Preview: Governed Data Stewardship on PrivaceraCloud

Table of Contents

Overview of Governed Data Stewardship on PrivaceraCloud

NOTE
Contact Privacera Support to request enabling this feature.

Your organization needs a flexible, collaborative solution to data governance problems such as the following:

- A centralized team building new products that stretch across multiple technological domains and datasets.
- A consumer-oriented group developing a data mesh composed of a large number of data sources and formats.
- Implementation of a bronze/silver/gold pattern for data access and use depending on specific needs.
- Data scientists building out new models for analytics or other needs.

Privacera Governed Data Stewardship's flexible architecture can enforce a separation of data, access control, and compliance adaptable to your data governance project needs.

The fundamental purposes of Privacera Governed Data Stewardship are to:

- Subdivide your data into fine-grained data domains of cloud applications that have your data.
- Create shared datasets based on those data domains.
- Define data owners, optional data stewards, and optional project leaders for those shared datasets and data domains.
- Control access by users, groups, and roles to data domains, shared datasets, and projects.

Privacera Governed Data Stewardship is based on PrivaceraCloud features such as resource policies, access policies, and permissions.

- Your account administrator sets up the basic PrivaceraCloud building blocks: connected applications and users.
- After these building blocks are available, Privacera Governed Data Stewardship simplifies and streamlines data governance, automatically creating policies and permissions as identified data owners define data domains, shared datasets, and projects.

Concepts in Governed Data Stewardship

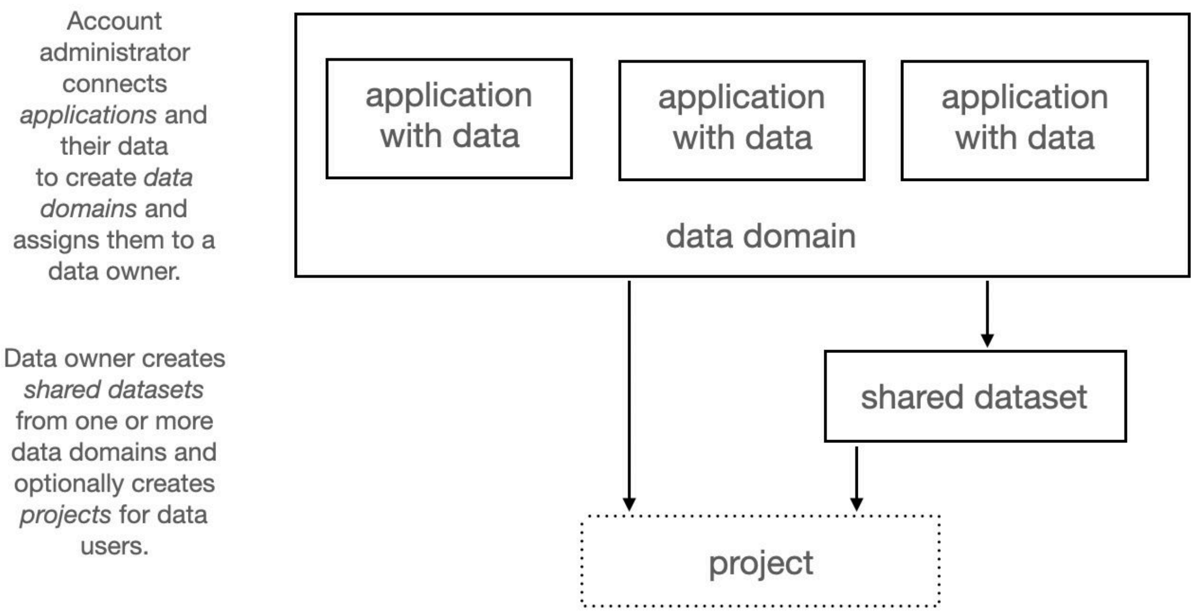
The high-level relationship among applications with data and the people involved in PrivaceraCloud Governed Data Stewardship is described in this section.

Data relationships

These diagrams present two different views of the relationships of data in Privacera Governed Data Stewardship. The functions of some of the roles that work with the data are further described in Hierarchy of Roles [5].

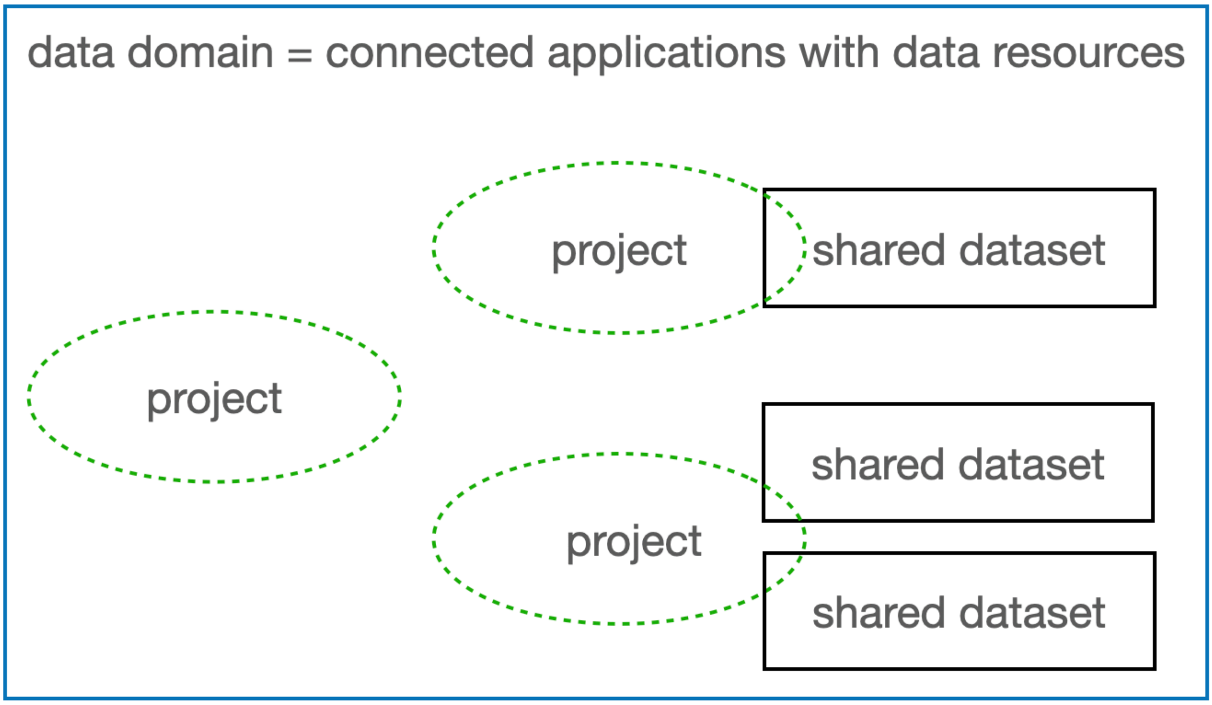
Dotted lines in these diagrams indicate features that are optional.

The following diagram represents how the data in Governed Data Stewardship is mechanically created by the various roles.



Account administrator connects *applications* and their data to create *data domains* and assigns them to a data owner.
application with data
application with data
application with data
data domain
Data owner creates *shared datasets* from one or more data domains and optionally creates *projects* for data users.
shared dataset
project

The following diagram represents how the data in Governed Data Stewardship might be used by your organization day-to-day.



Application with data resources

An application defines a third-party system that contains the data analyzed by PrivaceraCloud. The account administrator connects applications to PrivaceraCloud to make their data accessible. Depending on the type of application, it can contain resources or databases or tables.

- For a list of applications supported by Governed Data Stewardship, see Supported Applications [9].
- For details about adding applications to PrivaceraCloud, see Connect applications to PrivaceraCloud.

Resources is a generic term for the data made available to PrivaceraCloud by connecting an application. For example, files in an application, such as .csv or .json files in an S3 bucket, are resources.

A *database* is a single collection of data in an application, and a *table* is a subset of a database with a distinct schema.

Data domain and shared dataset

A *data domain* is a defined combination of applications with data that can be operated on as a whole for the purpose of access control or Privacera Discovery scans.

- A data domain is a logical abstraction, whereas an application with data represents a physical third-party system that has been connected to PrivaceraCloud by the account administrator.
- A data domain can include multiple applications with data of different types.
- The account administrator constructs a data domain and assigns the data owner of the domain.
- A *shared dataset* is composed of one or more data domains that the data owner or data steward shares with data users.

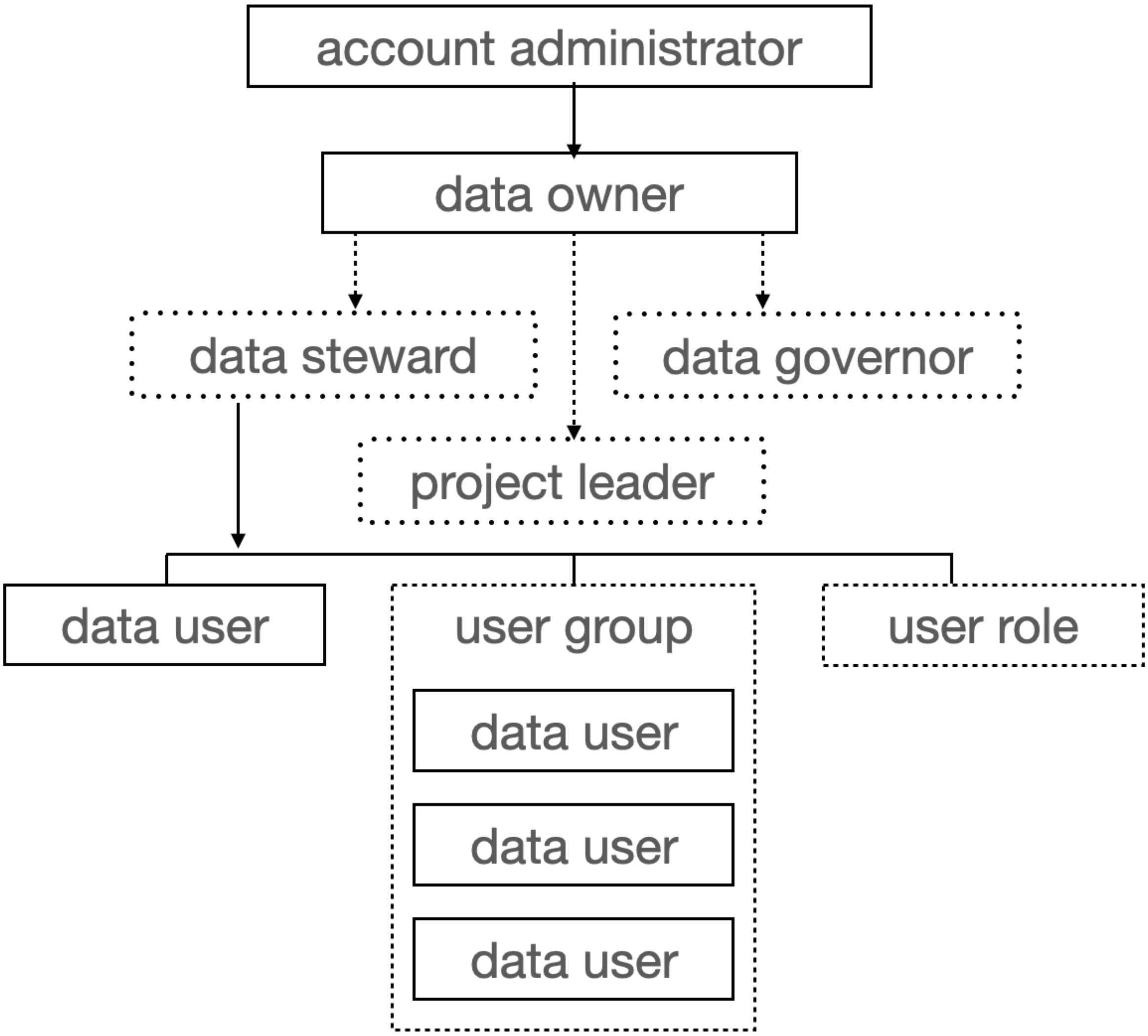
Project

With a specific goal defined by the data owner or data steward, a project can be created directly from a data domain or be composed of one or more shared datasets.

Hierarchy of roles

This diagram shows the logical hierarchy of relationships among the roles involved in Governed Data Stewardship.

Dotted lines in this diagram indicate features that are optional.



Account administrator

The account administrator is the first person who created an account for your organization on PrivaceraCloud.

The account administrator:

- Creates users and groups.
- Connects applications with data.
- Defines data domains based on applications.
- Assigns data domains to data owners.
- Can run Privacera Discovery scans on data domains.

Data owner and data steward

A data owner is a PrivaceraCloud user who has been assigned the data owner role by the account administrator for a particular data domain.

An optional data steward is a PrivaceraCloud user who has been assigned as a delegate by the data owner of a particular data domain.

There is no limit to the number of data owners or data stewards of a data domain.

A data owner or data steward:

- Creates and shares datasets composed of data domains.
- Can optionally delegate most of these functions to data stewards.
- Gives access to datasets to users, groups, or roles.
- Can make shared datasets or projects discoverable by data users.
- Accepts or rejects requests from data users to access shared datasets.
- Can optionally define projects.
- Can optionally assign project leaders to projects.
- Assigns users, groups, and roles to projects.
- Grants read/write access permissions to users, groups, or roles in datasets, resources in those datasets, or projects.
- A data owner can run Privacera Discovery scans on data domains, shared datasets, and projects.

> **NOTE**
> Except for running Privacera Discovery scans, a data steward has all the same functions of a data owner.

Project leader

An optional project leader is a PrivaceraCloud user assigned to projects defined by the data owner or data steward.

There is no limit to the number of project leaders of a project.

A project leader:

- Can add resources that they own to defined datasets.
- Can add users, groups, and roles to projects.
- Can accept or reject requests from data users to access shared datasets.

Data governor

A data governor is a PrivaceraCloud user who has been assigned this role by the data owner. Data governors have the function of an auditor.

A data governor:

- Can see all data in data domains, shared datasets, resources, projects, and discovery scan results to which they have been given access.
- Cannot change the data in any of those data domains, shared datasets, resources, or projects.
- Can initiate discovery scans.
- Can cancel discovery scans started by other users for data to which the data governor has access.

Data user

A data user is a PrivaceraCloud user who has been assigned certain Privacera system roles by the account administrator. A data user is given access to data domains, shared datasets or projects with certain permissions by a data owner, data steward, or project leader.

Data user is a general term for many different work functions that your organization might have. For example, you might have data analysts, ETL programmers, data scientists, and auditors.

For simplicity, Governed Data Stewardship abstracts these various functions into a single role: data user. Your organization's definition of these various possible functional roles is for you to decide.

A data user:

- Can request access to shared datasets that have been made discoverable by data owners or stewards.
- Can access shared datasets that they have been given permission to see.

Supported Applications

PrivaceraCloud Governed Data Stewardship supports the following applications.

Supported applications for Privacera Access Management:

- ADLS
- Athena
- Aurora DB
- Databricks
- Databricks SQL
- Databricks Unity Catalog
- Dataproc
- Dremio
- Dynamo DB
- EMR
- Files
- GCS
- Glue
- Microsoft SQL
- PostgreSQL
- Presto
- Redshift
- S3
- Snowflake
- Starburst Enterprise
- Starburst Enterprise Presto
- Trino
- Textract

Supported applications for Privacera Discovery:

- Databricks SQL
- Microsoft SQL
- Redshift
- S3
- Snowflake

Prerequisites and planning

Before you begin to define data domains, shared datasets, projects, or roles, prepare the following:

- Connect applications to PrivaceraCloud.
- Data access users, either by creating them manually, as described in Users, groups, and roles or by loading them from an external IdP, such as LDAP, Azure AD, or a SCIM server, as described in UserSync integrations.
- User role assignments, as listed in Map of Governed Data Stewardship roles to Privacera system roles [10].

High-level planning for Governed Data Stewardship

The general planning for Governed Data Stewardship is as follows:

1. Define easily-remembered names of the data domains to be created by the account administrator, including a useful description of the data domain.
2. For database applications, decide the depth of the resource to add to the data domain: the entire database, the tables, or the columns and rows, as described in Application with data resources [5].
3. Decide if the data owners should have optional data stewards to manage the data domain.
4. Identify data owners, optional data stewards, and optional project leaders to be given access to these data domains.
5. Assign appropriate Privacera system roles to the data owners, optional data stewards, and optional project leaders, as described in Map of Governed Data Stewardship roles to Privacera system roles [10].
6. Decide:
   - Which users, user groups, or roles to share the datasets with.
   - Which users, groups, or roles should have read/write permission to the data domain
   - Which users should have only read permission.
7. Optionally define and describe the projects that each data owner controls.
8. Decide if you want to create projects from a data domain, a shared dataset, or both.
9. Decide if you want to assign project leaders to the projects.
10. Determine which data domains, users, user groups, or roles should be assigned to which projects.

Map of Governed Data Stewardship roles to Privacera system roles

Each role for Privacera Governed Data Stewardship is a data access user that must be given a certain PrivaceraCloud system role, as shown in the table below.

For details on creating data access users and assigning roles, see Users, groups, and roles.

GDS Role	Privacera Role	Description
account administrator	ROLE_ACCOUNT_ADMIN	The default account administrator is the person who first signed up for PrivaceraCloud. Other account administrators can be created.
data owner and steward	ROLE_POLICY_ADMIN	Data owners and data stewards are portal users created manually or loaded from an external IdP and given this Privacera role.
data governor	ROLE_DISCOVERY_GOVERNANCE	A data governor is a portal user created manually or loaded from an external IdP and given this Privacera role.
project leader	ROLE_POLICY_ADMIN	A project leader is a portal user created manually or loaded from an external IdP and given this Privacera role.
data user	ROLE_USER	A data user is a portal user created manually or loaded from an external IdP and given this Privacera role, which is the default.

Additional features

Applications and database resources

Data domains can include any type of application with data that the account administrator has already connected. An application added to a data domain is called a *resource*.

By default, an application's resources are created for an entire database or storage location (for example, a PostgreSQL database or an S3 bucket). Additionally, resources can be created at a single storage path or at a particular database schema, table, or column.

Granular permissions on resources

A data owner or data steward can specify user permissions for either an entire dataset or individual resources in that dataset.

For example, the data owner might specify read permission for a dataset but read/write permission for individual resources.

The permissions on a resource override the permissions on the dataset.

Automatic expiry of access for shared datasets or projects

When creating a shared dataset or project, the data owner can specify a date when access expires. The default is **Never**.

At-a-glance dashboards by role

You can see an at-a-glance dashboard summary of your work in Governed Data Stewardship.

The dashboard summarizes the number of data domains, shared datasets, and projects. Also included are breakdowns of counts by connected application.

The details depend on your role, such as data owner, data steward, project leader, or data user. You can see only those details to which you have access.

To see the dashboard:

1. Go to **Datasets > Overview**.
2. Scroll down to see a graphical summary of data domains, shared datasets, and projects.

Optional data steward

A data owner can optionally delegate a data steward.

The data steward has all the same functions as a data owner except for running Discovery scans.

Privacera Discovery scans by admin or data owner

To classify sensitive data, the account administrator or data owner can initiate Privacera Discovery scans.

- The account administrator can scan an entire data domain.
- The data owner can scan a data domain, a dataset, a resource in a dataset, or a project.

Scan status is displayed in **Discovery > Scan Status**.

Real-time scanning is not supported.

Optional project leader

A data owner or data steward can optionally delegate to a project leader the function of managing data user access requests to datasets.

Optional terms of use

The data owner can optionally require that data users accept terms of use before they can access a shared dataset. The terms of use are freeform and can be any text the data owner wants.

Discoverability of shared datasets

The data owner can optionally specify that a shared dataset is discoverable by users in your Privacera-Cloud account.

User request access to datasets

Data users can request access to datasets that the data owner has configured as discoverable.

Notifications

All users of Governed Data Stewardship receive notifications about various actions by other users.

To view the notifications, do one of the following:

- See the bell icon in the upper right.
- Go to **Datasets > Notifications**.

On **Datasets > Notifications**, you can filter notifications by type, such as access requests or resources added to datasets.

Overview to examples by role

As background, see the following:

- Concepts in Governed Data Stewardship [4]
- Prerequisites and planning [10]

These examples of step-by-step procedures show how the various roles use PrivaceraCloud to implement portions of Governed Data Stewardship.

NOTE

- These example steps are presented in a logical sequence of an organization's hypothetical use.
- The steps do not show all possible features or functions of Governed Data Stewardship.
- In day-to-day work, you will discover additional sequences of steps not explicitly detailed here.

1. The account administrator [13] creates a data domain and assigns a data owner.
2. The data owner [14] creates a shared dataset and delegates it to a data steward [15].
3. The data owner [14] runs a Privacera Discovery scan on a dataset to discover and classify sensitive data.
4. The data steward [15] shares a data domain with data users.
5. The project leader [16] manages data users [17]' requests to access datasets.

Account administrator

This example is for the account administrator or delegate in Governed Data Stewardship.

As an account admin, you have the following general tasks:

- Connect applications
- Create users
- Define data domains for Governed Data Stewardship
- Assign data domains to data owners

Prerequisites: set up applications and users

Before creating a data domain, prepare the following:

- Connect applications
- Defined portal users with appropriate roles

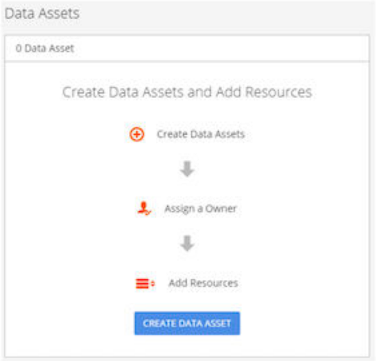
Define data domain, assign data owner, and add datasources

To create a data domain and assign a data owner:

1. Go to **Data Assets > Datasets**. You are presented with a wizard.

NOTE

This wizard appears only the first time you create a data domain. Thereafter, when you create other data domains, you use the tabs and fields that are displayed, not the wizard.



2. Click **CREATE DATA DOMAIN** at the bottom.
3. Enter an easy-to-remember name and a description of the data domain.
4. Click the **OWNER** tab.
5. Select **TYPEUser**, select the desired user as the data owner, and leave the default read/write permission or change the permission if necessary.
6. To assign additional data owners, click the plus sign on the right.
7. To create the data domain, click **CREATE**.
8. To add resources to the data domain, click **ADD RESOURCE**.

   You can instead click **BROWSE** to locate the desired application and resource.
9. From the displayed applications, select an application that you had already configured:
   - Required information varies by the type of application.
   - This example shows how to add a database table resource.
10. Enter the first characters of a database name to see and select the database.
11. Enter the first characters of a table name to see and select the table.
12. Click **ADD**. If this data domain needs additional applications, continue to add them.

The data domain has been created, your work is complete, and the data owner receives a notification.

Data owner

This example is for the data owner in Privacera Governed Data Stewardship.

See data domains

At **Datasets > Data Domains**, you can verify the work done by the account administrator so that the proper resources have been added to the data domain.

Create shared dataset

These steps assume you as data owner want to give only read-only permission to a user group, but you can also give read/write permission to individual users or roles.

To create a shared dataset:

1. Navigate to **Datasets > Shared Datasets.**
2. Click **Create Shared Dataset**.
3. In the **Name** and **Description** fields, add an easy-to-remember name and description, perhaps including who you are sharing with.

4. Rely on the defaults for other fields, or change them as desired:
   - Default: **Duration: Never Expires**.
   - Default: Unchecked **Discoverable**. The dataset is visible only to users with whom it has been shared.

Assign data steward

Continuing the example of creating a shared dataset, you assign a data steward.

To delegate to an optional data steward:

1. On the **DATA CO-OWNER** tab:
2. From the **TYPE** pulldown menu, select the desired **User**, **Group**, or **Role**.
3. In the **Select** field, find and select the required user, group, or role.
4. On the **ACCESS** tab, from the **TYPE** pulldown menu, select the desired **User**, **Group**, or **Role**.
5. In the **Select** field, find and select the required user, group, or role.
6. From the next field, select **Read** to disallow data users from being able to write.
7. Click **CREATE**.

You have created the shared dataset and delegated management functions to a data steward.

Data owner: Privacera Discovery scans

This example is for the data owner in Governed Data Stewardship.

To start a scan of a resource in a shared dataset, check the scan status, and view the results:

1. Go to **Datasets > Data Domains**.
2. Click the name of the desired data domain.
3. Click the desired shared dataset and drill down to see the resources included in that shared dataset.
4. To scan of the desired resource, on the right at the end of the row, click the scan icon:

   🔍

   The scan ID is displayed.
5. To check scan status, on the left, click **Discovery > Scan Status**, find the desired scan ID, and note its status.

After the scan is complete, a summary of the classification is viewable at **Datasets > Shared Datasets** > *name of desired shared datatset* > **INFO** tab.

Data steward

This example is for the data steward in Governed Data Stewardship.

Define Terms of Use

Continuing the example of the data owner creating a shared dataset [14], you as the data-steward define terms of use that the data users must agree to before accessing the dataset.

> **NOTE**
>
> **Terms of Use** are optional and are activated only if you enter text in the **Terms of Use** text box.

To define terms of use:

1. Click the **TERMS OF USE** tab**.**
2. Enter optional terms of use details in the **Terms of Use** text box.
3. Click **SAVE**.
4. On the **RESOURCES** tab, in the upper right corner, click **BROWSE** to browse the defined resources for this data domain.
5. For this resource, use the checkboxes on the left to drill down to the depth of the resource. Check the checkboxes for the desired data tables.

You have defined the terms of use for this shared dataset.

Share Dataset with data users as read-only

Continuing the example of creating a shared dataset, you share the dataset as read-only.

To share a dataset as read-only:

1. In the upper right corner, click the plus sign.
2. From the **Select a dataset** pulldown, select the appropriate data domain to share with QA.
3. Click **SHARE WITH**.
4. Select type **Group**, the name of the group, and read-only permission.
5. Click **ADD**.

The dataset has been shared as read-only.

Create a project, assign a project leader, and add users to the project

Projects can be created for data domains, shared datasets, or both. The steps here create a project from a data domain.

To create a project and add a data domain to it:

1. Navigate to **Datasets > My Projects**.
2. Click **CREATE PROJECT**.
3. Enter an easy-to-remember name and description for the project.
4. Select the data domains to assign to the project.
5. Leave the default duration for the project, which is **Never Expired**, or specify a time period.
6. To add a project leader, click the **LEADERS** tab.
7. From the pulldown, select **User**, **Group**, or **Role**.
8. Enter the name of the user, group, or role.
9. Add as many other project leaders as desired.
10. Click **CREATE**.

You have defined a project, assigned a project leader, and added users to the project.

Project leader

This example is for the project leader in Governed Data Stewardship.

Manage approvals for access requests from data users

As described in Discoverability of shared datasets [12], data users can request access [12] to shared datasets and projects that have been made discoverable.

Notifications are sent about these access requests.

You can view and filter notifications as described in Notifications [12].

To filter and view notifications of dataset access requests:

1. Go to **Datasets > Notifications.**
2. Click the **Access Request** filter.
3. Cycle through the request notifications and approve or reject them.

Data user

This example is for the data user in Governed Data Stewardship.

After a data owner gives you access to a shared dataset or project, you receive a notification. You then work with that data just as you normally would.

Discover shared datasets

To see the shared datasets you can access, go to **Datasets** > **Shared With Me**.

You can also search for shared datasets and request access to them from the data owner or data steward.

Request access to shared datasets

After you have discovered shared datasets, you can request access to them. These requests are approved or denied by the data owner or data steward.

Agree to terms of use

A data owner might have required optional **Terms of Use**, which you must accept before you can access a shared dataset. The first time you access the shared dataset, you are prompted to agree.