts associated with the Telco Churn category.



- Finally click the **Select** button to return to the Workflow Conditions. You will see all 6 artifact types along with their events.

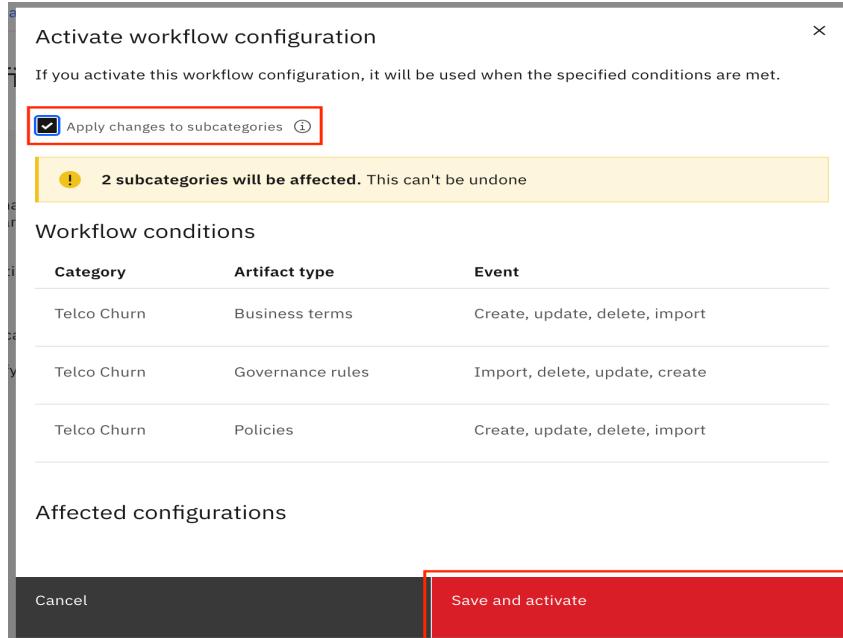
Category	Artifact type	Event
Telco Churn	Business terms	Create, Update, Delete, Import
Telco Churn	Governance rules	Create, Update, Delete, Import
Telco Churn	Policies	Create, Update, Delete, Import
Telco Churn	Classifications	Create, Update, Delete, Import
Telco Churn	Data classes	Create, Update, Delete, Import
Telco Churn	Reference data	Create, Update, Delete, Import

- On the **Publish** step, make sure the check box for **Automatically publish artifacts. Skip publishing approval** is selected. Click **Activate**.

Automatically publish artifacts. Skip publishing approval.

12. Click the checkbox for **Apply changes to subcategories**.

Click **Save and activate** in the pop-up window to confirm that you want to activate this workflow configuration for the specified conditions.



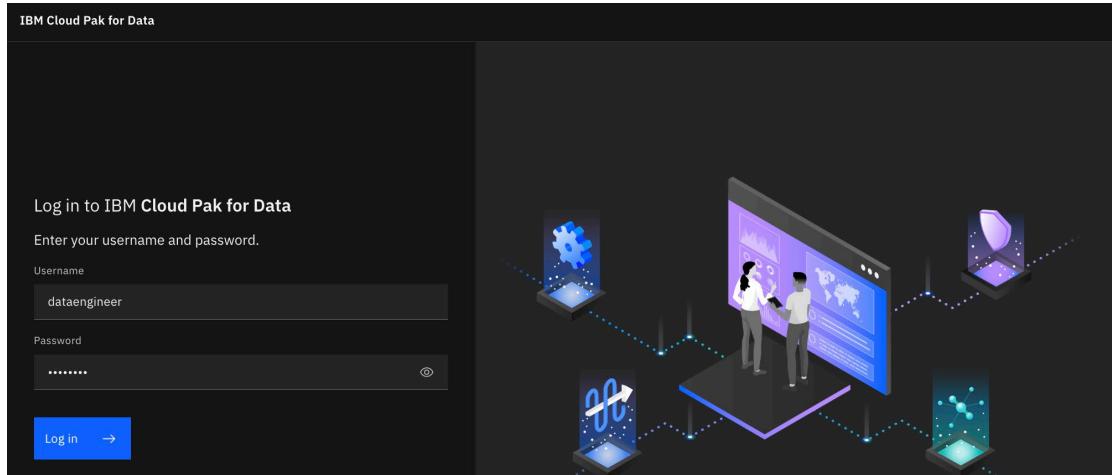
13. The processes to govern how artifacts get added to the Telco Churn category as well as its subcategories - “Personal Info” and “Customer Data” - have now been added. If you wish to, you can repeat these steps to create similar workflows for the “Data Privacy” category.
14. Once the workflow is created, the dataengineer can submit a request for adding other governance artifacts to the categories.

Business terms

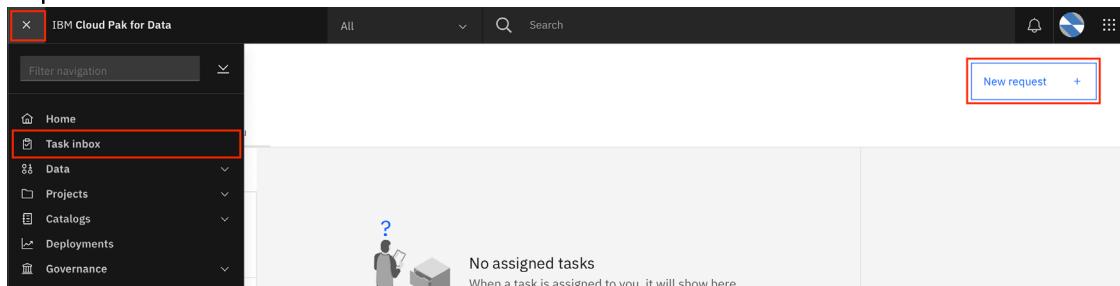
Business terms are used to create an ontology around the business. They are used to characterize other artifacts as well as assets and columns in the catalog. Using business terms for business concepts ensures that the enterprise data is described in a uniform manner and is easily understood across the entire enterprise. Business terms can be used to describe the contents of the data, the sensitivity of the data, or other aspects of the data, such as the subject or purpose of the data. You can assign one or more business terms to individual columns in relational data sets, to other governance artifacts, or to data assets.

Request to create Business terms

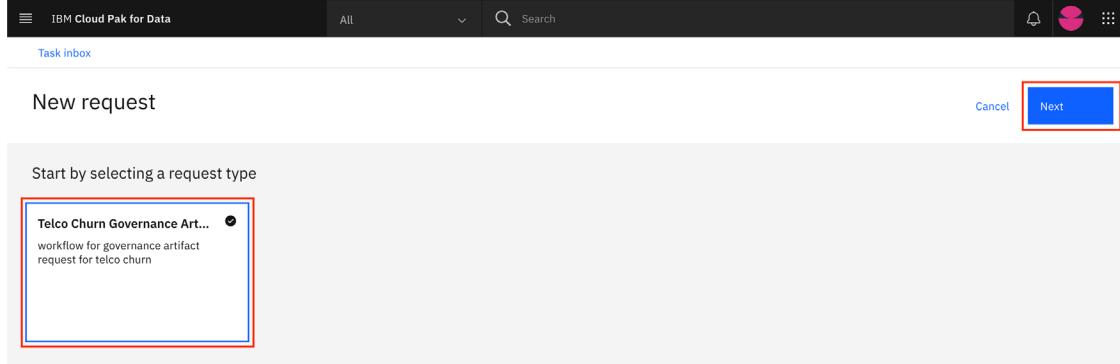
1. Log into Cloud Pak for Data as the dataengineer user.



2. Navigate to the **hamburger (≡)** menu in the upper-left corner and click on **Task inbox**. Click on **New request +**.



- Select the request type **Telco Churn Governance Artifact Request Process** and click **Next**.



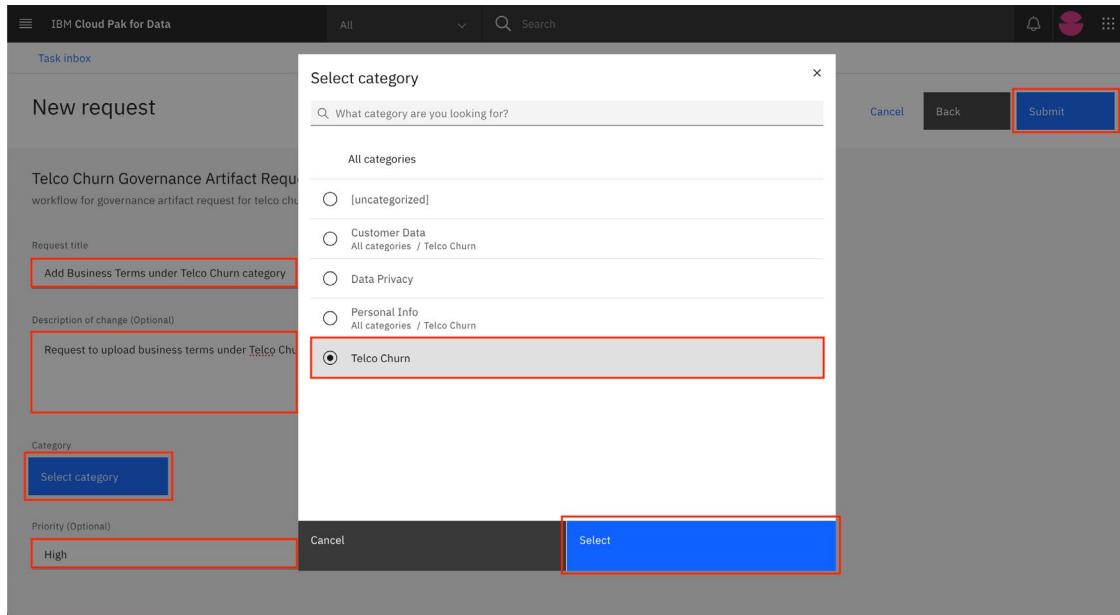
- Provide a Request **title** (Add Business Terms under Telco Churn category) and an optional Description of change (Request to upload business terms under Telco Churn category) for the request.

Click the **Select category** button.

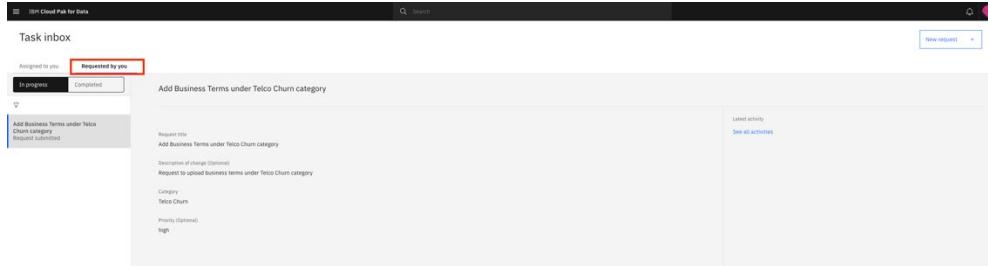
In the pop-up window, select the **Telco Churn** category and click **Select**.

Back on the “New Request” page, optionally select a Priority Level of **High**.

then click **Submit**.



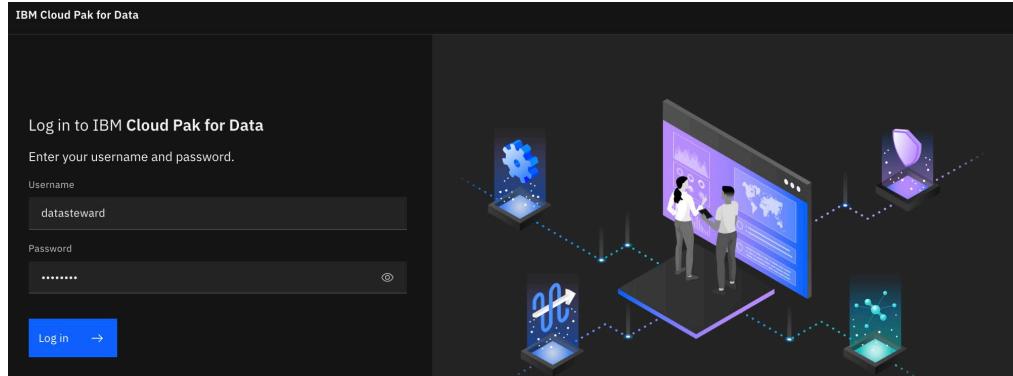
5. This request will be sent to the “Telco Churn” category owner to fulfill this request. You can view your requests by clicking on the **Requested by you** tab in the **Task inbox**.



Approve the Business term creation request

1. Log out of Cloud Pak for Data and log back in as the datasteward user.

Remember that the datasteward user was assigned as owner of the “Telco Churn” category and the workflow was configured for requests to be assigned to the category owner.



2. On your landing page, note the notification that you have been assigned a task to “Approve GCR Add Business Terms under Telco Churn category” (annotated with red rectangle).

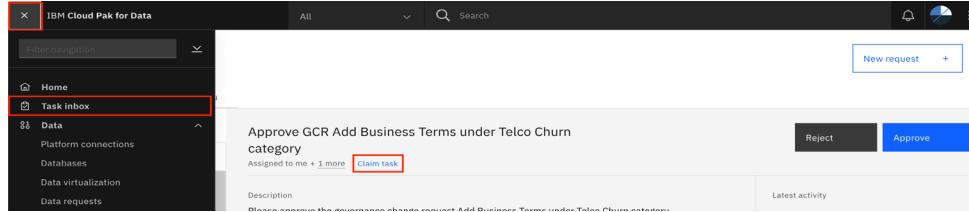
Additionally, you can view a summary of tasks assigned to you or requested by you on your landing page (annotated with red oval).

Also note that you can always customize the look and feel of your landing page when you log into Cloud Pak for Data.

Category	Value
Assigned to me	1
Requested by me	0

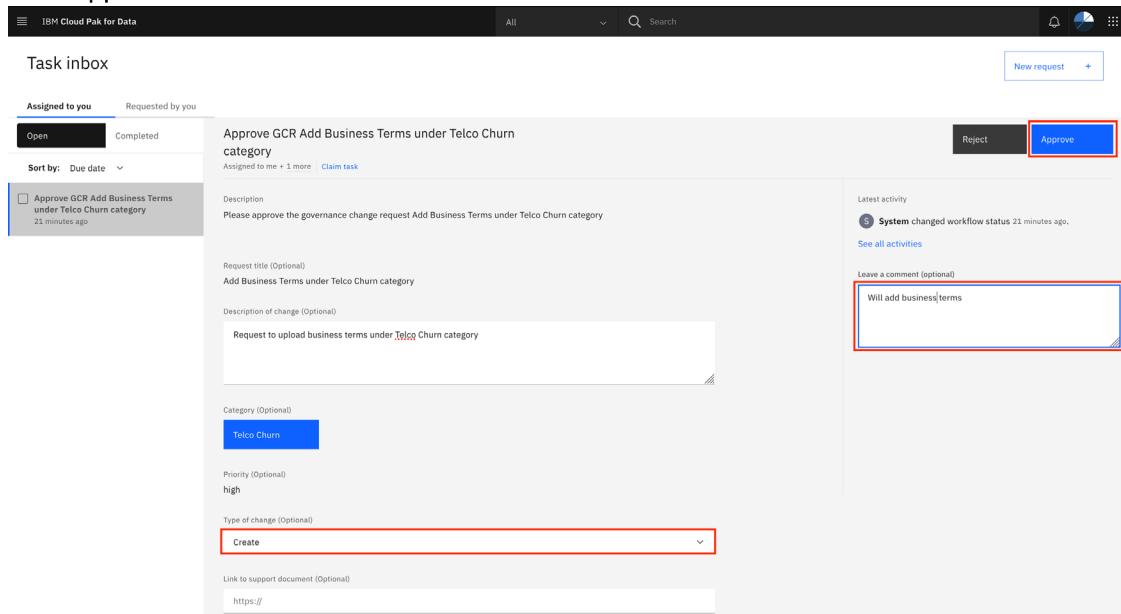
- Click on the **link** to access your new task from the landing page. Alternatively, you can navigate to the **hamburger (≡)** menu in the upper-left corner and click on **Task inbox** where you would see that the “Approve GCR Add Business Terms under Telco Churn category” task has been assigned to you.

Click on **Claim task** to show any other assignees that you will be completing the task.

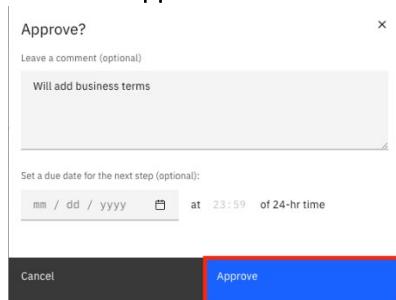


- Scroll down and add optional details such as the Type of change (**Create**), Link to support document option, and leave a comment such as “**Will add business terms**”.

Click **Approve**.



- Also click **Approve** on the confirmation window that appears.



Create business terms by importing a file

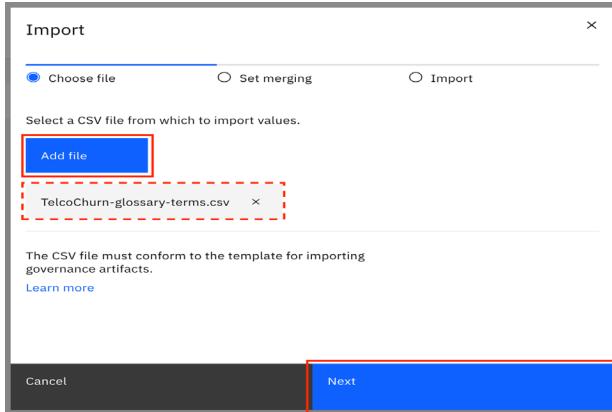
You will now upload business terms under the Telco Churn category to fulfill the request that you approved.

1. Download the [TelcoChurn-glossary-terms.csv](#) file.
2. Navigate to the hamburger () menu in the upper-left corner, expand **Governance** and click on **Business terms**.

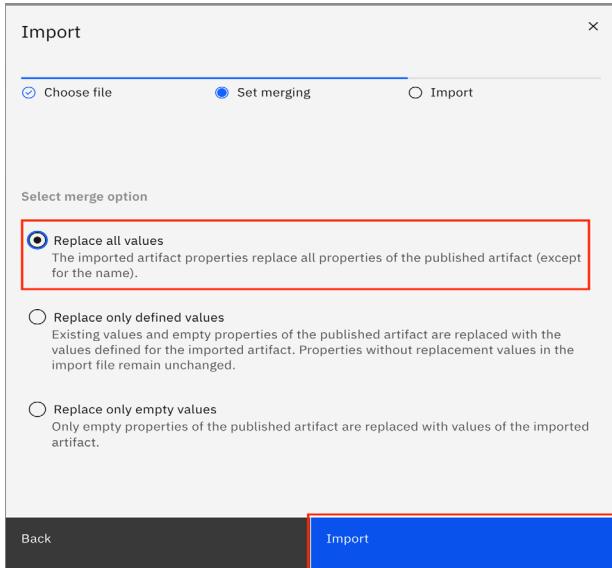
3. Expand the **Add business term** drop down and click on **Import from file**.

Note that you can also add business terms manually term by term, if you choose to.

- In the pop-up window, click **Add file** to browse and select the “TelcoChurn-glossary-terms.csv” file which you had downloaded earlier and click **Next**.

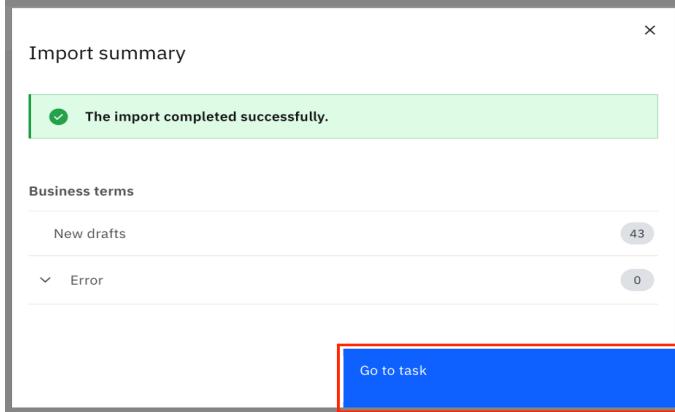


- Select the radio button to **Replace all values** and click **Import**.



- Once import is complete, you get a message that “The import completed successfully”.

Click the **Go to task** button to review and approve the imported business terms.



Note: All governance artifacts are first saved as draft to allow for edits and updates as needed before publishing.

- On the **Assigned to you** tab, you will see a new task to “Publish Business terms”.

Click **Claim task** to let other assignees know that you will handle this task.

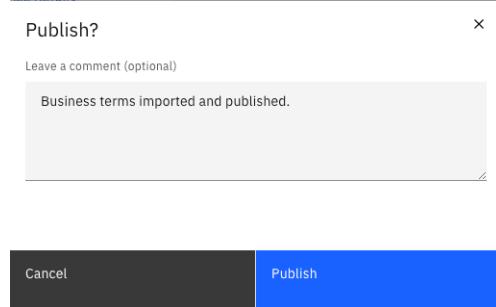
The “Description” field explains that the task is to review each business term and then publish or delete all the business terms together. It also explains that the datasteward imported these business terms.

Review the imported business terms and click **Publish**.

The screenshot shows the "Task inbox" page in IBM Cloud Pak for Data. The top navigation bar includes "IBM Cloud Pak for Data", "All", "Search", and a user profile. The main area shows the "Assigned to you" tab selected. A task titled "Publish Business terms" is listed, with a red box highlighting its title and the "Claim task" button. The task description states: "datasteward imported these Business terms. Review each Business term and then publish or delete all of them together." To the right, there is a "Delete draft" button and a large blue "Publish" button, also highlighted with a red box. On the far right, there is a sidebar with "Latest activity" showing a system status update, and a comment input field with "Leave a comment (optional)" and "Write a comment...".

It is also a good practice to add a comment.

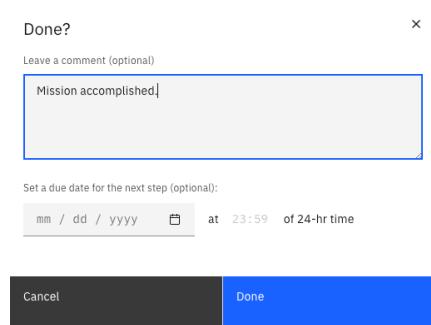
Enter a comment, for example “**Business terms imported and published.**” and click **Publish**.



8. You will now return to the Task inbox, task “Deliver GCR Add Business Terms under Telco Churn category” task. Click **Done** to mark the task as Completed.

Assigned to you	Requested by you	
Open	Completed	
Sort by: Due date ▾		
<input type="checkbox"/> Deliver GCR Add Business Terms under Telco Churn category <small>1 hour ago</small>		
Deliver GCR Add Business Terms under Telco Churn category <small>Assigned to me + 1 more Claim task</small>		
Description <small>Please approve the governance change request Add Business Terms under Telco Churn category</small>		
<small>Request title (Optional)</small> <small>Add Business Terms under Telco Churn category</small>		
<small>Latest activity</small> datasteward <small>Changed workflow status 1 hour ago</small> <small>Approve</small> <small>See all activities</small>		

Optionally, enter a description on the confirmation window and click **Done**.

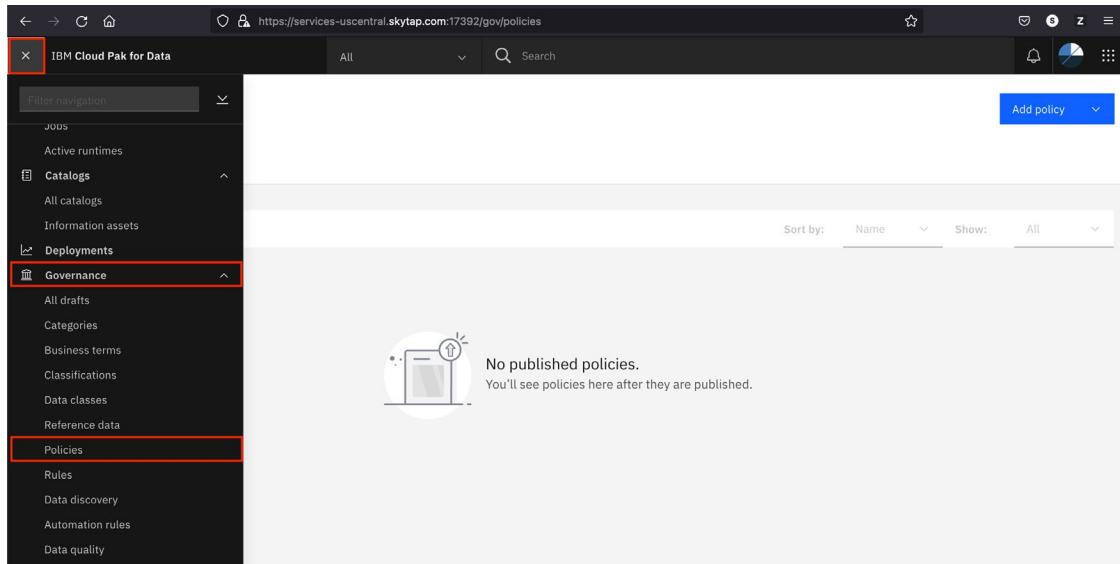


Governance Policies

Policies capture important initiatives driven by the organization. You will create a policy that requires all analytics teams to follow the same standards for US State and County codes.

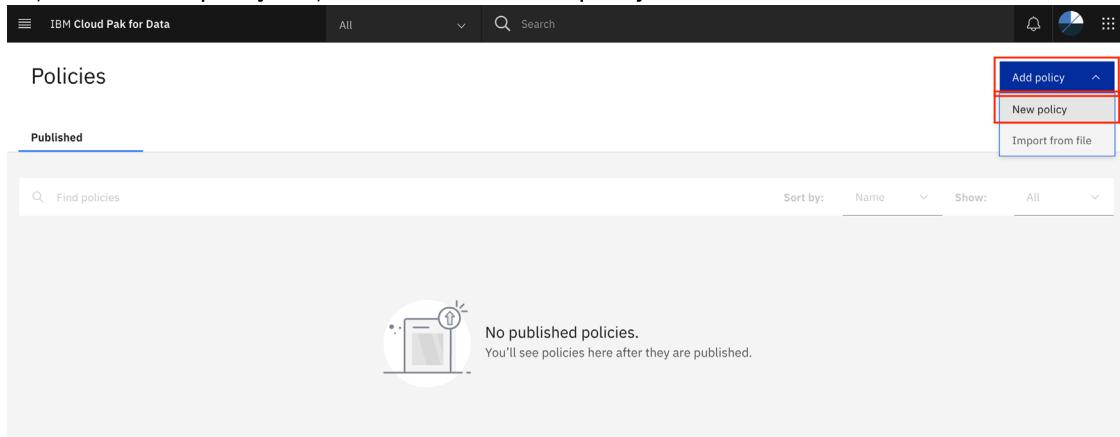
Create a policy manually

1. Navigate to the hamburger (\equiv) menu in the upper-left corner, expand **Governance** and click on **Policies**.



The screenshot shows the IBM Cloud Pak for Data interface. The left sidebar has a 'Governance' section expanded, with 'Policies' highlighted by a red box. The main area displays a message: 'No published policies. You'll see policies here after they are published.' with an upward arrow icon.

2. Expand the **Add policy** drop down and click **New policy**.



The screenshot shows the 'Policies' page with the 'Published' tab selected. The 'Add policy' dropdown menu is open, showing 'New policy' highlighted with a red box. The main area displays a message: 'No published policies. You'll see policies here after they are published.' with an upward arrow icon.

- Provide a Policy name (Use of US State and County codes) and an optional Description (Follow the standards outlined in this policy for the Telco Churn category and its subcategories).

Click on the **Select** button against Primary category and select the **Telco Churn** category to apply this policy to.

Click **Save as draft**.

New policy

Create a policy in a category where you have the rights.

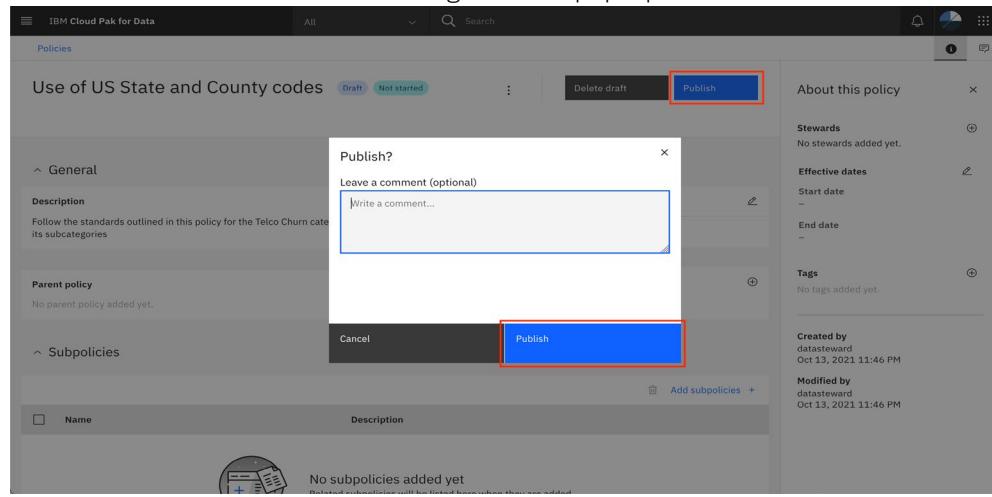
Name
Use of US State and County codes

Primary category ⓘ
Select a category **Select**

Description (optional)
Follow the standards outlined in this policy for the **Telco Churn** category and its subcategories

Cancel Save as draft

- Click **Publish** and then click **Publish** again in the pop-up window.



5. You can review activities associated with governance artifacts by clicking the activities icon to view the details of the activities.

The screenshot shows the IBM Cloud Pak for Data Policies interface. On the left, there's a main panel for a policy named "Use of US State and County codes" (status: Published). The panel includes sections for General (Description: Follow the standards outlined in this policy for the Telco Churn category and its subcategories; Primary category: Telco Churn), Parent policy (No parent policy added yet), and Subpolicies (No subpolicies added yet). On the right, a modal window titled "Activities" is open, containing a list of audit logs:

- datasteward created the policy Use of US State and County codes. 1 minute ago. See details
- datasteward updated the policy Use of US State and County codes. 1 minute ago. See modifications
- datasteward performed action Publish on policy Use of US State and County codes. 52 seconds ago. Published. See modifications

Below the log list is a comment input field with placeholder text "Write a comment..." and a "Comment" button.

Create a policy by importing a file

1. Download the [TelcoChurn-policies.csv](#) file.
2. Navigate back to Policies by clicking on Policies in the breadcrumbs.

The screenshot shows the IBM Cloud Pak for Data interface. In the top navigation bar, 'IBM Cloud Pak for Data' is selected. Below it, 'Policies' is highlighted in red. The main content area displays a policy titled 'Use of US State and County codes' which is 'Published'. The policy details include a description: 'Follow the standards outlined in this policy for the Telco Churn category and its subcategories', a primary category 'Telco Churn', and secondary categories. The 'Activities' sidebar on the right shows three recent actions: 'datasteward created the policy Use of US State and County codes.', 'datasteward updated the policy Use of US State and County codes.', and 'datasteward performed action Publish on policy Use of US State and County codes.'.

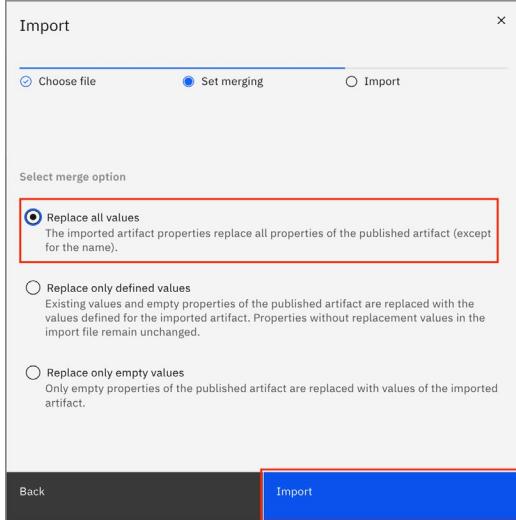
3. Expand the Add policy dropdown and click Import from file.

The screenshot shows the 'Policies' page with the 'Published' tab selected. The 'Add policy' dropdown menu is open, showing 'Add policy', 'New policy', and 'Import from file', with 'Import from file' highlighted in red. The main content area shows one published policy: 'Use of US State and County codes'.

4. Click Add file and browse to select the "TelcoChurn-policies.csv" file. Click Next.

The screenshot shows the 'Import' dialog box. It has three options: 'Choose file' (selected), 'Set merging', and 'Import'. Below this, it says 'Select a CSV file from which to import values.' and shows a file named 'TelcoChurn-policies.csv' with a red dashed border around it. A note below states: 'The CSV file must conform to the template for importing governance artifacts.' and a 'Learn more' link. At the bottom are 'Cancel' and 'Next' buttons, with 'Next' highlighted in red.

5. Select the radio button to Replace all values and click Import.



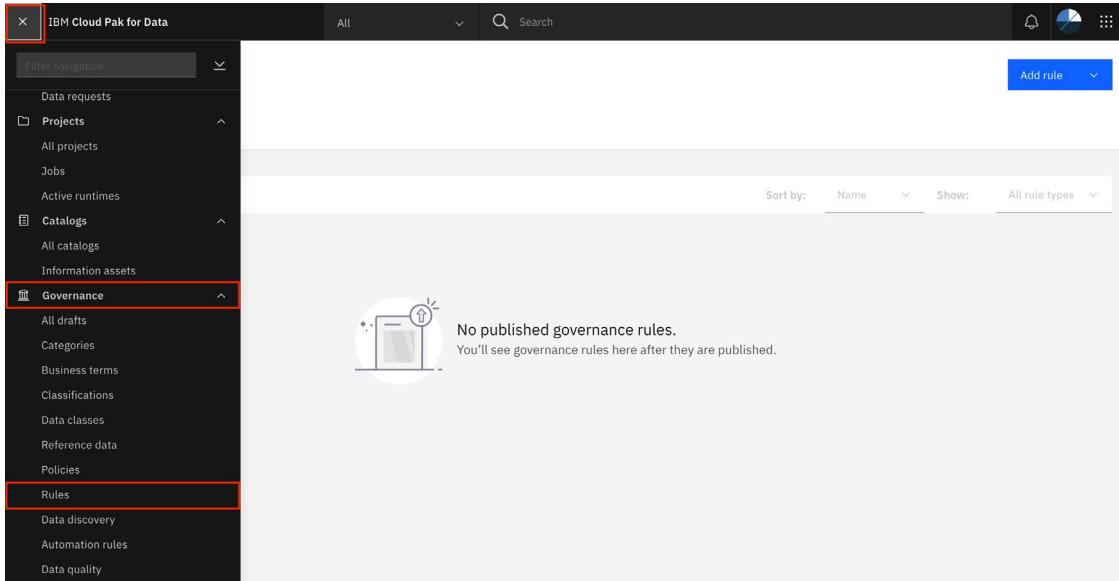
6. Once import is complete, you get a message that the import has completed successfully.
Click on **Go to task** to review and publish the policies in a manner similar to what was done earlier with the business terms.

Governance rules

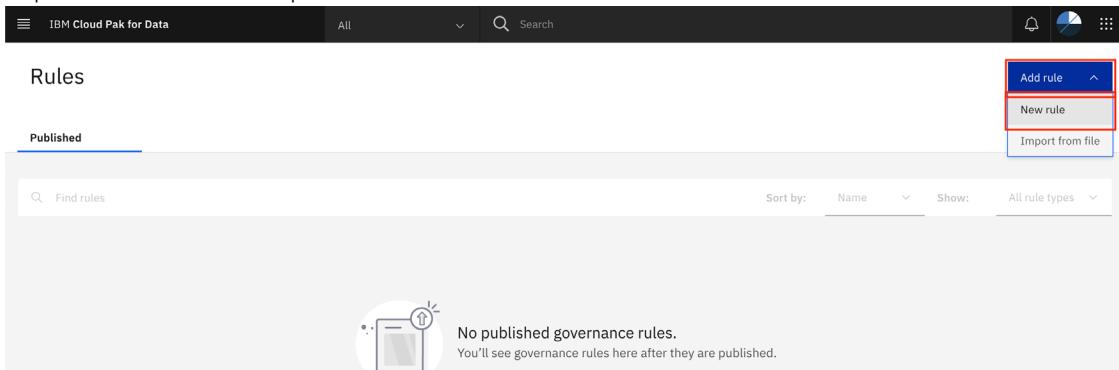
Watson Knowledge Catalog supports two types of rules that can be added to policies: governance rules and data protection rules. Data protection rules automatically mask or restrict access to data assets. Governance rules explain specific requirements to follow a policy.

Create a governance rule manually

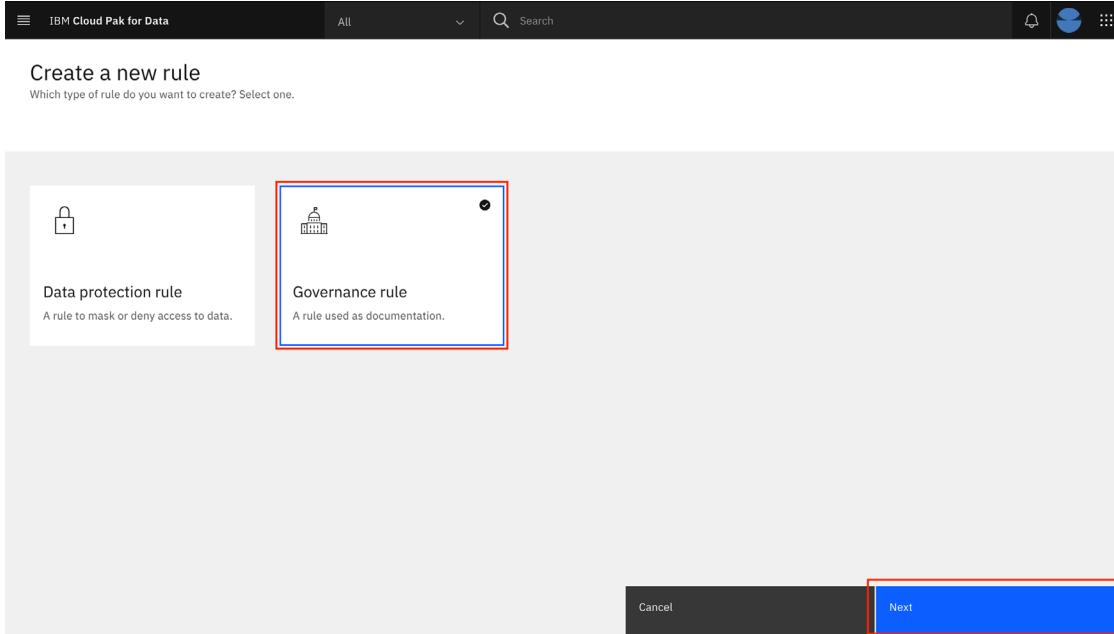
1. Navigate to the hamburger (\equiv) menu in the upper-left corner, expand **Governance** and click on **Rules**.



2. Expand the **Add rule** dropdown and select **New rule**.



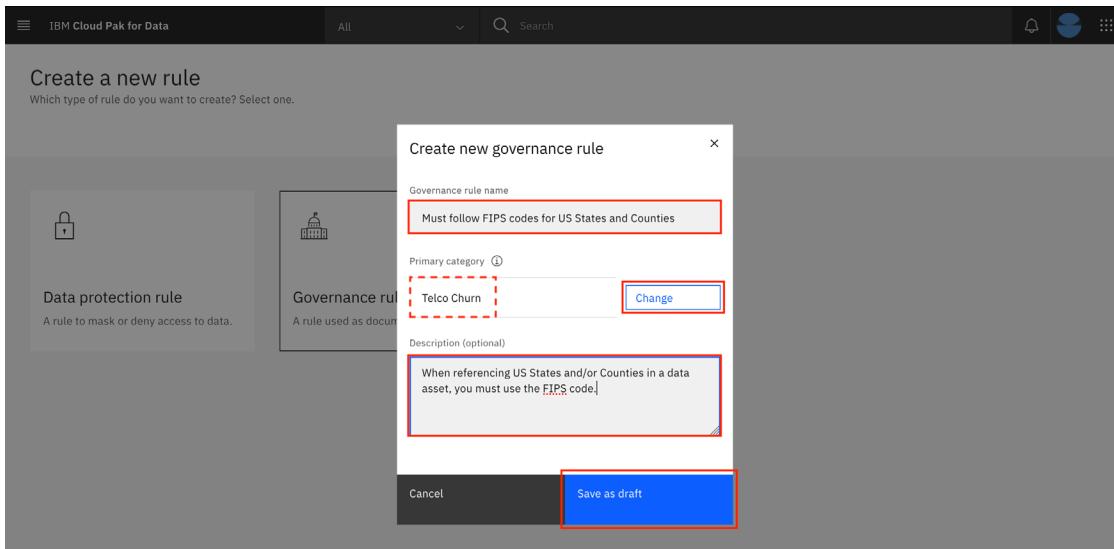
- Click on the tile for **Governance rule** and click **Next**.



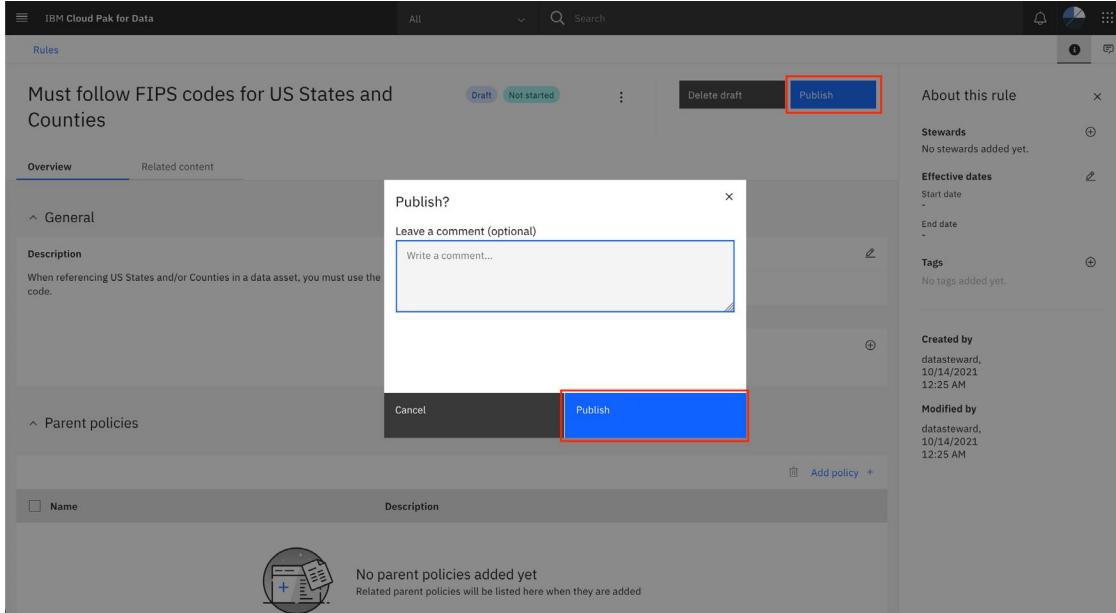
- Provide rule details such as the Governance rule **name** (Must follow FIPS codes for US States and Counties) and an optional Description (When referencing US States and/or Counties in a data asset, you must use the FIPS code.)

Select the Category to which the rule applies by clicking **Change** and selecting the **Telco Churn** category.

Click Save as draft.



5. Click **Publish** and then click **Publish** again in the pop-up window to publish the rule.



Associate a governance rule with a policy

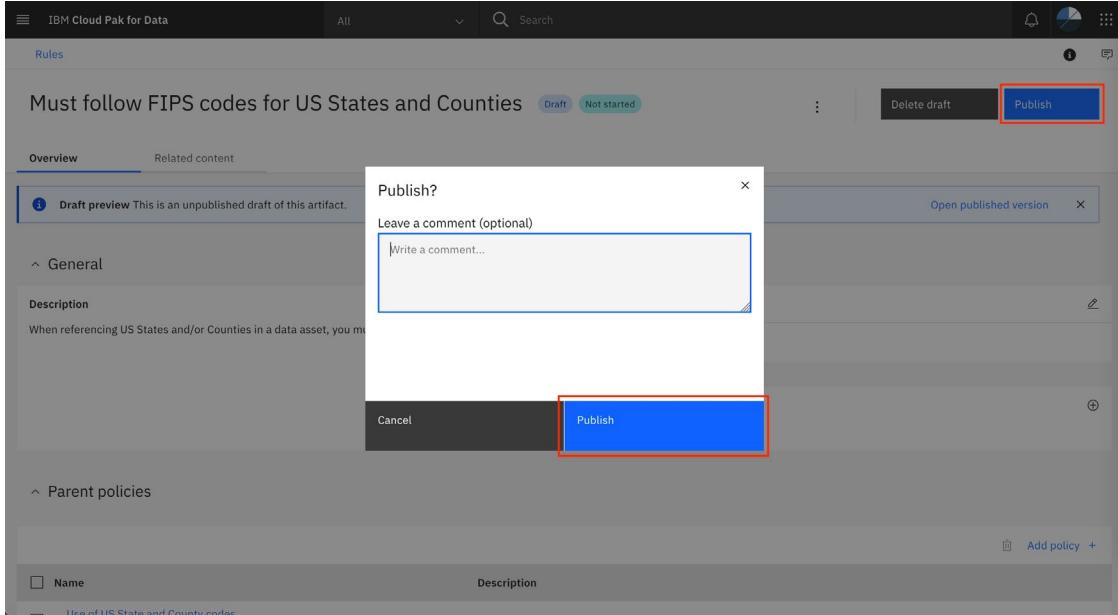
- Once the rule is published, you can associate it with the policy created earlier using one of two methods – either by adding the parent policy to the rule, or by adding the rule to the policy.
- For the first method, in the rule page, click on **Add policy +** in the “Parent Policies” section.

The screenshot shows the 'Rules' section of the IBM Cloud Pak for Data interface. A specific rule titled 'Must follow FIPS codes for US States and Counties' is displayed. In the 'Parent policies' section, there is a red box around the 'Add policy +' button. The 'About this rule' and 'Effective dates' sections are also visible on the right side of the page.

In the pop-up, select the **Use of US State and County codes** policy and click **Add**.

The screenshot shows a 'Add policies' pop-up window. On the left, there's a list of policies: 'Data Privacy' (unchecked) and 'Use of US State and County codes' (checked). On the right, under 'Selected policies', 'Use of US State and County codes' is listed with a sub-note 'All categories / Telco Churn'. At the bottom right of the pop-up is a large blue 'Add' button.

- Publish the rule again to publish this association between the policy and the rule.



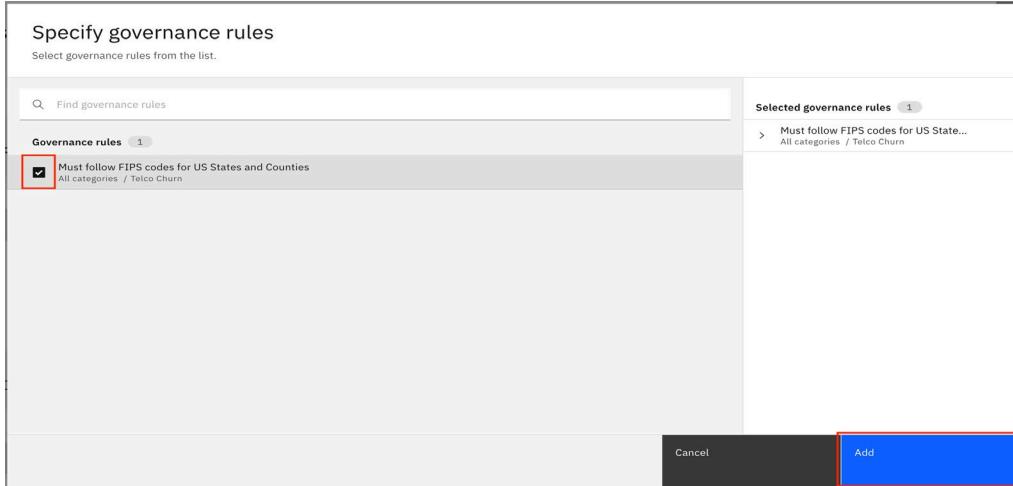
- Alternatively, you can update the policy and add the rule to it. Navigate back to the governance policies by clicking on the hamburger (☰) menu in the upper-left corner, expanding **Governance** and clicking on **Policies**. On the Policies page, click on the **Use of US State and County codes** policy.

Name	Description	Last modified
Data Privacy	Company-wide data privacy policy for securing private data	Oct 14, 2021
Use of US State and County codes	Follow the standards outlined in this policy for the Telco Churn category and its subcategories	Oct 13, 2021

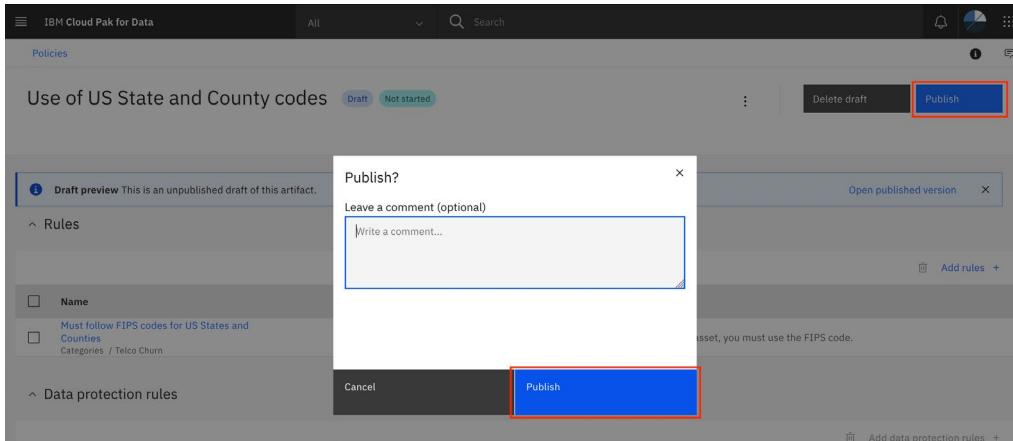
- Scroll down to the “Rules” section and click **Add rules +**.

Name	Description
No rules added yet	Related rules will be listed here when they are added

- Note that if you've already assigned the rule to the policy, you will not see any rules to add here. If you have not, then select the check box next to **Must follow FIPS codes for US States and Counties** governance rule. Click **Add**.



- Once the policy is updated, you need to publish the updated policy.



Create governance rules by importing a file

Next, you will define additional governance rules by importing a file.

1. Download the [TelcoChurn-rules.csv](#) file.
2. Navigate back to governance rules by going to the **hamburger (≡)** menu in the upper-left corner, expanding **Governance** and clicking on **Rules**.

The screenshot shows the IBM Cloud Pak for Data interface. On the left, there is a navigation sidebar with sections like Data requests, Projects, Catalogs, Governance, and Rules. The 'Rules' section under Governance is highlighted with a red box. The main content area displays a message: 'No published governance rules. You'll see governance rules here after they are published.' There is also a small icon of a document with a gear.

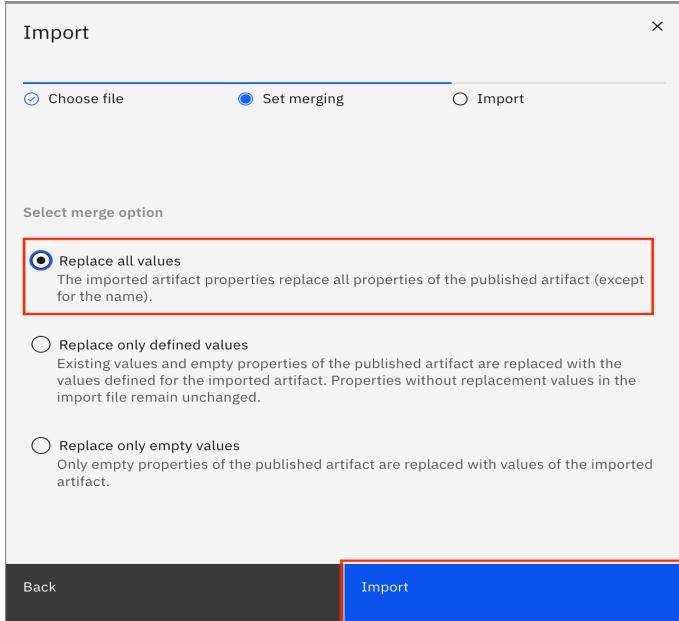
3. Expand the **Add rule** dropdown and select **Import from file**.

The screenshot shows the governance rules list page. The 'Published' tab is selected. In the top right, there is an 'Add rule' dropdown with options 'New rule' and 'Import from file', where 'Import from file' is highlighted with a red box. Below the dropdown, there is a table listing a single rule: 'Must follow FIPS codes for US States and Counties'. The table includes columns for 'Name', 'Last modified', and 'Edit'.

4. Click **Add file** and browse to select the “TelcoChurn-rules.csv” file that was downloaded earlier. Click **Next**.

The screenshot shows the 'Import' dialog box. It has three radio button options: 'Choose file' (selected), 'Set merging', and 'Import'. Below this is a section to 'Select a CSV file from which to import values.' with a 'Add file' button highlighted with a red box. A dashed red box surrounds the file 'TelcoChurn-rules.csv' listed in the file selector. At the bottom, there is a note: 'The CSV file must conform to the template for importing governance artifacts.' followed by a 'Learn more' link. The 'Cancel' and 'Next' buttons are at the bottom, with 'Next' highlighted with a red box.

5. Select the radio button to **Replace all values** and click **Import**.



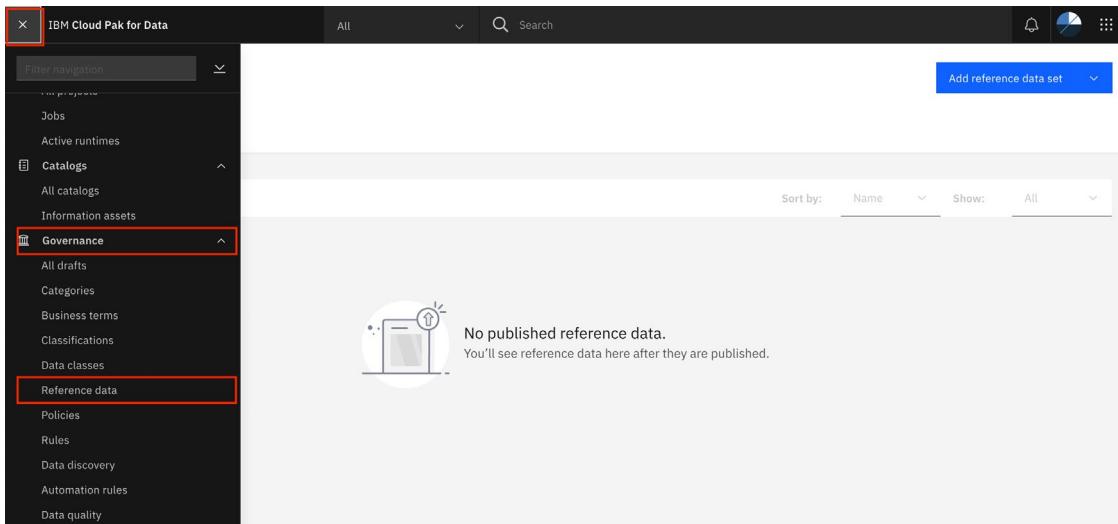
6. Once import is complete, you get a message that the import has completed successfully.
Click on **Go to task** to review and publish the rules in a manner similar to what was done earlier with the business terms.

Reference data sets and hierarchies

Reference Data Sets are common codes and values used across the organization. It is important to centrally manage reference data, so that the organization can use codes and values consistently in different systems and track how reference data sets used in different systems relate to each other.

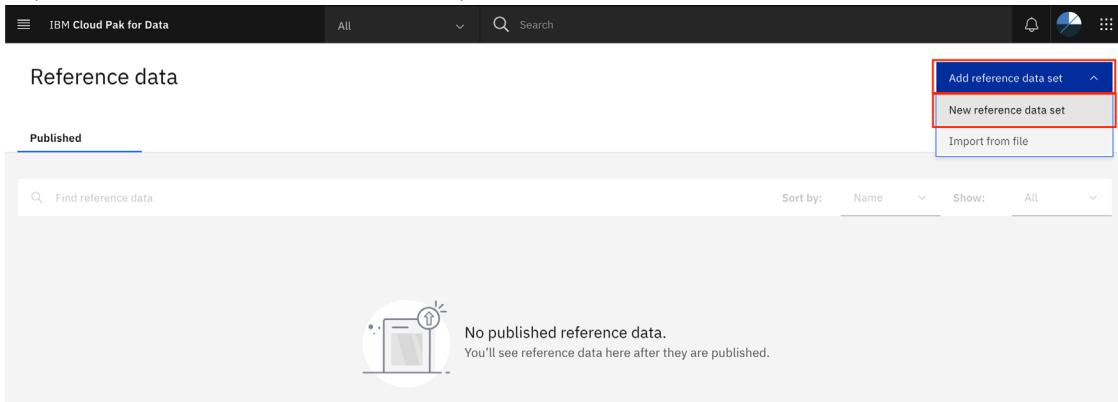
Create reference data sets

1. Download the [USStateCodes.csv](#).
2. Navigate to the hamburger () menu in the upper-left corner, expand **Governance** and click on **Reference data**.



The screenshot shows the IBM Cloud Pak for Data interface. The top navigation bar includes the logo, a search bar, and a 'Bell' icon. On the left, a sidebar menu is open under 'Governance'. The 'Reference data' option is highlighted with a red box. The main content area displays a message: 'No published reference data. You'll see reference data here after they are published.' with an icon of a document with arrows.

3. Expand the **Add reference data set** dropdown and select **New reference data set**.



The screenshot shows the 'Reference data' page. The top navigation bar includes the logo, a search bar, and a 'Bell' icon. The sidebar on the left has 'Published' selected. The main content area displays a message: 'No published reference data. You'll see reference data here after they are published.' with an icon of a document with arrows. A dropdown menu is open at the top right, showing 'Add reference data set' with a sub-item 'New reference data set' highlighted with a red box. Other options in the dropdown include 'Import from file'.

- Upload the “USStateCodes.csv” file which includes the codes for the US State Codes reference data set. Provide a Name (US State Codes) and an optional Description (FIPS codes for US States) for the reference data set.

Click **Select** to select the Telco Churn category.

Click **Next**.

New reference data set

Add file (Optional)
Upload a CSV file or add values manually later.
Drop a file here or browse for a file to upload

USStateCod...

Reference data name: US State Codes | Reference data set type: Text

Primary category: Select a category

Description (Optional): FIPS codes for US States

Cancel

- Ensure that the **First row as column header** option is set to “On”, so it picks up the titles of the columns in your file.

Set the **Name** column to be **Value** and the **FIPS Code** column to be **Code**.

Click **Next**.

New reference data set

Add file
Header option
First row as column header On

Columns from file	Target columns
Name	Value <input style="border: 2px solid red; width: 100px; height: 20px;" type="text"/>
Abbreviation	Select column <input style="width: 100px; height: 20px;" type="text"/>
FIPS Code	Code <input style="border: 2px solid red; width: 100px; height: 20px;" type="text"/>

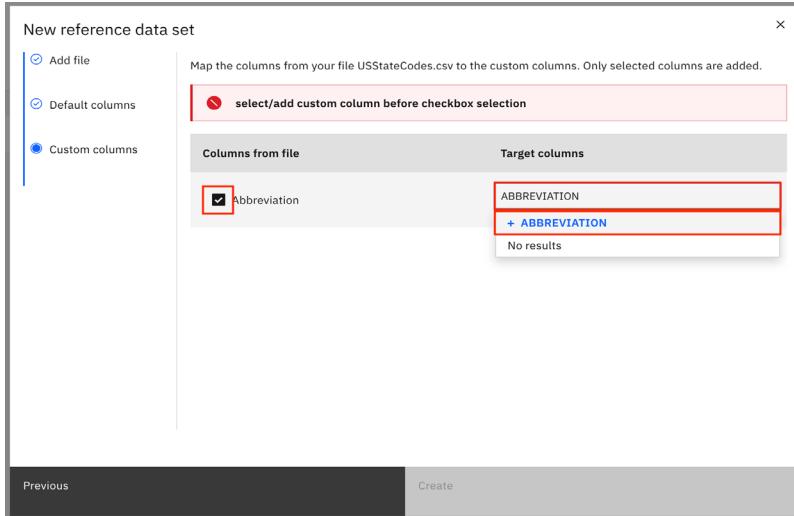
Previous

Note: Watson Knowledge Catalog offers the following default columns: Code, Value, Description, Parent (when creating a set-level hierarchy) and Related terms (when relating values to Business Terms).

- Since “Abbreviation” is not one of the default columns supported in Watson Knowledge Catalog, you will need to create a custom column.

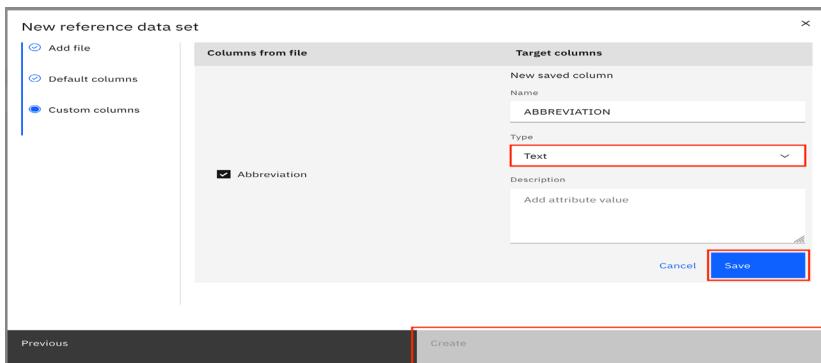
Select the check box next to **Abbreviation** then type in a name (**ABBREVIATION**) for the column. The name of the column is shown in the list with a plus sign.

Click to add the custom column.



- Complete the fields in the next prompt by specifying the **Type** as **Text** and providing an optional description, then click on **Save** to save that column as a re-usable column. This will help the organization be more consistent in how they name columns used in reference data sets.

Click **Create**.



8. You will be brought to the Reference Data Set that you just created. For each value, you will see the default columns plus the new column you created for Abbreviation.

Add **datasteward** as the steward and feel free to add churn as a tag.

The screenshot shows the 'Add stewards' dialog for the 'US State Codes' reference data set. The 'datasteward' option is selected and highlighted with a red box. The 'Add' button at the bottom right of the dialog is also highlighted with a red box.

9. Click **Publish** and then click **Publish** again to publish this reference data set.

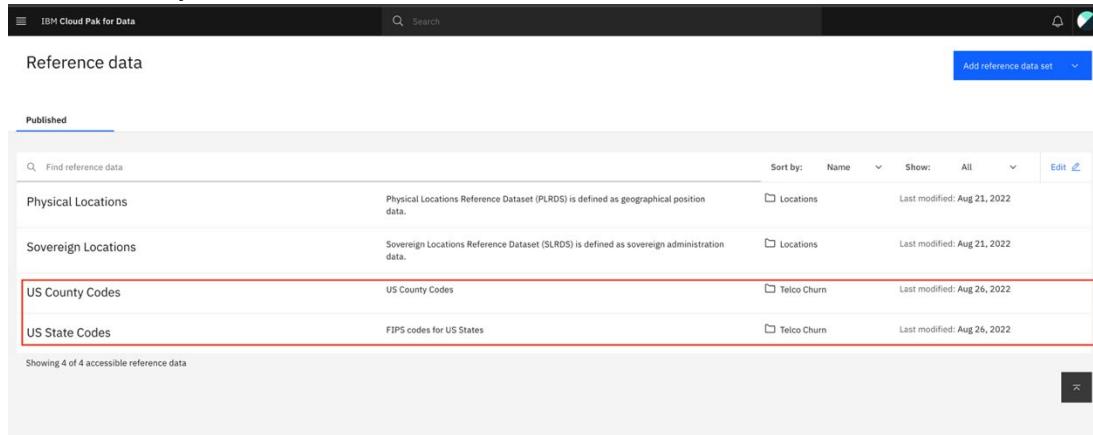
The screenshot shows the 'Publish?' dialog for the 'US State Codes' reference data set. The 'Publish' button at the bottom right of the dialog is highlighted with a red box.

10. Repeat the steps above to import the [USCountyCodes.csv](#) file into a new “US County Codes” reference data set. Select target column Code for the FIPS column and Value for the Name column. As in case of the Abbreviation column in the USStateCodes reference data set, you will create a custom column (STATE) for the State column in the USCountyCodes reference data set.

Note: Importing the USCountyCodes.csv file will take a few minutes because it is a larger data set. Refresh the webpage periodically to check if it finished importing.

Once imported, add **datasteward** as the Steward and select **churn** as the tag for the reference data set. Finally, **Publish** the reference data.

When returning to the Reference data sets (click on the **Reference data** breadcrumb) you will find both **US County Codes** and **US State Codes** reference data sets.



The screenshot shows the IBM Cloud Pak for Data interface with the "Published" tab selected under "Reference data". The page displays four reference data sets:

Name	Description	Type	Last modified
Physical Locations	Physical Locations Reference Dataset (PLRDS) is defined as geographical position data.	Locations	Aug 21, 2022
Sovereign Locations	Sovereign Locations Reference Dataset (SLRDS) is defined as sovereign administration data.	Locations	Aug 21, 2022
US County Codes	US County Codes	Telco Churn	Aug 26, 2022
US State Codes	FIPS codes for US States	Telco Churn	Aug 26, 2022

A red box highlights the "US County Codes" and "US State Codes" rows. At the bottom left, a message says "Showing 4 of 4 accessible reference data".

Map related values

Related values are used to map reference data values to other reference data values either within the same reference data set or in different reference data sets. You will see how to map a county to a US state, and this illustrates how to map one reference data set (in this case the US County Codes) to another reference data set (US State Codes).

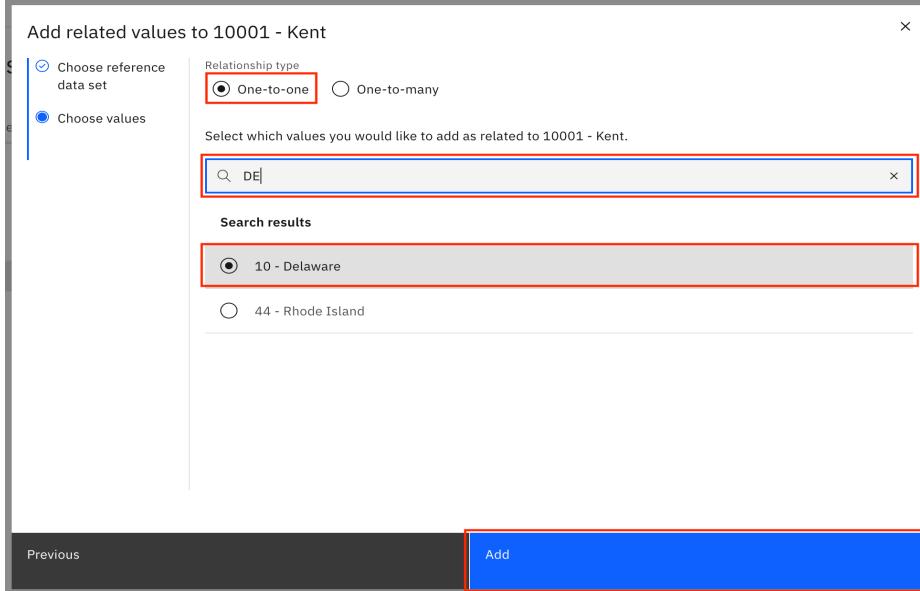
- In the US County Codes Reference data set, go to the first reference data value. Click the + sign next to Related values.

The screenshot shows the 'Reference data' section of the IBM Cloud Pak for Data interface. The 'US County Codes' dataset is selected. In the 'Values (3228)' list, the entry '10001 - Kent' is highlighted with a red box. To the right, under 'Related values', there is a '+ Add' button which is also highlighted with a red box.

- Select the **US State Codes** reference data set and click **Next**.

The screenshot shows a modal dialog titled 'Add related values to 10001 - Kent'. On the left, there are two radio button options: 'Choose reference data set' (selected) and 'Choose values'. Below is a search bar and a list of 'All reference data sets'. Under 'All reference data sets', the 'US County Codes' and 'US State Codes' options are listed. The 'US State Codes' option is selected and highlighted with a red box. At the bottom right of the modal, the 'Next' button is also highlighted with a red box.

3. Select **One-to-one** mapping. One-to-one mapping means that one value in the current reference data set (US County codes) can map to only one value in the related reference data set (US State codes). Next, find the value in the related dataset. You can also use the search bar to filter results. Select **10 - Delaware** as the US State to map Kent county to and click **Add**.



You can add related values for all the other US County Codes reference data values in a similar manner, if you wish to. In practice, you would use the Watson Knowledge Catalog API to populate the related values in an automated manner.

Note: While you have mapped the related values, these changes are still in draft mode, and will not be published until you publish the reference data set in the next section.

Set reference data hierarchy

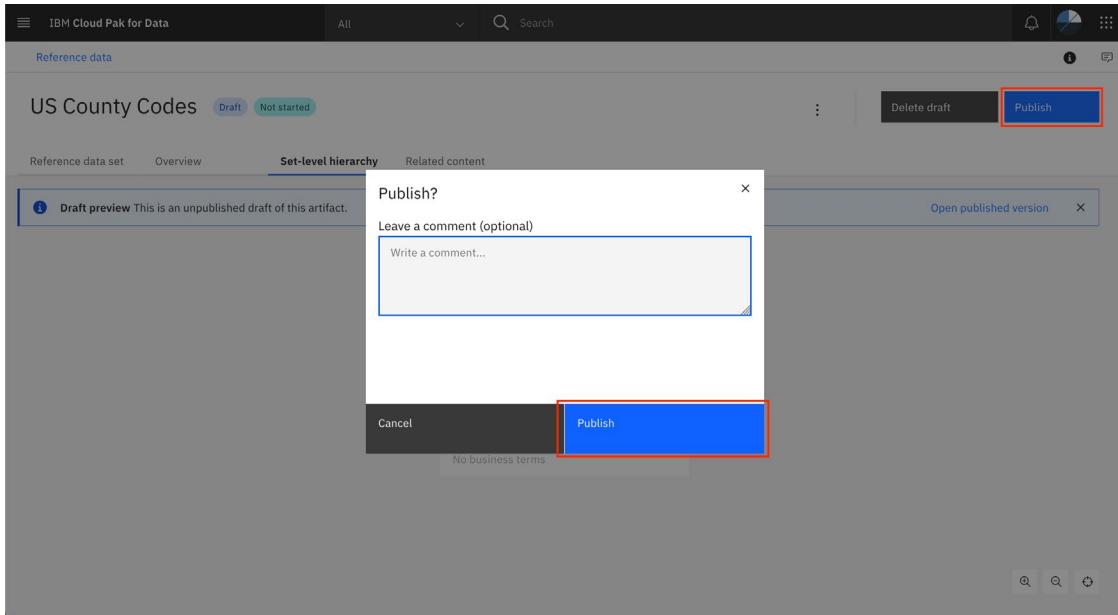
- Select the **Set-level hierarchy** tab in the “US County Codes” reference data set to create a child-parent relation between the “US County Codes” and “US State Codes” reference data sets. Click on **Add parent set** button.

The screenshot shows the IBM Cloud Pak for Data interface. At the top, there's a navigation bar with 'IBM Cloud Pak for Data', a search bar, and various icons. Below it, a card for 'US County Codes' shows it's in 'Draft' mode. The 'Set-level hierarchy' tab is highlighted with a red box. Below the tabs, a message says 'Draft preview This is an unpublished draft of this artifact.' and 'Open published version'. On the right, there are 'Delete draft' and 'Publish' buttons. The main content area displays a circular icon with a plus sign and a document, followed by the text 'No hierarchy created yet' and instructions to add either a parent or dependent set. Two buttons are shown: 'Add parent set' (highlighted with a red box) and 'Add dependent set'.

- Select the **US State Codes** reference data set by clicking the check box next to it and then click **Add**. This sets the child-parent relationship between “US County Codes” and “US State Codes”.

The screenshot shows the 'Edit reference data' dialog. On the left, there's a search bar and a list of 'Reference data' items. One item, 'US State Codes', has a checked checkbox and is highlighted with a red box. On the right, a 'Selected reference data' list shows 'US State Codes' with a count of 1, also highlighted with a red box. At the bottom, there are 'Cancel' and 'Save' buttons, with 'Save' highlighted with a red box.

- Finally, click **Publish** and then **Publish** again in the pop-up window to publish the “US County Codes” reference data set.



You have now completed the task of creating reference data sets that the analytics team will need to follow as standards. They will be able to export these reference data sets as a CSV file or use the Watson Knowledge Catalog APIs to connect to and consume the reference data values.

Data classes

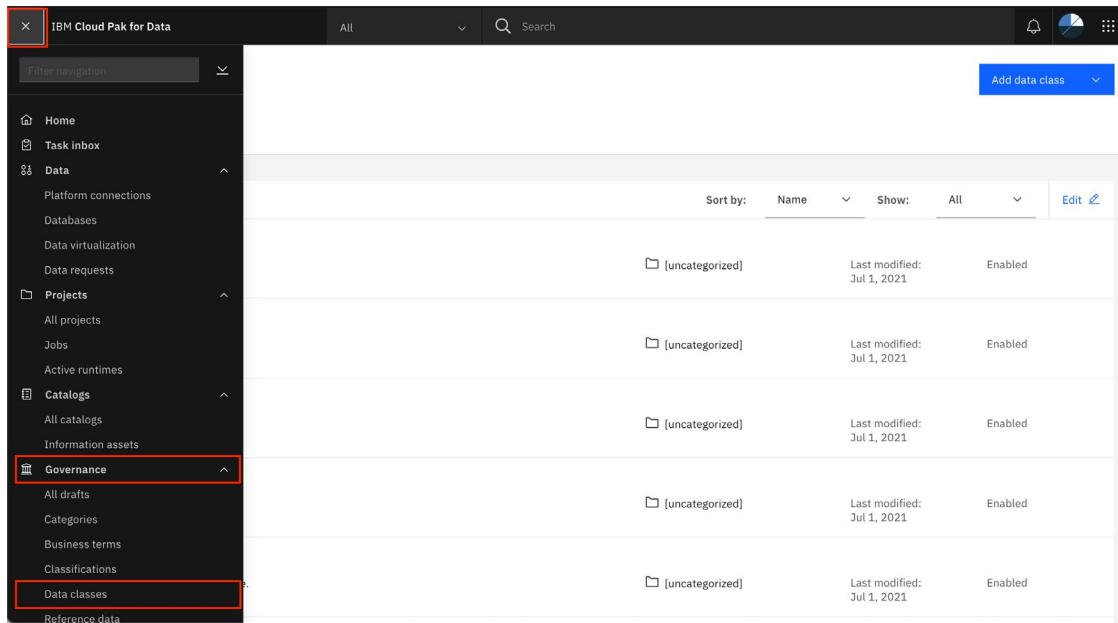
Data Classes are governance artifacts used to automatically profile catalog assets during data discovery and metadata enrichment. They can be created using a regular expression, Java class, column name, data type, list of values, or reference data set. It saves data stewards a lot of time by automatically finding key data elements and understanding where PII (Personal Identifiable Information) is stored.

Additionally, data quality analysis uses data classes to find anomalies that do not follow that regular expression, Java class logic, list of values, or reference data set codes.

You will create a data class using the US State Codes reference data set. By doing this, data stewards will be able to profile data assets to find columns that consume the values in that reference data set as well as find data quality issues where fields in those columns do not completely follow that standard.

This way, the data steward curating a catalog for Telco Churn analysis will be able to ensure that all columns referencing US State Codes are following the standard that has been created and fix any that do not follow it.

1. Navigate to the **hamburger (≡)** menu in the upper-left corner, expand **Governance** and click on **Data classes**.



The screenshot shows the IBM Cloud Pak for Data interface. The left sidebar is expanded to show the 'Governance' section, which is highlighted with a red box. The 'Data classes' item under 'Governance' is also highlighted with a red box. The main area displays a list of data classes, each with a checkbox, name, last modified date (Jul 1, 2021), and an 'Enabled' status. The list includes:

	Name	Last modified	Status
<input type="checkbox"/> [uncategorized]	Jul 1, 2021	Enabled	
<input type="checkbox"/> [uncategorized]	Jul 1, 2021	Enabled	
<input type="checkbox"/> [uncategorized]	Jul 1, 2021	Enabled	
<input type="checkbox"/> [uncategorized]	Jul 1, 2021	Enabled	
<input type="checkbox"/> [uncategorized]	Jul 1, 2021	Enabled	

- Watson Knowledge Catalog comes with 160+ data classes out of the box such as Address Line 1, State Driver's License (for multiple states), Credit Card Number, US Social Security Number, and more.

You can create a new data class by expanding the **Add data class** dropdown and selecting **New data class**.

The screenshot shows the 'Data classes' list in the IBM Cloud Pak for Data interface. The 'Published' tab is selected. A dropdown menu titled 'Add data class' is open, with 'New data class' highlighted. The list contains three items: 'Account Number', 'Address Line 3', and 'Address Line 1'. Each item has a 'Sort by:' dropdown, a 'Name' column, a 'Show:' dropdown, and a 'Last modified:' and 'Enabled' column. The 'Address Line 1' item is currently selected.

- Provide a Name (US State Code) and an optional Description (Data class to identify US State Codes) for the new data class.

Click **Change** and select the **Telco Churn** category to associate this data class with this category.

Click **Save as draft**.

The dialog box is titled 'Create new data class'. It contains the following fields:

- Data class name:** US State Code
- Primary category:** Telco Churn (highlighted with a dashed red box)
- Description (optional):** Data class to identify US State Codes
- Buttons:** Cancel and Save as draft (highlighted with a solid blue box)

4. Click the + next to “Matching method” to select how to do the matching for this data class.

The screenshot shows the 'Data classes' interface in IBM Cloud Pak for Data. A data class named 'US State Code' is selected. The 'Matching method' section is expanded, showing a toggle switch labeled 'Enabled' which is currently turned on (indicated by a green circle). Other sections like 'Primary category' and 'Secondary categories' are also visible.

5. From the available matching methods, select **Match to reference data**. Click **Next**.

The screenshot shows the 'Data matching' configuration dialog. Under the 'Matching method' section, the 'Match to reference data' option is selected and highlighted with a red box. The other options ('No automatic matching', 'Match to list of valid values', 'Match to criteria in regular expression', 'Match to criteria in deployed Java class', and 'Other matching criteria') are shown but not selected. At the bottom of the dialog, the 'Next' button is also highlighted with a red box.

6. The other matching methods are as follows:

- No automatic matching: This means that the data class can be manually selected to assign to a column, but the system will not automatically profile any columns as matching that data class.
- Match to list of valid values: Selecting this will allow you to provide a list of values to match to. It is recommended to save these values as reference data sets instead, but this is a good option if you prefer not to have so much formal management of that list.
- Match to reference data: While profiling columns, the system will evaluate against values in a reference data set.
- Match to criteria in a regular expression: A regular expression is used to determine if values match a data class.
- Match to criteria in deployed Java class: The logic specified in a Java class determines if each value of a database column or the whole database column belongs to the data class.
- Other matching criteria: Matching is based on criteria about the name and/or the data type of the column only. There are no additional criteria to evaluate the values of the column.

7. Select the **US State Codes** reference data set.

Keep the default percentage match threshold of **80%**. This means that, if 80% of the values in a column match the values in that reference data set, then that column will be classified as the “US State Code” data class.

Click **Next**.

The screenshot shows a configuration interface for 'Data matching'. At the top, there are four tabs: 'Select matching method' (which is selected), 'Define data matching', 'Other matching criteria (optional)', and 'Matching priority (optional)'. Below these tabs is a section titled 'Match to reference data' with a search bar labeled 'Find reference data sets'. Underneath the search bar is a list titled 'Reference data sets' containing 'US County Codes' and 'US State Codes'. The 'US State Codes' option is selected and highlighted with a red box. At the bottom of the interface is a 'Percentage match threshold (optional)' input field with the value '80', accompanied by minus and plus buttons. At the very bottom are two buttons: 'Back' on the left and 'Next' on the right, with the 'Next' button also highlighted with a blue box.

- Watson Knowledge Catalog supports additional matching criteria based on column name and/or data types. However, for now, you can leave those empty (default values) and only leverage the reference data set to determine the data class.

Click **Next**.

Data matching

Select matching method Define data matching Other matching criteria (optional) Matching priority (optional)

Other matching criteria

Column name criteria ⓘ

Type a regular expression

Test column name criteria

Sample column name

Column name criteria not specified

Column data type

Any data type

Minimum length of data value Maximum length of data value

Back Next

- Finally, you can change the priority of this rule related to other matching rules which provides more granular control over to which data class a column will be mapped.

Data matching

Select matching method Define data matching Other matching criteria (optional) Matching priority (optional)

Matching priority

Data class priority

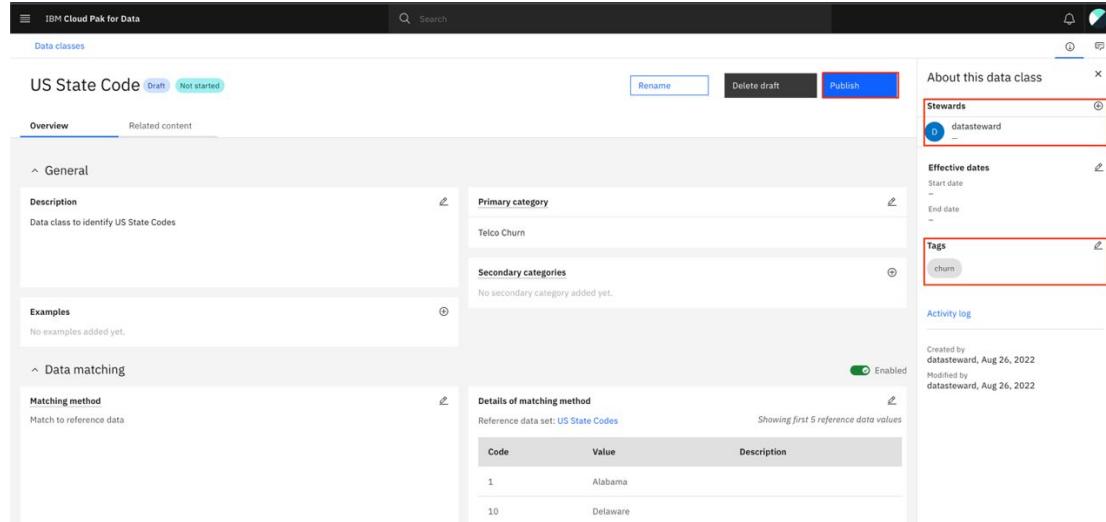
10

Change affects order relative to other data classes
Make sure you have assessed that the changed priority affects the overall order as intended.

Back Save

10. Add the **datasteward** user as Steward, add churn tag to the Tags.

Click **Publish** and then click **Publish** again in the pop-up window to publish the data class.



The screenshot shows the 'Data classes' section of the IBM Cloud Pak for Data interface. A data class named 'US State Code' is selected. The top navigation bar includes 'IBM Cloud Pak for Data', a search bar, and a 'Publish' button which is highlighted with a red box. Below the navigation, there are tabs for 'Overview' and 'Related content'. The 'Overview' tab is active, showing the following details:

- General**:
 - Description: Data class to identify US State Codes
 - Primary category: Telco Churn
 - Secondary categories: No secondary category added yet.
 - Examples: No examples added yet.
- Data matching**:
 - Matching method: Match to reference data
 - Enabled: Enabled (button)
 - Details of matching method:
 - Reference data set: US State Codes
 - Showing first 5 reference data values

Code	Value	Description
1	Alabama	
10	Delaware	

On the right side of the interface, there is a sidebar titled 'About this data class' containing sections for 'Stewards' (with 'datasteward' listed), 'Effective dates' (both start and end date fields are empty), and 'Tags' (with 'churn' listed). At the bottom right, there is an 'Activity log' section showing creation and modification logs by 'datasteward' on August 26, 2022.

Summary

In this tutorial, you learned how to set up data governance for your enterprise. The enterprise teams can now efficiently apply their data analysis and data science models for improved Telco Customer Churn prediction while meeting the governance and compliance requirements for the enterprise.

You have learned how to leverage Watson Knowledge Catalog to perform the following tasks:

- Create categories to define logical structure of governance artifacts.
- Configure workflows for governance artifact requests and assign user permissions.
- Create governance artifacts such as policies, rules, business terms, data classes and reference data sets. You learned how to do this manually through the UI and by importing CSV files.
- Update workflow requests to publish and deliver requests as you perform the various tasks associated with the governance artifact requests.