Informatica® Cloud Data Quality
July 2024

Dictionaries

Informatica Cloud Data Quality Dictionaries
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

Informatica Resources 4

- Informatica Documentation 4
- Informatica Intelligent Cloud Services web site 4
- Informatica Intelligent Cloud Services Communities 4
- Informatica Intelligent Cloud Services Marketplace 4
- Informatica Knowledge Base 5
- Informatica Intelligent Cloud Services Trust Center 5
- Informatica Global Customer Support 5

Chapter 1: Introduction to dictionaries 6

Types of dictionary data 6

Dictionary operations 7

Dictionary configuration options 8

Creating a dictionary 9

Adding data to a dictionary 10

- Importing dictionary values from a file 10
- Rules and guidelines for importing data 11

Downloading dictionary data 11

Finding and filtering data values 12

Working with very large dictionaries 13

- Running a query on a large dictionary 14

Verifying the dictionary download location 15

Index 16

Preface

Refer to *Dictionaries* for information on how to create and manage dictionaries in Data Quality. A dictionary is a reference data object that you can use in a data quality asset to verify and enhance the accuracy and usability of your data.

Informatica Resources

Informatica provides you with a range of product resources through the Informatica Network and other online portals. Use the resources to get the most from your Informatica products and solutions and to learn from other Informatica users and subject matter experts.

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.

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 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 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 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.

CHAPTER 1

Introduction to dictionaries

A dictionary is a reference data set that you can use to evaluate data in a mapping. Use dictionaries to verify that the data values on a data source or another object in a mapping are accurate and correctly formatted.

You add a dictionary to an asset in Data Quality, and you add the asset to the corresponding transformation in a mapping in Data Integration. For example, you can configure a rule specification to read a dictionary, and you can add the rule specification to a Rule Specification transformation in a mapping. You can also add a dictionary to a cleanse asset and add the asset to a Cleanse transformation in a mapping.

When you run a mapping that contains the transformation, the mapping compares each input data value to the dictionary data and performs actions that you specify. For example, a rule specification might return a value that indicates a match between the input data value and the dictionary data. Or, a cleanse asset might return a preferred or correct version of the input value.

A dictionary can contain multiple columns. At least one column must contain the standard or preferred versions of a set of values that the organization uses. Other columns can contain alternative or related versions of the values, including versions that might appear in the source data. The column that contains the standard or preferred values is called the valid column. You can populate a dictionary with any combination of values that suits your organization and your project requirements.

Types of dictionary data

You can create dictionaries with any data that is useful to the data project that you work on. The dictionary does not evaluate the data that you enter, and the values in each column do not need to be objectively correct or incorrect.

Consider the following examples:

- The dictionary might include alternative versions of the same data, such as *Massachusetts* and *MA*.
- The dictionary might include values that relates to each other in a way that is relevant to your business, such as *Massachusetts* and *Boston*.
- The dictionary might include correct and incorrect spellings of the same data, such as *Massachusetts* and *Massachusets*.

Each row in a dictionary can have a different number of values. For example, the number of telephone area codes in each state of the United States varies from state to state. If you create a dictionary that matches the area codes to their respective states, the rows for each state will have different numbers of values.

The following table shows a fragment of a dictionary of United States telephone area codes:

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7
Mississippi	228	601	662	769		
Missouri	314	417	573	636	660	816
Montana	406					
Nebraska	308	402	531			
Nevada	702	725	775			
New Hampshire	603					

Dictionary operations

When you run a mapping with a transformation that specifies a dictionary, the transformation compares the input field data to the data in the dictionary. If the transformation finds a match between an input value and a dictionary value, the transformation performs an action that the asset in the transformation defines. For example, the transformation might return the value in the valid column that corresponds to the input value.

Valid column behavior

The transformation searches every column in the dictionary except the valid column that the asset specifies. If you want the search operation to include the valid column data, add a copy of the valid column to the dictionary. By default, the valid column is the first or left-most column in a dictionary. You can create a dictionary that contains two identical columns of data.

Note: A test that you run in an asset in Data Quality reads a dictionary in the same way as a transformation reads a dictionary.

You can select any column with values that suit your data objectives as the valid column. You might specify different columns as valid in different rule statements. The contents of the dictionary columns determine if the dictionary can support different valid columns in different business cases.

You can design a useful dictionary with multiple occurrences of a single value in the valid column. For example, you might design a dictionary that identifies a list of cities by time zone. Avoid designing a dictionary with multiple occurrences of a single value in non-valid columns.

Dictionary download considerations

The Secure Agent can download a dictionary that an asset specifies when you test the asset or run a mapping or profile that reads the asset.

The Secure Agent uses the following guidelines to download dictionary data:

- When you test a step in a cleanse asset that reads a dictionary, the Secure Agent downloads every dictionary that the cleanse asset specifies.

  When you test any other Data Quality asset, the Secure Agent downloads any dictionary that the test logic requires. If the asset specifies a dictionary that is not required during the test, the Secure Agent does not download the dictionary.

- When you run a profile or a mapping than references an asset that reads a dictionary, the Secure Agent downloads every dictionary that the asset specifies.

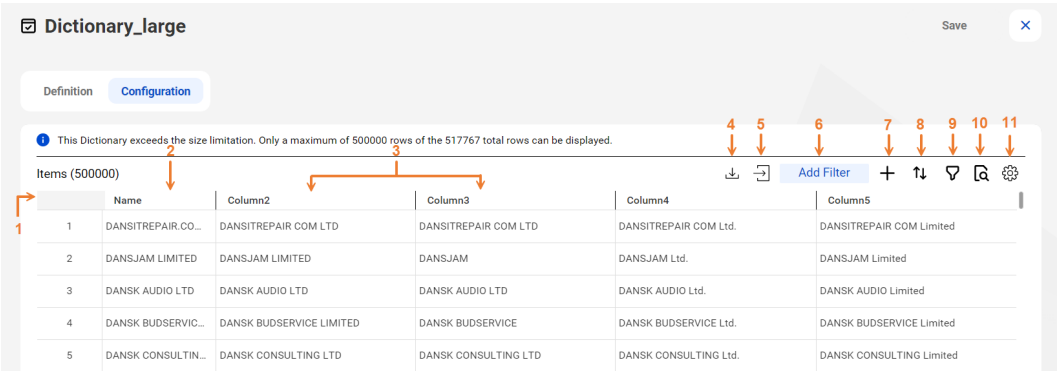
The Secure Agent downloads dictionaries to a directory on its host machine. You may decide to periodically check the download directory, for example to review the amount of disk space that the directory uses. For more information about dictionary downloads, see "Verifying the dictionary download location" on page 15.

Dictionary configuration options

You define and update the content and structure of a dictionary on the **Configuration** view.

You can select the **Configuration** view when you create or open a dictionary.

The following image shows the options that you can use to configure the dictionary:



The dictionary contains the following options and elements:

1. Row ID.
   Shows the current sequence of the dictionary rows. To highlight a row, select its row ID.
2. Default valid value column.
   Contains the values that the transformation returns at run time by default. You can specify a different valid column when you configure the asset.
3. Alternative value columns.
   Contains alternative or incorrect versions of the values in the default valid column.
4. Download data to a file.
   Downloads the data in a dictionary to a comma-separated file.
5. Import data from a file.
   Imports data to the dictionary.
6. Add a filter query.
   Opens the **Add Filter** query options.

   The **Add Filter** query option appears in a dictionary that contains very large quantities of data. For information about very large dictionaries, see "Working with very large dictionaries" on page 13.
7. Add a row.
   Adds an empty row after the current final row in the dictionary.
8. Sort the dictionary data.
   Sorts the data rows by values in one or more columns that you specify.
9. Filter the current page of dictionary data.

Displays the rows that contain a value that you specify in a column that you select. The option applies a filter on the current page in a dictionary.

10. Find and replace data.
    Finds and replaces a value that you specify.

    Note: The Filter and Find and Replace options are linked. When the filter is active, you can find and replace data within the filtered results only.

11. Settings menu.
    Determines the dictionary row height.

Creating a dictionary

When you click **New**, Data Quality prompts you to create asset. When you select the dictionary option, Data Quality opens a Dictionary window.

The **Dictionary** window contains a **Definition** tab and a **Configuration** tab. Use the Definition tab to define the name and the location of the dictionary. Use the Configuration tab to build the dictionary structure and add data.

Note: You can also open the Definition and Configuration tabs when you open the asset from the **Explore** page.

1. In the Definition tab, enter a name for the dictionary.
2. Optionally, enter a description.
3. Select a location for the dictionary.

   You can ignore the Asset References fields when you create a dictionary. A new object contains no asset references.
4. Save the dictionary.

   Data Quality replaces the Asset References fields with fields that include the creation date and the name of the asset creator.
5. Optionally, add a tag to the dictionary. You can search for assets with a common tag on the **Explore** page.
6. Select the Configuration tab.

   Data Quality displays the design workspace for the dictionary. The workspace is a grid of rows and columns.
7. Enter the dictionary data.

   Begin in the first row of column 1. The dictionary can have different numbers of values on each row.

   Column 1 is the default valid value column. Populate column 1 with the values that a transformation will return by default at run time.

   Note: You can design a dictionary in which more than one column might serve as a valid column. When you add a dictionary to an asset, you can select any column as the valid column.
8. Save the dictionary.

You can also populate the dictionary by "Importing dictionary values from a file" on page 10.

Adding data to a dictionary

When you create a dictionary, the **Configuration** view displays fifty empty rows by default.

You can enter data into any cell in any row. As best practice, enter data to the first cell in the first available row. After you add the data, save the dictionary.

You can add data to a dictionary in the following ways:

- Select an empty cell in the dictionary, and type data into the cell.

  If the dictionary does not contain an empty row, use the **Add Row** option to add rows to the dictionary. The rows appear at the bottom of the dictionary.
- Copy and paste data from a list or table in another application, such as a web page.

  You can paste the data to empty rows in the dictionary, or you can overwrite the current dictionary data.

  You can also copy and paste data within the dictionary and paste data from another dictionary.

  You do not need to add rows to accommodate data that you paste into a dictionary. The dictionary expands to accommodate the data that you add.
- Import data from a delimited file or a Microsoft Excel file.

  For more information on importing data, see "Importing dictionary values from a file" on page 10.

Updating data values

To update the value in a cell, select the cell and type a new value. You can select cells or rows in a sequence and delete the data that the cells or rows contain.

Use Windows shortcuts to copy, paste, and delete data.

Importing dictionary values from a file

You can import dictionary values from a Microsoft Excel file or any file that uses a delimiter that Data Quality recognizes. You can import the file data to an empty dictionary or to a dictionary that contains data. When you import the file data, Data Quality adds the data to the first empty row in the dictionary.

If you import data from a file with multiple columns, verify that the valid data column in the file matches the valid data column in the dictionary. The first column is the valid value column by default.

1. From the **Explore** page, select and open a dictionary.

   Or, click **New** and create a dictionary.

   If you create a dictionary, complete the fields on the Definition tab.
2. Select the Configuration tab.
3. Select the option to import data from a file.
4. In the **Import a flat file** dialog box, complete the following steps:
   - Choose the file that contains the data to import.
   - Select or clear the option to import column names. The option is cleared by default.
   - Select the line in the file from which to begin the data import. By default, Data Quality imports data from the first line in the file.
   - If you import data from a delimited file, select the delimiter. By default, Data Quality uses a comma delimiter.
   - If the file uses single quotes or double quotes as a text qualifier, select the text qualifier. By default, Data Quality does not recognize text qualifiers.

Note: A dictionary can contain a maximum of 42 columns. If you try to import a file with additional columns, Data Quality prompts you to reduce the number of columns in the file.

5. Click **Import**.

   Data Quality adds the file contents to the dictionary.

6. Save the dictionary.

   Note: When you import a very large dictionary, Data Quality may save the dictionary when the import operation is complete. A very large dictionary may contain up to 55 MB of data.

Rules and guidelines for importing data

Consider the following rules and guidelines when you import data:

- The import option supports CSV and Microsoft Excel files. The file must use the UTF-8 character encoding.
- You can import up to 55 MB of data from a file.
  To add very large quantities of data to a dictionary, use the file import mechanism. Do not cut and paste very large quantities of data into a dictionary.
- You can preview up to 100 lines of data before you import.
- If a dictionary contains more than 100,000 rows, Data Quality enables pagination and displays rows on multiple pages. Each page displays 100,000 rows.
- Data Quality automatically saves a dictionary when the quantity of data that you import exceeds the pagination threshold for dictionaries. For information about pagination in dictionaries, see "Working with very large dictionaries" on page 13.
- Data Quality can paginate and display a maximum of approximately 500,000 rows of dictionary data. The physical limit for the dictionary depends on the quantity of data that the dictionary contains. You can use the query options to find and work with any data that the dictionary stores beyond the pagination limit. The complete dictionary data set remains available to assets that you test in Data Quality and to transformations that contain Data Quality assets.

  The dictionary configuration pane displays the number of rows in the dictionary.

Downloading dictionary data

You can download the data in a dictionary to a comma-separated file.

1. From the **Explore** page, select and open a dictionary.
2. Select the **Configuration** tab.
3. Click the **Download** icon.

   By default, the name of the downloaded file is same as the dictionary name.

Note the following points when you download the dictionary data:

- The operation downloads the complete contents of a dictionary and disregards any filter that you apply to the dictionary before you download.
- The data might contain characters that spreadsheet applications can interpret as formulas.

Finding and filtering data values

Use the find and replace options to update one or more values in a dictionary. Filter the dictionary rows to review the data values that match your criteria and to ensure that you update only the values that you require.

Note: The find, replace, and filter options operate on the page of data that Data Quality currently displays for the dictionary. Additionally, you can filter the complete contents of a very large dictionary. The **Add Filter** option appears when you can filter data across the complete dictionary.

Finding and replacing data values

You can find and replace data values throughout a dictionary, or you can find and replace data values in one or more rows or columns that you select.

1. To display the find and replace options, click **Find**.
2. Optionally, select one or more rows or columns to restrict the find and replace operations.

   To select a row, click the row ID. To select a column, click the column name. Use the Control or Shift keys to select multiple rows or columns.
3. Enter a value in the Find field and click the search arrow.

   Data Quality highlights any matching dictionary value in yellow. Data Quality also selects the cell that contains the first instance of the value.
4. Perform one of the following actions:
   - To replace a single value, type in the cell that contains the value. Or, enter the new value in the Replace field and click **Replace**.

     When you click **Replace**, Data Quality updates the current cell and selects the next cell in the dictionary that contains the value.
   - To replace successive instances of the value, click **Replace** each time that Data Quality selects a cell.

     To skip a cell, or to navigate through the search results, use the Up and Down arrows beside the Find field.
   - To replace all instances of the value in the dictionary, click **Replace All**.

     Before you click **Replace All**, review the dictionary to ensure that you want to replace every instance of the value. The Find field displays the number of times that the value occurs in the dictionary.
5. After you replace the data values, save the dictionary.

Note: Data Quality finds and replaces the exact set of characters that you enter. If the characters occur within a longer string in a cell, Data Quality replaces the characters that you search for and leaves the other characters in the string.

If the characters appear more than once in a cell, Data Quality replaces all instances of the characters in the cell.

In cases where the characters appear more than once in a cell, Data Quality includes each instance in the results in the Find field.

Filtering data values

When you apply a filter to a dictionary, Data Quality displays the rows that match the filter that you set and hides any other row.

You can set a filter on each column in a dictionary. When you set multiple filters, Data Quality displays the rows that match all of the filters.

The filter options are hidden by default. To display the options, click **Filter**. When you click **Filter**, the **Add Field** options appear.

1. Click **Add Field** and select a column from the column list.
2. Enter a filter value in the **Value** field, and click **OK**.

   The dictionary displays the rows that contain the filter value in the column that you specify.

   To add another filter, repeat the steps.

When you add one or more filters to a dictionary, Data Quality restricts the find and replace operations to the data rows that meet the filter criteria.

Note: You can also select one or more cells, rows, or columns when you apply a filter. Data Quality performs find and replace operations within the cells, rows, or columns that you select.

Working with very large dictionaries

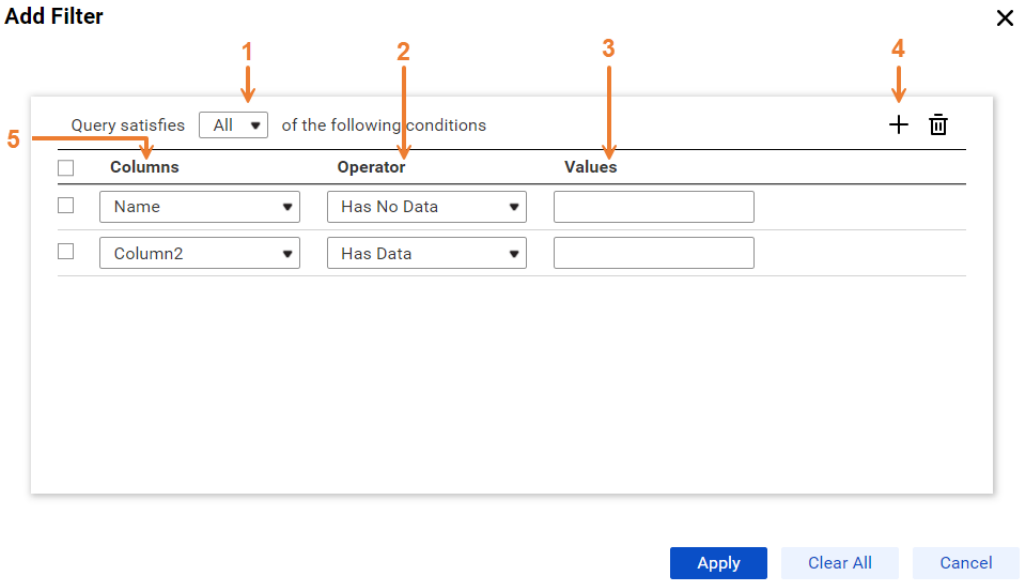
Data Quality can paginate up to 500,000 rows of dictionary data at a time, based on five columns of dictionary data. The physical limit for your dictionary depends on the number of rows and columns in the dictionary and the number of characters in each cell. The limit governs any operation to view or update data, such as a search and replace operation, that you perform in the dictionary.

When a dictionary contains more data than Data Quality can paginate in a single operation, you can use the filter query options to retrieve a subset of the dictionary data. Data Quality applies the filter to the complete dictionary, regardless of the quantity of data that the dictionary may contain. Use the **Add Filter** option to define the filter query.

A single filter query can return data up to the pagination limit. You can define multiple filters if necessary to cover the complete contents of your dictionary.

Note: The complete dictionary data set is available to assets that you test in Data Quality and to transformations that contain Data Quality assets.

The following image shows the options that you use to create a filter query:



The dialog box contains the following options and attributes:

1. Query scope option.
   Determines whether the dictionary data that the query retrieves must meet any condition that you add or all conditions that you add.
2. Operator attribute.

   Indicates the operator for each condition. The operator determines how the query analyzes the column data. You can select the following operations:

   - Equals
   - Not Equals
   - Contains
   - Not Contains
   - Has Data
   - Has No Data
3. Values attribute.

   Specifies the value that the condition looks for in a column when you select one of the following operators: Equals, Not Equals, Contains, or Not Contains.
4. Add option.

   Adds a condition to the query. You can add one or more conditions.
5. Columns attribute.

   Indicates the column to which each condition applies.

Running a query on a large dictionary

You can create a filter query to retrieve data from the full set of dictionary values. The filter options are available in very large dictionaries.

For information about very large dictionaries and the pagination limits that can apply to them, see "Working with very large dictionaries" on page 13.

You build a filter query on dictionary column values. For example, you can create a query that can retrieve rows with given values in a column that you select.

1. On the **Configuration** view, click **Add Filter**.
2. In the **Add Filter** dialog box, click the '+' symbol to add a condition.
3. Choose a column and an operator for the condition. If required, add a value that the query can look for in the column.
4. Add more conditions if required.
5. If you add multiple conditions, choose one of the following options:
   - All. Data Quality retrieves any row that meets all of the conditions.
   - Any. Data Quality retrieves any row that meets at least one of the conditions.
6. Click **Apply**.

Note: To edit a query, click **Edit Filter** in the **Configuration** pane.

Verifying the dictionary download location

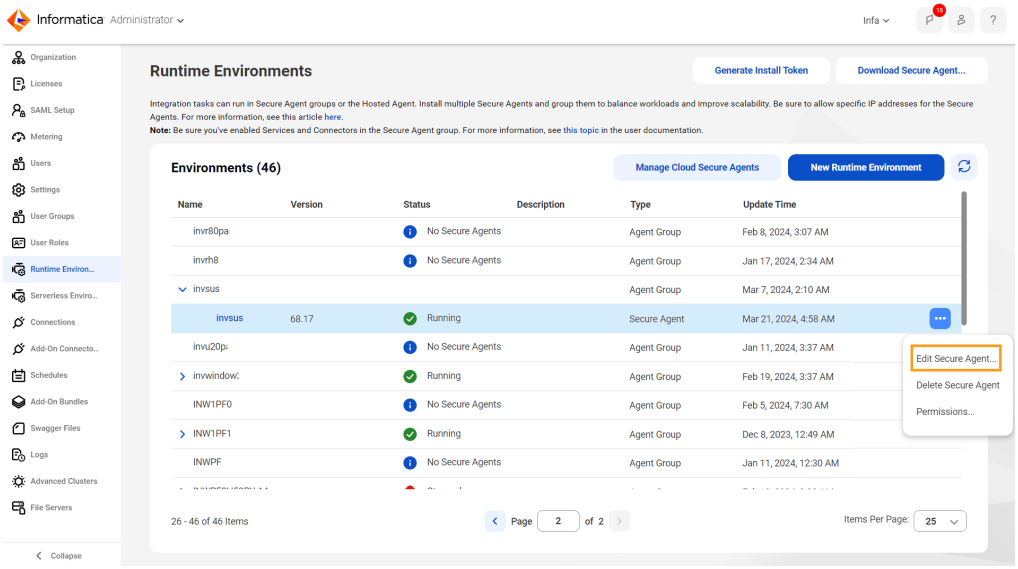
The Secure Agent can download a dictionary when you test an asset that specifies a dictionary or when you run a mapping or profile that reads a dictionary. You can verify or update the download location in the Administrator service.

If you plan to update the download location, confirm the update with an Admin user for your organization.

1. From the **My Services** page, select the **Administrator** service.
2. Choose the **Runtime Environments** option.
3. Select the **Secure Agent** that will test or run the assets, profiles, or mappings.

   Note: Each Secure Agent has a separate set of properties.
4. Hover over the **Actions** icon for the Secure Agent, and select the **Edit Secure Agent** option.

   The following image shows the option:



   The **Secure Agent** page appears.
5. Under **System Configuration Details**, select the following options:
   - Select **Data Integration Server** in the **Service** field.
   - Select **IDQ** in the **Type** field.

   The **System Configuration Details** pane returns a list of properties based on the type that you specified.
6. Find the *ReferenceDataLocation* property in the **Name** column.

   The corresponding entry in the **Value** column shows the current download directory.

For more information about dictionary downloads, see "Dictionary operations" on page 7.

INDEX

C

Cloud Application Integration community
  URL 4
Cloud Developer community
  URL 4

D

Data Integration community
  URL 4
dictionaries
  adding data 10
  configuration options 8
  creating a dictionary 9
  download considerations 7
  filtering data 12
  finding and replacing data 12
  introduction 6
  valid column 7
  working with large dictionaries 13

I

Informatica Global Customer Support
  contact information 5
Informatica Intelligent Cloud Services
  web site 4

M

maintenance outages 5

S

status
  Informatica Intelligent Cloud Services 5
system status 5

T

trust site
  description 5

U

upgrade notifications 5

W

web site 4