



Informatica® Cloud Data Quality
July 2024

Introduction

Informatica Cloud Data Quality Introduction

July 2024

July 2024

© Copyright Informatica LLC 2019, 2024

This software and documentation are provided only under a separate license agreement containing restrictions on use and disclosure. No part of this document may be reproduced or transmitted in any form, by any means (electronic, photocopying, recording or otherwise) without prior consent of Informatica LLC.

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, duplication, disclosure, modification, and adaptation is subject to the restrictions and license terms set forth in the applicable Government contract, and, to the extent applicable by the terms of the Government contract, the additional rights set forth in FAR 52.227-19, Commercial Computer Software License.

Informatica, Informatica Cloud, Informatica Intelligent Cloud Services, and the Informatica logo are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. A current list of Informatica trademarks is available on the web at <https://www.informatica.com/trademarks.html>. Other company and product names may be trade names or trademarks of their respective owners.

Portions of this software and/or documentation are subject to copyright held by third parties.

The information in this documentation is subject to change without notice. If you find any problems in this documentation, report them to us at infa_documentation@informatica.com.

Informatica products are warranted according to the terms and conditions of the agreements under which they are provided. INFORMATICA PROVIDES THE INFORMATION IN THIS DOCUMENT "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

Publication Date: 2024-07-11

Table of Contents

| | |
|---|-----------|
| Preface | 4 |
| Chapter 1: Introduction to Data Quality | 5 |
| Informatica Intelligent Cloud Services. | 5 |
| Data Quality. | 6 |
| Data quality life cycle. | 7 |
| Dimensions of data quality. | 7 |
| Data Integration. | 9 |
| Data Integration tasks | 9 |
| Data Profiling. | 9 |
| Administrator. | 11 |
| Monitor. | 12 |
| Data Observability. | 12 |
| Informatica Intelligent Cloud Services security. | 12 |
| Learning resources for Data Quality. | 13 |
| Chapter 2: Data Quality tools | 14 |
| Projects. | 14 |
| Assets. | 14 |
| Explore page. | 15 |
| Chapter 3: Informatica resources | 18 |
| Informatica Intelligent Cloud Services web site. | 18 |
| Informatica Intelligent Cloud Services Communities. | 18 |
| Informatica Intelligent Cloud Services Marketplace. | 18 |
| Informatica Intelligent Cloud Services Trust Center. | 18 |
| Informatica Product Availability Matrices. | 19 |
| Informatica Documentation. | 19 |
| Data Integration connector documentation. | 19 |
| Informatica Knowledge Base. | 19 |
| Informatica Global Customer Support. | 20 |
| Index | 21 |

Preface

Refer to the *Introduction* for an overview of Informatica Intelligent Cloud Services™ Data Quality.

CHAPTER 1

Introduction to Data Quality

Data Quality is one of several services available in Informatica Intelligent Cloud Services. Data Quality is an on-demand subscription service that provides all the features that you need to verify, standardize, and improve the quality of your business data.

Informatica Intelligent Cloud Services

When you log in to Informatica Intelligent Cloud Services, the **My Services** page opens onscreen. In addition to Data Quality, the **My Services** page might include other services that you subscribe to and trial versions of other services.

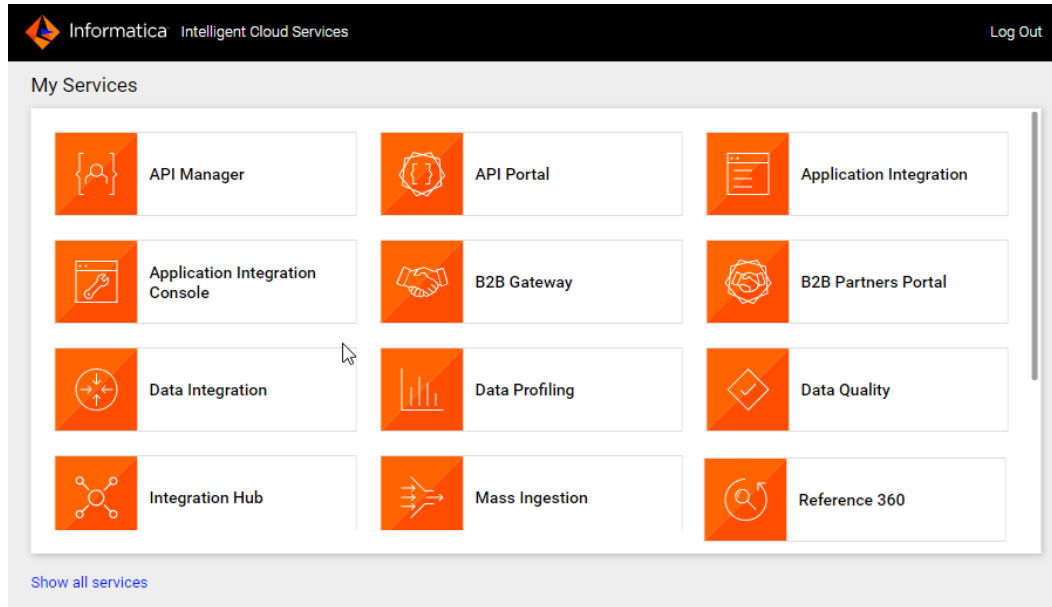
As a Data Quality user, you might work in the following services:

- Data Quality. Create assets to verify, standardize, and improve the quality of your data.
- Data Profiling. Create and run data profiling tasks.
- Data Integration. Manage data integration projects, create assets such as mappings, tasks, and taskflows, and run tasks and taskflows.

Note: When you define a mapping in Data Integration, you can include transformations that contain data quality assets.

- Administrator. Perform administrative tasks, for example to configure address verification properties in your runtime environment.

- Monitor. Check the status of tasks that are in progress and tasks that have completed.

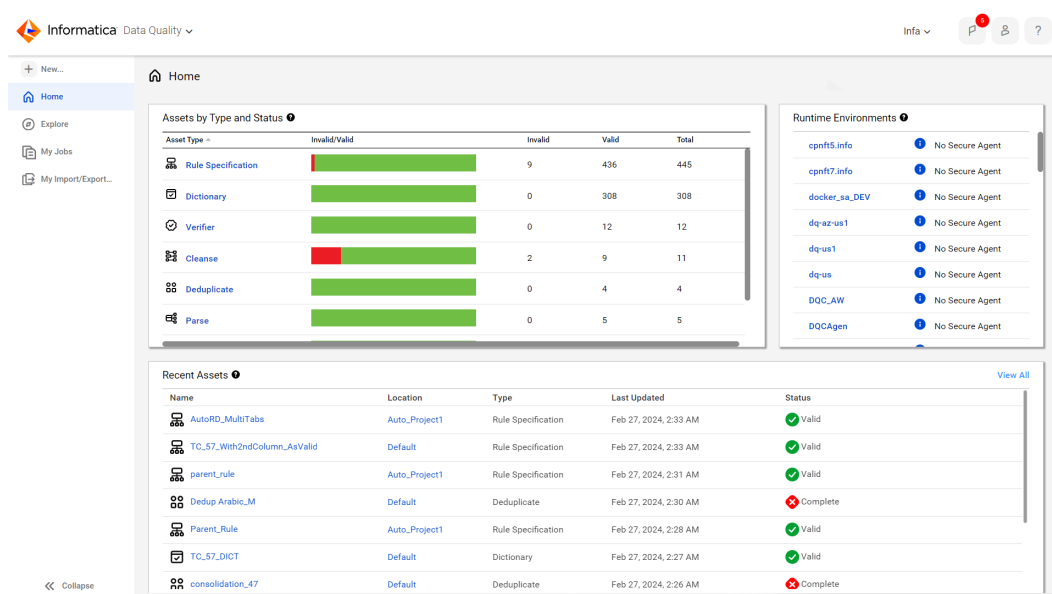


Data Quality

Use Data Quality to create data quality assets. Add the assets to transformations in a mapping in Data Integration.

When you select Data Quality from the **My Services** page, the **Home** page appears.

The following image shows the Data Quality **Home** page:



The **Home** page displays the following panels:

- Runtime Environments. The status of the runtime environments that are available to you.
- Assets by Type and Status. A summary of the assets that you can access.
- Recent Assets. The assets that you most recently modified.

You can access the following pages from the navigation bar for Data Quality:

- New. Create dictionaries, cleanse assets, deduplicate assets, parse assets, labeler assets, rule specification assets, and verifier assets.
- Home. Return to the Home page.
- Explore. Manage project and project folders. Find and open data quality assets, data profiling tasks, and exception tasks.
- My Jobs. View the status of the jobs that run in your organization.
- My Import/Export Logs. View the status of your imports and exports.

When you switch from Data Quality to another service, the panels and the options in the navigation bar change to suit the service.

Data quality life cycle

The assets that you configure for your data quality projects constitute a set of operations that you can perform across Informatica Intelligent Cloud Services.

To understand and improve the quality of your data, you can move the data through the following stages:

1. **Discover.** Analyze the content and structure of your source data.
To analyze the content and structure, create a profile in Data Profiling.
Note: You can open and run profiles from the Explore page in both Data Profiling and Data Quality.
2. **Design.** Create assets to address the issues that you find in the source data.
Create the assets in Data Quality.
3. **Apply.** Add the assets to one or more mappings, and run the mappings on the data.
Design and run the mappings in Data Integration.
4. **Measure.** Run profiles to review the results of the mappings.
Optionally, update the assets that you created in Data Quality and run the mappings again to optimize the quality of the data.

Dimensions of data quality

Your organization may target a range of objectives when you build data quality initiatives into your data systems. For example, you may need to eradicate duplicate records in order to comply with regulatory standards. Or, you might recognize that postal address accuracy is sub-optimal across your records. Or, you might decide to mine additional information and value from your current data.

The needs of every organization are unique, but the data quality issues that your data may demonstrate can fall into a range of common categories. Data Quality assets can identify these categories as *dimensions*.

You can set the **Dimension** option in data quality assets that correspond to passive transformations in Data Integration. Set the option to specify the data quality issue that you want the asset to address. You can set the Dimension option in a cleanse, labeler, parse, rule specification, and verifier asset. A scorecard can read the dimension that you set on a rule specification asset.

You can select one of the following dimensions:

Accuracy

Select Accuracy when the asset logic is primarily concerned with establishing the accuracy of data values. Data is accurate when it matches a known data fact that the asset can verify.

For example, a business rule may require that each employee in an organization has the correct data security clearance for their role. The organization maintains a set of personnel records that includes the security clearance level and job title of each employee. You can configure an asset to compare the security clearance data to the job title data in each record and to verify that the values match accurately.

You might use dictionaries that contain the job titles and the security clearance levels to verify that the respective data values are correct.

Validity

Select Validity when the asset logic is primarily concerned with establishing the validity of the data. Data is valid when it meets the formal and structural requirements of a business rule that your organization defines. For example, valid data might use the data type and conform to the character length that the business rule expects.

Note: Validity and consistency are similar dimensions. However, data values can be consistent but not valid. Consistency is a measure of the similarity in form between the data values in a column. Validity is a measure of the correspondence between the formal aspects of the column data and the format that your organization requires.

Completeness

Select Completeness when the asset logic is primarily concerned with establishing the completeness of the data.

For example, a business rule in your organization might require that one or more data columns do not contain null data. You can configure a rule specification with one or more rule statements that search the relevant columns for null data.

Consistency

Select Consistency when the asset logic is primarily concerned with establishing the consistency of the data within one or more columns. The data in a column is consistent when the column values conform to a uniform character format. Additionally, column data can be consistent in the use of an agreed set of terms for different pieces of information. For example, you might configure a cleanse asset to standardize street descriptors such as Street and Road to ST and RD.

Uniqueness

Select Uniqueness when the asset logic is primarily concerned that a data set does not contain duplicate records. Two or more records are duplicates of each other when they refer to the same data entity with substantially the same data. To report on the uniqueness of the records, use a deduplicate asset.

A deduplicate asset applies a threshold score to the results of the comparisons that it makes between pairs of records in a data set. You can feed the output from a Deduplicate transformation to a Rule Specification transformation in a mapping, and you can configure the Rule Specification transformation to apply a status value to records according to their threshold scores. You can assign the Uniqueness dimension to the rule specification asset in the Rule Specification transformation.

Timeliness

Select Timeliness when the primary purpose of the asset is to verify that the record data is current. Current data represents the most recent version of a data fact.

For example, a retail organization might require that warehouse inventory records are updated every day. You can define a rule specification to check that the date stamp on each inventory record matches the current date.

Rules and guidelines for dimensions

Consider the following rules and guidelines when you add a dimension to an asset:

- You select a dimension as an optional step when you configure an asset. By default, an asset does not specify a dimension.
- An asset may contain steps that examine data in multiple dimensions. Select the dimension that best describes the purpose of the asset.
- The dimension that you set is a metadata value. It does not affect the asset logic or the analysis that the asset performs.
- The list of issues on the **Dimensions** menu does not represent a complete range of the data quality issues that you can investigate. Likewise, the meaning and impact of the dimensions may be different in your organization. Select the dimensions that are right for your data.

Data Integration

Use Data Integration to load, transform, and integrate your data. You can create data integration projects and run data integration tasks.

Data Integration is an on-demand subscription service that provides a complete platform for cloud data integration and management. You can use Data Integration to build and run advanced, complex integrations using pre-built functions and templates for mappings and transformations. Data Integration has the features you need to create a variety of data integration projects to fit your business needs.

Data Integration tasks

A Data Integration task is a process that you configure to analyze, extract, transform, and load data. You can run individual tasks manually or set tasks to run on a schedule.

For example, you might schedule a mapping task to run a mapping that includes a Rule Specification transformation.

You can use multiple types of task to integrate data in Data Integration. For more information, see the Data Integration online help.

Note: Data Integration tasks that contain Cleanse, Deduplicate, Parse, Labeler, Rule Specification, or Verifier transformations are not certified for serverless runtime execution. If you run a Data Integration task that contains a data quality asset in a serverless runtime environment, the task can count towards your serverless compute hour usage and quota.

Data Profiling

Use Data Profiling to create and run data profiling tasks. In Data Profiling, data profiling tasks are also called profiles.

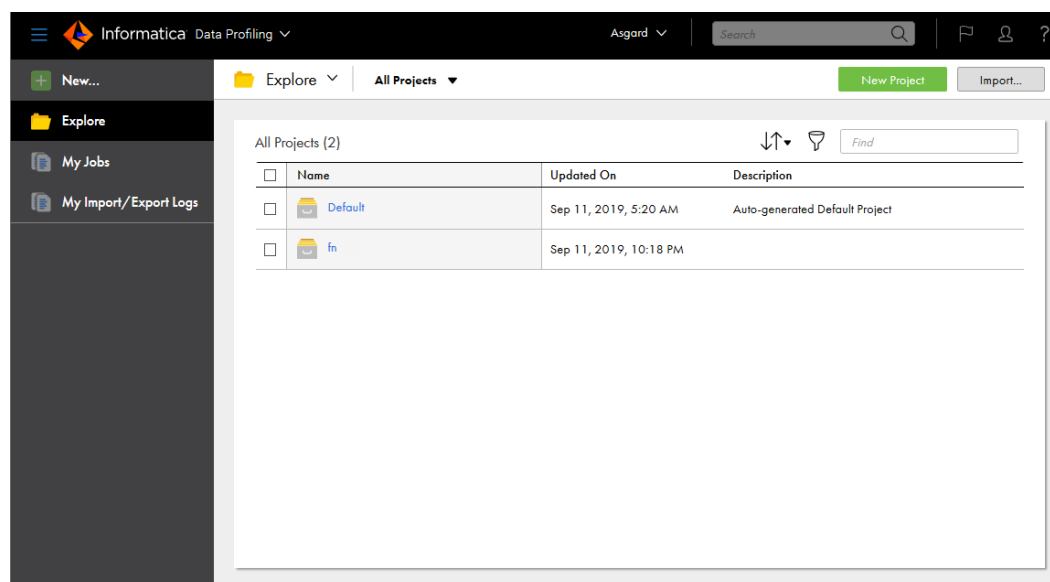
When you run a profile on a source object, the results include the following column statistics:

- Number of distinct, non-distinct, and null values

- Percentage of distinct, non-distinct, and null values
- Documented and inferred data types
- Number of patterns
- Percentage of top pattern
- Maximum and minimum length of values
- Maximum and minimum values
- Average, sum, and standard deviation for numeric data types
- Value frequencies
- Outliers

You can run data profiling tasks on relational data sources and flat file data sources.

When you select Data Profiling from the **My Services** page, the **Explore** page appears as shown in the following image:



You can access the following pages from the navigation bar for Data Profiling:

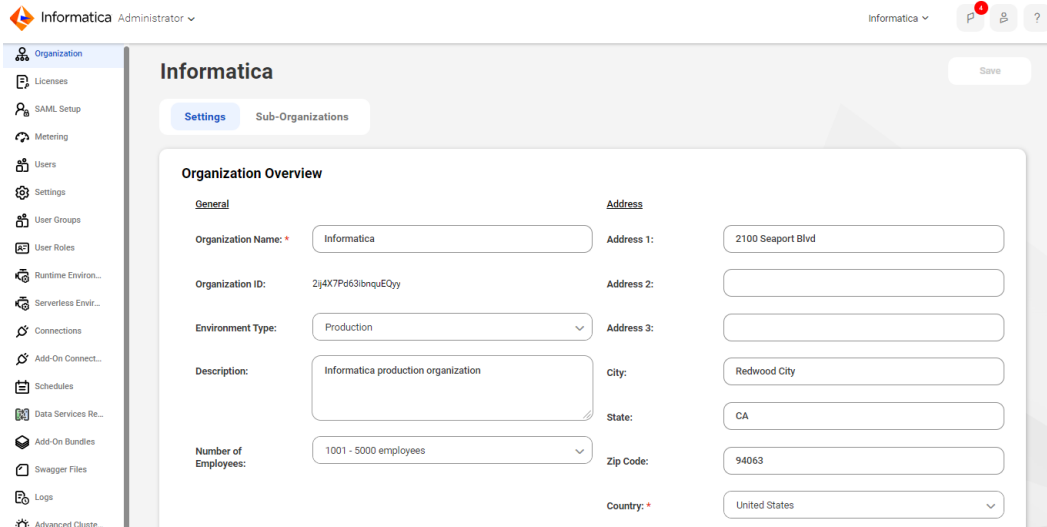
- New. Create data profiling tasks.
- Explore. Create and manage projects, folders, and assets.
- My Jobs. View the status of your data profiling and exception jobs.
- My Import/Export Logs. View the status of your imports and exports.

Note: You can open a profiling task and an exception task from the Explore page in Data Quality.

Administrator

Administrative tasks for Informatica Intelligent Cloud Services are performed in the Administrator service.

When you select Administrator from the **My Services** page, the navigation bar provides options for Informatica Intelligent Cloud Services administration, as shown in the following image:



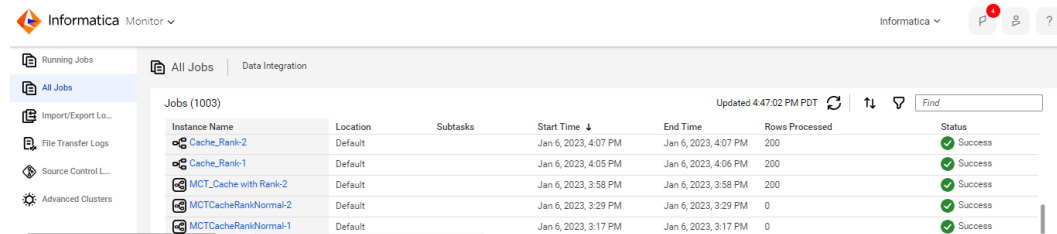
The following options are available based on your licenses:

- Organization. Manage your organization details and your sub-organizations.
- Licenses. View licenses for your organization and manage your sub-organization licenses.
- Ecosystem single sign-on. When you create your organization through Microsoft Azure, configure some single sign-on properties for Azure users.
- SAML Setup. Configure SAML single sign-on settings.
- Settings. Configure source control settings and upgrade settings for Secure Agent services.
- Users. Add new users. Configure user groups and roles for individual users.
- User Groups. Create user groups. Assign users to user groups.
- User Roles. Manage user roles for your organization. Configure the privileges associated with each role.
- Runtime Environments. Configure Secure Agents and Secure Agent groups.
- Connections. Create and manage connections.
- Add-On Connectors. Install add-on connectors.
- Schedules. Create and manage schedules for tasks.
- Add-On Bundles. Install, uninstall, upgrade, and copy bundles for your organization.
- Swagger Files. Generate swagger files for REST V2 connections.
- Logs. View asset and security logs.
- File servers. Configure file transfer servers and file transfer users to allow remote partners to send files to your organization.

Monitor

Use Monitor to monitor jobs, imports, and exports in your organization. A job is an instance of a mapping, task, or taskflow.

When you select Monitor from the **My Services** page, the navigation bar provides options for monitoring activity, as shown in the following image:



The navigation bar provides the following options:

- **Running Jobs.** Provides run-time details about the Data Integration jobs that are running or have completed within the last five minutes.
- **All Jobs.** Provides details about all Data Integration jobs in the organization.
- **Data Ingestion.** Provides details about the Data Ingestion and Replication jobs in the organization.
- **Import/Export Logs.** Provides details about imports and exports that are running and that have completed.
- **File Transfer Logs.** Provides details about the file transfers in the organization.
- **Source Control Logs.** Provides a log of actions on source-controlled objects in the organization.

To view details about a specific job, import instance, or export instance, click the instance name.

Data Observability

Use data observability to identify anomalies in the characteristics of the data in your source system.

Data observability analyzes a sudden deviation from a usual trend within a source system and generates events for the anomalies detected within your source system data. To use data observability, you can enable the data observability capability for your source system in Metadata Command Center. Then use Data Governance and Catalog to visualize and evaluate the events generated for the data observability anomalies identified.

For more information about enabling data observability, see the *Administration help* module for *Metadata Command Center*.

For more information about visualizing and evaluating data anomalies, see the *Introduction and Getting Started* and *Working with Assets* help modules for *Data Governance and Catalog*.

Informatica Intelligent Cloud Services security

Informatica Intelligent Cloud Services uses authentication and encryption to ensure that data is secure and available only to users within the organization. Informatica Intelligent Cloud Services uses industry approved

algorithms to encrypt sensitive information. All customer data is encrypted at rest using an AES-256 key. Data in transit is encrypted using the TLS 1.2 protocol.

For more information, see the [Informatica Intelligent Cloud Services Trust Center](#).

Learning resources for Data Quality

You can explore tutorials designed to provide you with hands-on experience for an end-to-end data quality project in Informatica Intelligent Cloud Services.

Check out the following tutorials to help you on your data quality learning journey:

- [Introduction to Data Quality](#)
- [Understanding your source data](#)
- [Standardizing data values](#)
- [Verifying postal addresses](#)
- [Labeling data values](#)

CHAPTER 2

Data Quality tools

Informatica Intelligent Cloud Services provides the following tools that you can use to satisfy your data integration business needs:

- Projects
- Assets
- Explore page

Projects

Informatica Intelligent Cloud Services assets are organized in projects. You can create multiple folders within a project to contain the assets that you need for your project.

For example, you might have a business need to provide monthly analyses of your company's sales activity. You create a project called Monthly Sales Analysis. You create folders that contain the assets that you use on a monthly basis to integrate data from various systems to perform your analysis.

The types of assets that you use in your projects depends on the services that you use. For example, a Data Integration project can include assets such as mappings, tasks, and taskflows.

Assets

In Data Quality, an asset is an object that defines part or all of a data quality operation.

In Data Integration, assets are mappings, taskflows, and tasks such as mapping tasks, synchronization tasks, and replication tasks. Data Integration assets also include components such as business services, mapplets, and hierarchical schemas.

You can create the following assets in Data Quality: cleanse, deduplicate, parse, labeler, rule specification, and verifier. Add the assets to a Cleanse transformation, Deduplicate transformation, Parse transformation, Labeler transformation, Rule Specification transformation, and Verifier transformation respectively when you design a mapping in Data Integration. You can add each transformation to a non-elastic mapping. You can add a Cleanse transformation, Parse transformation, Rule Specification transformation, and Verifier transformation to an elastic mapping.

You can also create dictionaries in Data Quality that other data quality assets can use. A dictionary is a table of reference data that a mapping can read to evaluate the accuracy or usefulness of source data.

Create assets to satisfy your data objectives. Templates are available for certain asset types. You can use a template as-is, or you can use it as a head start and customize it to meet your needs.

You can use assets in multiple projects. All assets must be part of at least one project. If you create an asset without specifying a project, the asset is created in the default project.

Explore page

Use the **Explore** page to work with your Data Quality projects and assets. The **Explore** page shows the assets that you can work with in Data Quality.

Finding projects and assets on the Explore page

Use any of the following methods to find your projects and assets on the **Explore** page:

- Explore by projects and folders. View all projects or select a particular project.
- Explore by asset types. View all assets or view assets of a particular type.
- Explore by tags. View assets associated with a particular tag.
- Search for projects, folders, or assets. To search for projects, folders, and assets in the organization, view the **Explore** page by **All Projects**, and then enter a name or description in the Find box. Or, view the **Explore** page by **Asset Types** and to narrow your search, select an asset type from the **All Assets** list. Then, in the Find box, enter a name or description in full or part.
- Sort the search results. Sort the **Explore** page by name, asset type, last update date, create date, or description. When you sort by type, the **Explore** page groups assets by asset type. It does not list the asset types in alphabetical order.

- Filter the objects on the page. To filter objects, click the **Filter** icon. To apply a filter, click **Add Filter**, select the property to filter by, and then enter the property value. The filters that are available depend on how you view the page. You can specify multiple filters.

For example, to find all the assets in your organization with the tag SW Region, view the **Explore** page by **All Assets** and then click the **Filter** icon. Add the **Tags** filter and enter "SW Region."

If the result size is large, for example, over 1000 objects, all the objects appear on the **Explore** page but the total number of objects that displays at the top and bottom of the page can be approximate for a minute or so.

Tip: Filtering is available on other pages in addition to the **Explore** page. For example, on the **Import Assets** page, you can filter by status to find the assets that imported successfully.

| <input type="checkbox"/> | Name ↑ | Type | Updated On | Location | Tags |
|--------------------------|----------------|--------------|-----------------------|---------------------|----------------------------|
| <input type="checkbox"/> | EC_AcctExps | Mapping | Oct 1, 2020, 2:34 PM | ECDistr_SC\EC_Accts | map, east, map, east, l... |
| <input type="checkbox"/> | MapLkUpSQLOR | Mapping | Mar 21, 2024, 4:51 PM | Default\Tag | east |
| <input type="checkbox"/> | MapLkUpSQLOR | Mapping | Mar 19, 2024, 5:05 PM | Default | east |
| <input type="checkbox"/> | MapLkUpSQLOR - | Mapping | Mar 21, 2024, 4:49 PM | Default | east |
| <input type="checkbox"/> | MappingTask41 | Mapping Task | Mar 21, 2024, 4:51 PM | Default\Tag | map, east |

You can see projects, folders, and assets for all of the services that you use.

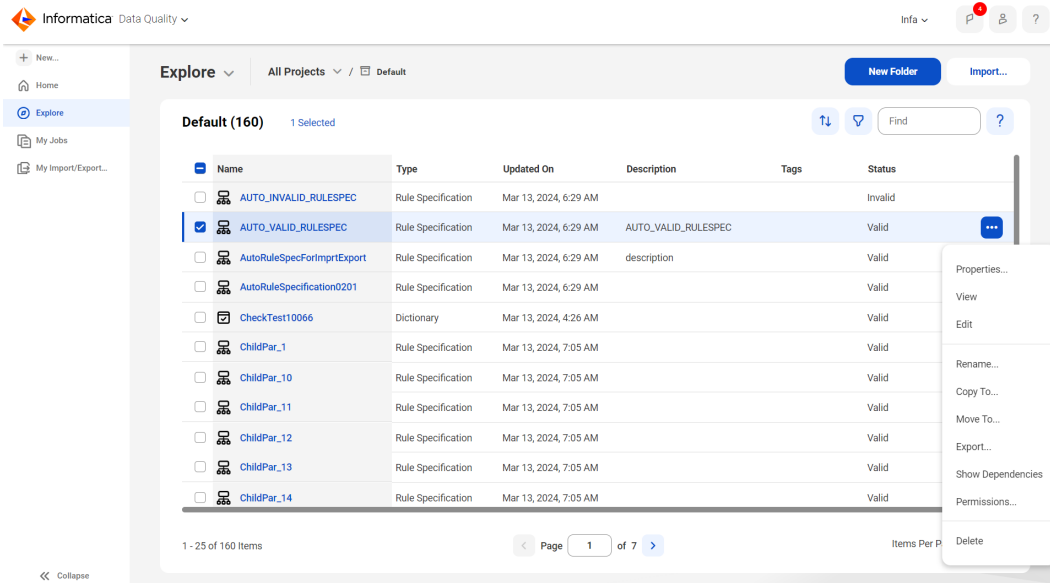
You can't use the following characters:

? ' | { } " ^ & [] / \

Working with projects and assets on the Explore page

You can perform actions on projects and assets on the Explore page. To see the actions that you can perform on a project or asset, click **Actions** in the row that contains the project or asset.

The following image shows the menu of actions that you can perform on an asset:



The **Actions** menu lists the actions you can perform based on your user role privileges and the permissions specified for the selected asset.

You can also perform an action on multiple objects at one time. Select the check box to the left of each object, or select the Select All check box to select all of the objects that are displayed on the current page. After you select the objects, right-click the objects to open a menu and choose the action

Customizing the Explore page

You can display, hide, or rearrange object properties on the Explore page. To display or hide properties, right-click the column heading area and check or uncheck the properties.

To rearrange columns, click a column heading and drag it to a different location.

CHAPTER 3

Informatica resources

In addition to the online help, you can find information about Informatica Intelligent Cloud Services using the following resources.

Informatica Intelligent Cloud Services web site

You can access the Informatica Intelligent Cloud Services web site at <http://www.informatica.com/cloud>.

Informatica Intelligent Cloud Services Communities

Use the Informatica Intelligent Cloud Services Community to discuss and resolve technical issues. You can also find technical tips, documentation updates, and answers to frequently asked questions.

Access the Informatica Intelligent Cloud Services Community at:

<https://network.informatica.com/community/informatica-network/products/cloud-integration>

Developers can learn more and share tips at the Cloud Developer community:

<https://network.informatica.com/community/informatica-network/products/cloud-integration/cloud-developers>

Informatica Intelligent Cloud Services Marketplace

Visit the Informatica Marketplace to try and buy Data Integration Connectors, templates, and mapplets:

<https://marketplace.informatica.com/>

Informatica Intelligent Cloud Services Trust Center

The Informatica Intelligent Cloud Services Trust Center provides information about Informatica security policies and real-time system availability.

You can access the trust center at <https://www.informatica.com/trust-center.html>.

Subscribe to the Informatica Intelligent Cloud Services Trust Center to receive upgrade, maintenance, and incident notifications. The [Informatica Intelligent Cloud Services Status](#) page displays the production status of all the Informatica cloud products. All maintenance updates are posted to this page, and during an outage, it will have the most current information. To ensure you are notified of updates and outages, you can subscribe to receive updates for a single component or all Informatica Intelligent Cloud Services components. Subscribing to all components is the best way to be certain you never miss an update.

To subscribe, on the [Informatica Intelligent Cloud Services Status](#) page, click **SUBSCRIBE TO UPDATES**. You can choose to receive notifications sent as emails, SMS text messages, webhooks, RSS feeds, or any combination of the four.

Informatica Product Availability Matrices

Product Availability Matrices (PAMs) indicate the versions of the operating systems, databases, and types of data sources and targets that a product release supports. You can browse the Informatica PAMs at <https://network.informatica.com/community/informatica-network/product-availability-matrices>.

Informatica Documentation

Use the Informatica Documentation Portal to explore an extensive library of documentation for current and recent product releases. To explore the Documentation Portal, visit <https://docs.informatica.com>.

If you have questions, comments, or ideas about the product documentation, contact the Informatica Documentation team at infa_documentation@informatica.com.

Data Integration connector documentation

You can access documentation for Data Integration Connectors at the Documentation Portal. To explore the Documentation Portal, visit <https://docs.informatica.com>.

Informatica Knowledge Base

Use the Informatica Knowledge Base to find product resources such as how-to articles, best practices, video tutorials, and answers to frequently asked questions.

To search the Knowledge Base, visit <https://search.informatica.com>. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team at KB_Feedback@informatica.com.

Informatica Global Customer Support

You can contact a Global Support Center through the Informatica Network or by telephone.

To find online support resources on the Informatica Network, click **Contact Support** in the Informatica Intelligent Cloud Services Help menu to go to the **Cloud Support** page. The **Cloud Support** page includes system status information and community discussions. Log in to Informatica Network and click **Need Help** to find additional resources and to contact Informatica Global Customer Support through email.

The telephone numbers for Informatica Global Customer Support are available from the Informatica web site at <https://www.informatica.com/services-and-training/support-services/contact-us.html>.

INDEX

A

assets [14](#), [15](#)

C

Cloud Application Integration community
URL [18](#)
Cloud Developer community
URL [18](#)

D

Data Integration community
URL [18](#)
Data Quality
overview [5](#)
data quality dimensions [7](#)

E

Explore page [15](#)

I

Informatica Global Customer Support
contact information [20](#)
Informatica Intelligent Cloud Services
web site [18](#)

M

maintenance outages [18](#)

P

project folders [14](#), [15](#)
projects [14](#), [15](#)

S

security [12](#)
status
Informatica Intelligent Cloud Services [18](#)
system status [18](#)

T

TLS
security [12](#)
trust site
description [18](#)

U

upgrade notifications [18](#)

W

web site [18](#)