



Informatica™

Informatica® Cloud Data Quality
Summer 2019 July

Dictionaryes

Informatica Cloud Data Quality Dictionaries
Summer 2019 July
July 2019

© Copyright Informatica LLC 2019

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, Big Data Management, 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. Required third party notices are included with the product.

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: 2019-08-28

Table of Contents

- Preface 4**
- Informatica Resources. 4
 - Informatica Network. 4
 - Informatica Knowledge Base. 4
 - Informatica Documentation. 4
 - Informatica Product Availability Matrices. 5
 - Informatica Velocity. 5
 - Informatica Marketplace. 5
 - Informatica Global Customer Support. 5

- Chapter 1: Introduction to dictionaries..... 6**
- Dictionary operations. 6
- Dictionary configuration options. 7
- Creating a dictionary. 8
- Adding data to a dictionary. 9
 - Importing dictionary values from a file. 10
- Finding and filtering data values. 11

- Index..... 13**

Preface

The *Informatica Dictionaries Guide* contains information on how to create and manage dictionaries in Informatica Cloud. A dictionary is a reference data object that you can use to a rule specification 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 Network

The Informatica Network is the gateway to many resources, including the Informatica Knowledge Base and Informatica Global Customer Support. To enter the Informatica Network, visit <https://network.informatica.com>.

As an Informatica Network member, you have the following options:

- Search the Knowledge Base for product resources.
- View product availability information.
- Create and review your support cases.
- Find your local Informatica User Group Network and collaborate with your peers.

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 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 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 Velocity

Informatica Velocity is a collection of tips and best practices developed by Informatica Professional Services and based on real-world experiences from hundreds of data management projects. Informatica Velocity represents the collective knowledge of Informatica consultants who work with organizations around the world to plan, develop, deploy, and maintain successful data management solutions.

You can find Informatica Velocity resources at <http://velocity.informatica.com>. If you have questions, comments, or ideas about Informatica Velocity, contact Informatica Professional Services at ips@informatica.com.

Informatica Marketplace

The Informatica Marketplace is a forum where you can find solutions that extend and enhance your Informatica implementations. Leverage any of the hundreds of solutions from Informatica developers and partners on the Marketplace to improve your productivity and speed up time to implementation on your projects. You can find the Informatica Marketplace at <https://marketplace.informatica.com>.

Informatica Global Customer Support

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

To find your local Informatica Global Customer Support telephone number, visit the Informatica website at the following link:

<https://www.informatica.com/services-and-training/customer-success-services/contact-us.html>.

To find online support resources on the Informatica Network, visit <https://network.informatica.com> and select the eSupport option.

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 a rule specification in Data Quality, and you add the rule specification to a Rule Specification transformation in Data Integration.

When you run a mapping that contains the Rule Specification transformation, the rule specification compares each input data value to the dictionary data and performs actions that you specify. The rule specification might return a value that indicates a match between the input data value and the dictionary data. Or, the rule specification might return a preferred or corrected version of the input value.

A dictionary contains one or more columns. At least one column must contain a set of standard or required values as defined by a business rule. 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 correct or required values is called the valid column. You can populate a dictionary with any combination of values that suits your business rule logic.

Dictionary operations

When you run a mapping with a Rule Specification 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 rule specification defines.

You add the dictionary to a rule specification that the Rule Specification transformation reads. You can select a dictionary when you configure a condition or an action in a rule statement in the rule specification. You can add a dictionary to any condition or action in a rule specification. A rule specification can read one or more dictionaries.

When you add a dictionary to a condition, Data Quality compares the input data to the values in each column in the dictionary and passes the results to the action in the rule statement.

When you add a dictionary to an action, Data Quality compares the input data to the values in each column in the dictionary and returns a value according to the operation that you select in the **Define Action** dialog box. For example, the action might return the value in the valid column on the dictionary row that contains the input value.

By default, the valid column is the first or left-most column in a dictionary. In a single-column dictionary, all data is valid data. When you add a dictionary to an action in a rule statement, you can select any column as the valid column. Select the column that suits the acceptance criteria for the action. 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 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 the dictionary:

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					

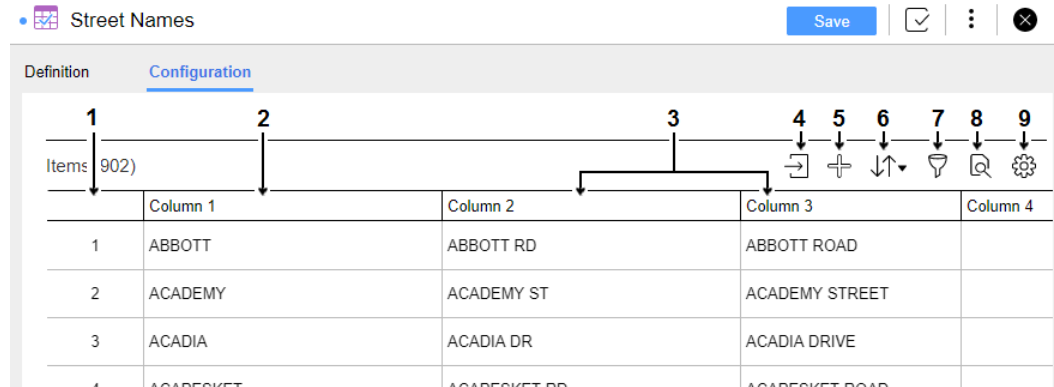
You can design a useful dictionary with multiple occurrences of a single value in the valid column. For example, you might design a dictionary that identifies a list of cities by time zone. Avoid designing a dictionary with multiple occurrences of a single value in non-valid columns.

Dictionary 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 Rule Specification transformation returns at run time by default. You can select a different valid column when you add the dictionary to a rule specification.
3. Alternative value columns.
Contains alternative or incorrect versions of the valid values.
4. Import data from a file.
Imports data to the dictionary.
5. Add a row.
Adds an empty row after the current final row in the dictionary.
6. Sort the dictionary data.
Sorts the data rows by values in one or more columns that you specify.
7. Filter the dictionary data.
Displays the rows that contain a value that you specify in a column that you select. Hides other rows.
8. Find and replace.
Finds and replaces a value that you specify.
9. **Note:** The Filter and Find options are linked. When the filter is active, you can find and replace data within the filtered results only.
10. Settings menu.
Determines the dictionary row height.

Creating a dictionary

When you click **New**, Data Quality prompts you to create a rule specification or a dictionary. 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 identify 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 the rule specification can return 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 a rule specification, you can select any column as the valid column within the rule specification.

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.

Finding and filtering data values

Use the find and replace options to update a single dictionary value or multiple 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.

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.

INDEX

D

dictionaries

- adding data [9](#)
- configuration options [7](#)
- creating a dictionary [8](#)

dictionaries (*continued*)

- filtering data [11](#)
- finding and replacing data [11](#)
- introduction [6](#)
- valid column [6](#)