



Informatica® Data Quality for Microsoft
Excel

1.2

Data Quality for Microsoft Excel Guide

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, the Informatica logo, and PowerCenter 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.

See patents at <https://www.informatica.com/legal/patents.html>.

DISCLAIMER: Informatica LLC provides this documentation "as is" without warranty of any kind, either express or implied, including, but not limited to, the implied warranties of noninfringement, merchantability, or use for a particular purpose. Informatica LLC does not warrant that this software or documentation is error free. The information provided in this software or documentation may include technical inaccuracies or typographical errors. The information in this software and documentation is subject to change at any time without notice.

NOTICES

This Informatica product (the "Software") includes certain drivers (the "DataDirect Drivers") from DataDirect Technologies, an operating company of Progress Software Corporation ("DataDirect") which are subject to the following terms and conditions:

1. THE DATADIRECT DRIVERS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT.
2. IN NO EVENT WILL DATADIRECT OR ITS THIRD PARTY SUPPLIERS BE LIABLE TO THE END-USER CUSTOMER FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, CONSEQUENTIAL OR OTHER DAMAGES ARISING OUT OF THE USE OF THE ODBC DRIVERS, WHETHER OR NOT INFORMED OF THE POSSIBILITIES OF DAMAGES IN ADVANCE. THESE LIMITATIONS APPLY TO ALL CAUSES OF ACTION, INCLUDING, WITHOUT LIMITATION, BREACH OF CONTRACT, BREACH OF WARRANTY, NEGLIGENCE, STRICT LIABILITY, MISREPRESENTATION AND OTHER TORTS.

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: 2022-05-15

Table of Contents

Preface	5
Informatica Resources.	5
Informatica Network.	5
Informatica Knowledge Base.	5
Informatica Documentation.	5
Informatica Product Availability Matrices.	6
Informatica Velocity.	6
Informatica Marketplace.	6
Informatica Global Customer Support.	6
 Chapter 1: Overview.....	 7
Data Quality for Microsoft Excel Overview.	7
Data Quality Service Process.	7
 Chapter 2: Installation.....	 9
Prerequisites for Microsoft Excel Users.	9
Prerequisites for Data Quality Developers.	9
Installing Data Quality for Microsoft Excel.	10
 Chapter 3: Data Quality Services for Excel Users.....	 11
Data Quality Services for Excel Users Overview.	11
Data Quality for Microsoft Excel Ribbon.	11
Add and Remove Services.	12
Run a Service.	12
Service Window Properties.	12
Keyboard Shortcuts in the Service Window.	13
Service Inputs and Outputs.	14
Running a Data Quality Service.	17
Data Quality for Microsoft Excel Configuration.	17
Export and Import Settings.	18
Settings Window Properties.	18
Service Review.	19
Review Window Properties.	19
Troubleshooting.	19
 Chapter 4: Data Quality Service Example.....	 21
Data Quality Service Example Overview.	21
Input Data.	22
Using the Data Quality Service.	22
Output Data.	23

Chapter 5: Data Quality Services for Developers.....	24
Data Quality Services for Developers Overview.	24
Deploying Data Quality for Excel in an Organization.	25
Data Quality Service Requirements.	25
Data Quality Services Optimization.	25
Adding a Data Quality Service to a Data Integration Service.	26
Adding a Data Quality Service to a Web Services Hub.	26
Provide a Data Quality Service to Excel Users.	26
Copying Data Service Web Service URLs for Microsoft Excel Users.	27
Copying PowerCenter Web Services Hub URLs for Microsoft Excel Users.	27
Saving WSDL files for Data Quality for Microsoft Excel.	27
Index.	28

Preface

Read the *Data Quality for Microsoft Excel Guide* to learn how you can apply data quality rules to data in Microsoft Excel worksheets. The rules perform data quality operations such as parsing, cleansing, and standardization. To apply the rules, you run data quality services through Data Quality for Microsoft Excel.

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

Overview

This chapter includes the following topics:

- [Data Quality for Microsoft Excel Overview, 7](#)
- [Data Quality Service Process, 7](#)

Data Quality for Microsoft Excel Overview

Data Quality for Microsoft Excel is an Excel plug-in that applies data quality rules to the data in Excel worksheets.

As an Excel user, you apply the data quality rules by running data quality services in Data Quality for Microsoft Excel. A data quality service is a web service that applies business and data quality logic to data. You can run data quality services to perform data quality operations such as parsing, cleansing, and standardization.

You request data quality services from a developer. The developer creates the services and provides you with URLs or files that you access to run the services.

For example, consider an Excel worksheet that contains customer records. You need to standardize the United States state names in the records to the corresponding two-letter state abbreviations. You can use Data Quality for Microsoft Excel to run a web service that processes the state names and writes standardized output to another column or worksheet.

Data Quality Service Process

A data quality service applies business logic to data contained in Microsoft Excel worksheets. Excel users and data quality developers work together to define the business logic for data quality services.

The following steps describe the process of creating and using data quality services:

1. An Excel user identifies data quality requirements including information about input data, the types of data quality operations required, and the expected output.
2. The Excel user provides the data quality requirements to a developer.
3. The developer creates the data quality service and provides the Excel user with a URL or file path for the web service.
4. The Excel user adds the URL or file path to Data Quality for Microsoft Excel to store a link to the service.

5. The Excel user selects Microsoft Excel data to use as input for the data quality service and sets the service output options.
6. The Excel user runs the service.
7. The Excel user reviews the service results.

CHAPTER 2

Installation

This chapter includes the following topics:

- [Prerequisites for Microsoft Excel Users, 9](#)
- [Prerequisites for Data Quality Developers, 9](#)
- [Installing Data Quality for Microsoft Excel, 10](#)

Prerequisites for Microsoft Excel Users

Microsoft Excel users must install Excel version 2007 or later.

To run the Data Quality for Microsoft Excel plug-in, Excel requires the following components:

- Windows Installer 3.1
- .Net Framework 3.5
- Visual Studio Tools for the Microsoft Office system (version 3.0 Runtime Service Pack 1)

If the installer does not detect these components, it downloads them from Microsoft websites. The computer must have Internet access for the installer to download the components.

Prerequisites for Data Quality Developers

To create data quality services that Data Quality for Excel can run, data quality developers must install Informatica Services and Informatica Developer version 9.1.0 or later.

The Informatica Services license must include one of the following options:

- Data Quality Web Services Option 9.1 or later
- PowerCenter® Web Services Provider 8.6.1 or later

Installing Data Quality for Microsoft Excel

Run the installer executable and restart Excel to install Data Quality for Microsoft Excel.

To install Data Quality for Microsoft Excel, perform the following steps:

1. Verify that the Microsoft Excel application is not running.
2. Download and unzip the Data Quality for Microsoft Excel installer.
3. Run `Setup.exe`.
4. Open Excel.
5. Verify that an Informatica menu item is available.

CHAPTER 3

Data Quality Services for Excel Users

This chapter includes the following topics:

- [Data Quality Services for Excel Users Overview, 11](#)
- [Add and Remove Services, 12](#)
- [Run a Service, 12](#)
- [Data Quality for Microsoft Excel Configuration, 17](#)
- [Service Review, 19](#)
- [Troubleshooting, 19](#)

Data Quality Services for Excel Users Overview

You can use Data Quality for Microsoft Excel to add, remove, configure, and run data quality services.

Data quality developers create data quality services. You provide a data quality developer with the business logic that data quality services use. The developer creates data quality services and provides you with URLs or files for the services.

When you add a data quality service, you use the URL or file path provided by a developer to store a link to the service. When you run a data quality service, you select Microsoft Excel data to use as input for the data quality service and set the service output options.

Data Quality for Microsoft Excel Ribbon

The Data Quality for Microsoft Excel installer adds an Informatica menu item to Excel. When you select this menu item, the Informatica ribbon appears.

The following table lists the ribbon commands and the actions you can perform with the commands:

Command	Action
Use	Run a data quality service.
Add	Add a data quality service.

Command	Action
Remove	Remove data quality services.
Settings	Define settings such as default WSDL directory, the number of Excel inputs to send in one batch, service run-time options. You can also import or export Data Quality for Microsoft Excel properties.
Review	Review the results of the last data quality service that you ran.
Help	View online help.

Add and Remove Services

You can add links to data quality services to Data Quality for Microsoft Excel. You access these links in the **Use Service** window. Contact a data quality developer to get a URL or access to a WSDL file.

You can add data quality services in the following ways:

- You can use the **Add** command in the Informatica ribbon to add a URL or multiple WSDL files.
- In the **Settings** window, you can add all the WSDL files in a directory that you specify.

You can also remove a data quality service from the list of available services.

Run a Service

When you run a service, you choose the service to run against the data in the current worksheet. Choose to use columns or rows for inputs or outputs. You can also choose to create a worksheet for output data.

Service Window Properties

You can run services and configure service properties in the **Use Service** window.

You can configure the following properties:

Available services

Provides a list of the data quality services that you added to Data Quality for Microsoft Excel.

Put results on new sheet

Create an Excel worksheet for the output of the data quality service. You can configure a default value for this property in the **Settings** window.

Close window when finished

Closes the **Use Service** window when the data quality service finishes processing input data. You can configure a default value for this property in the **Settings** window.

Run

Runs the data quality service.

Undo

Removes the results of the data quality service operation from the Microsoft Excel worksheet.

Column input format

Reads data from input fields using a column orientation, reading from top to bottom.

Row input format

Reads data from input fields using a row orientation, reading from left to right.

Autofill input fields

Fills the input fields using the fields, rows, and columns selected in the Microsoft Excel worksheet. You can fill multiple input fields by selecting cell ranges with more than one column or row, or by selecting multiple areas in the worksheet.

Clear input fields

Resets all input fields.

Column output format

Writes data in output fields using a column orientation, writing from top to bottom.

Row output format

Writes data in output fields using a row orientation, writing from left to right.

Autofill output fields

Fills the output fields using the fields, rows, and columns selected in the Microsoft Excel worksheet. You can fill multiple output fields by selecting cell ranges with more than one column or row, or by selecting multiple areas in the worksheet.

Clear output fields

Resets all output fields.

Keyboard Shortcuts in the Service Window

You can use keyboard shortcuts to populate fields and run services in the **Use Service** window.

The following table describes the keyboard shortcuts:

Keyboard Shortcut	Description
ALT + A	When any field in the Input section has keyboard focus, this shortcut fills all input fields using the cell ranges highlighted in the Excel worksheet. When any field in the Output section has keyboard focus, this shortcut fills all output fields using the cell ranges highlighted in the Excel worksheet.
ALT + S	Populates the currently selected input or output field with the range selected in the Excel worksheet.
ALT + C	Clears all input or output fields.
ALT + D	Swaps the value in the currently selected field with the value in the next field.
ALT + U	Swaps the value in the currently selected field with the value in the preceding field.
ENTER	Runs the service selected in the Available Service field.

Service Inputs and Outputs

You can enter keyboard shortcuts or click an **Autofill** button to populate service inputs and outputs with contiguous or non-contiguous cell ranges. You use the **CTRL** key to select non-contiguous areas in Microsoft Excel worksheets.

After you select ranges, enter a keyboard shortcut or click an **Autofill** button in the Input section or Output section in the **Use Service** window.

If a service uses multiple inputs or output fields, you must populate the fields with ranges that contain the same number of cells. For example, if you populate two input fields with the ranges A1:A100 and B1:B99 and click **Run**, Data Quality for Excel returns the error "The number of cells in all inputs must be the same."

Selection Order

The Autofill function fills service fields based on the order that you select ranges in the worksheet.

The following figure shows the correlation between the selection order of Excel ranges and input fields in the **Use Service** window:

The figure shows an Excel worksheet and the 'Use Service' window. The Excel worksheet has columns A through K. Rows 1 through 10 are highlighted in blue. The data in the worksheet is as follows:

	A	B	C	D	E	F	G	H	I	J	K
1	FullName		CustNumber		PhoneNumber						
2	John Doe 1		100001		123-456-781						
3	John Doe 2		100002		123-456-782						
4	John Doe 3		100003		123-456-783						
5	John Doe 4		100004		123-456-784						
6	John Doe 5		100005		123-456-785						
7	John Doe 6		100006		123-456-786						
8	John Doe 7		100007		123-456-787						
9	John Doe 8		100008		123-456-788						
10	John Doe 9		100009		123-456-789						

The 'Use Service' window is open, showing the 'DQ!StandardizeCustomerData' service. The input fields are populated with the following ranges:

- FullName: Sheet1!A3:A8 (labeled 1)
- PhoneNumber: Sheet1!C3:C8 (labeled 2)
- CustomerNumber: Sheet1!E3:E8 (labeled 3)

The output fields are empty. The 'Run' button is highlighted. The 'Input Format' is set to 'column' and 'row'. The 'Output Format' is set to 'column' and 'row'. The 'Autofill' button is circled in red.

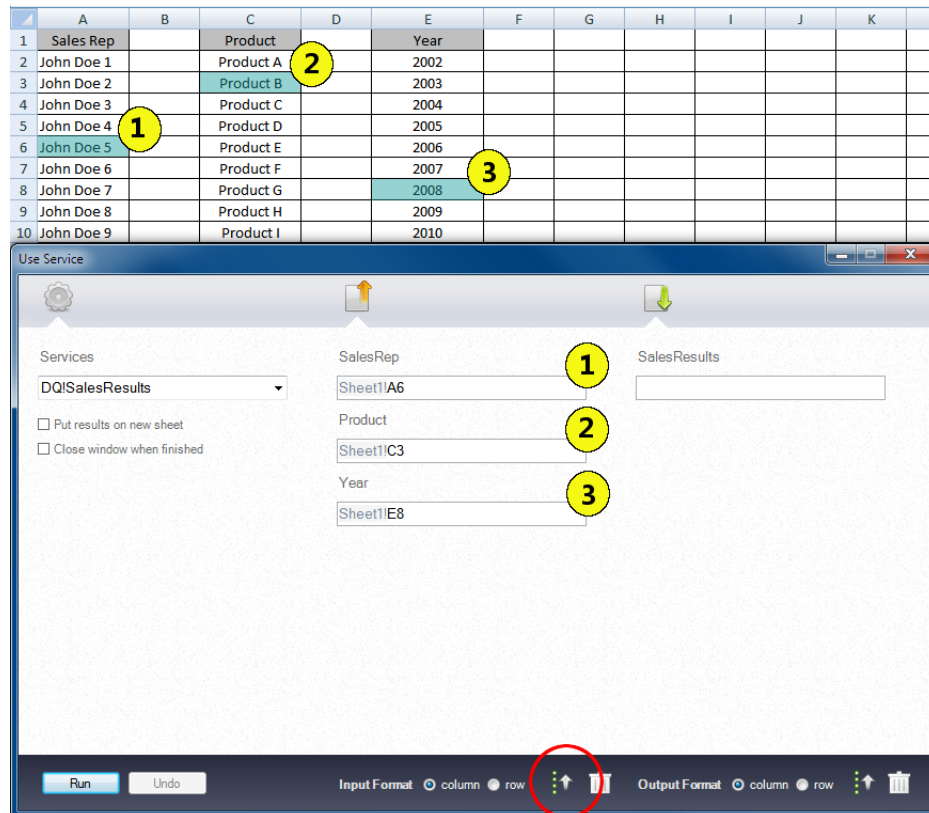
The highlighted blue areas in the Excel worksheet indicate the cell ranges that you select. The numbers in the Excel worksheet indicate the order that you select the ranges. The numbers in the **Use Service** window indicate the order in which the Data Quality for Excel plug-in populates the input fields.

If you click the **Autofill** button in the Input section of the **Use Service** window, the plug-in populates the input fields with the cell ranges indicated by the corresponding numbers. The **Autofill** button is indicated by the red circle in the figure.

Unaligned Selections

You can select unaligned cell ranges to populate multiple input and output fields. Selections are unaligned if the cell ranges do not use the same rows or columns for the start or end of the range.

The following figure shows the correlation between unaligned selections in a worksheet and output fields in the **Use Service** window:



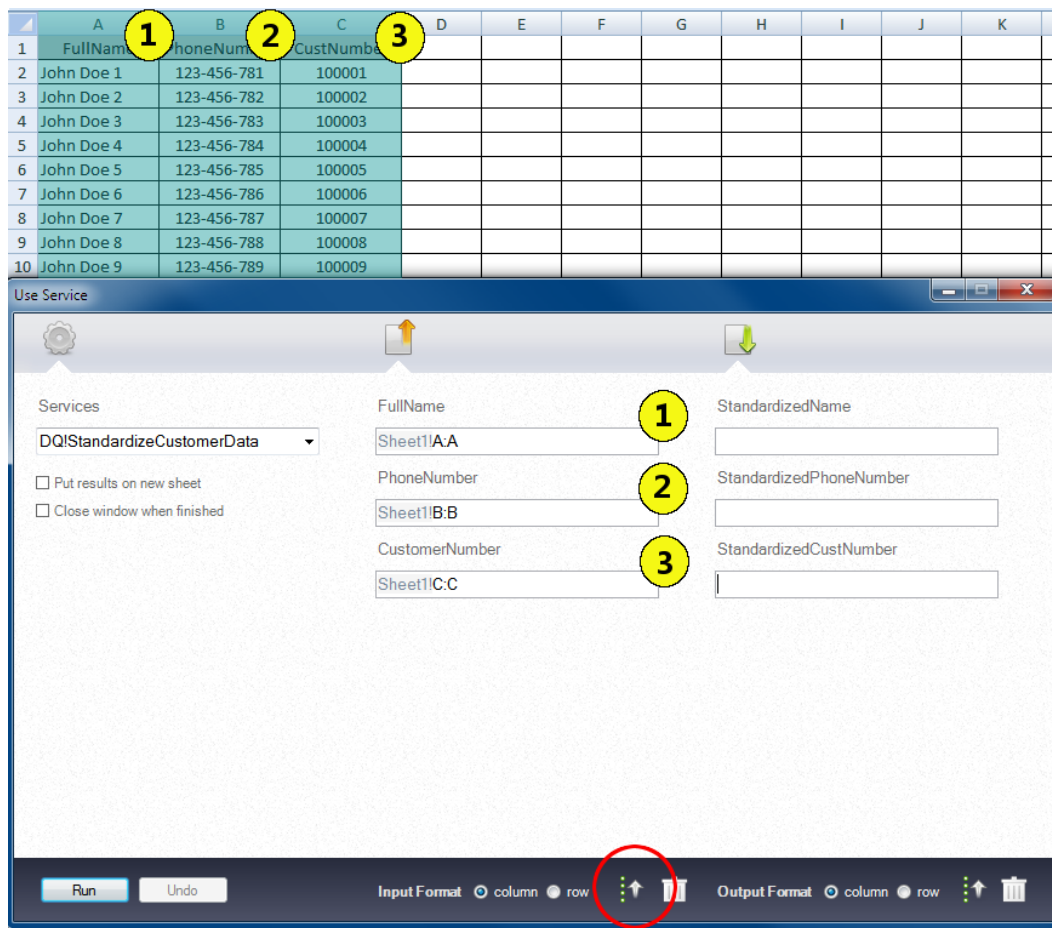
The highlighted blue areas in the Excel worksheet indicate the cell ranges that you select. The numbers in the Excel worksheet indicate the order that you select the ranges. The numbers in the **Use Service** window indicate the order in which the Data Quality for Excel plug-in populates the input fields.

If you click the **Autofill** button in the Input section of the **Use Service** window, the plug-in populates the input fields with the cell ranges indicated by the corresponding numbers. The **Autofill** button is indicated by the red circle in the figure.

Column and Row Selections

You can select column and row headers in to populate input and output fields with entire column ranges.

The following figure shows the correlation between selected columns in a worksheet and output fields in the **Use Service** window:



The highlighted blue area in the Excel worksheet indicates columns that you select. The numbers in the Excel worksheet indicate the order that you select the columns. The numbers in the **Use Service** window indicate the order in which the Data Quality for Excel plug-in populates the input fields.

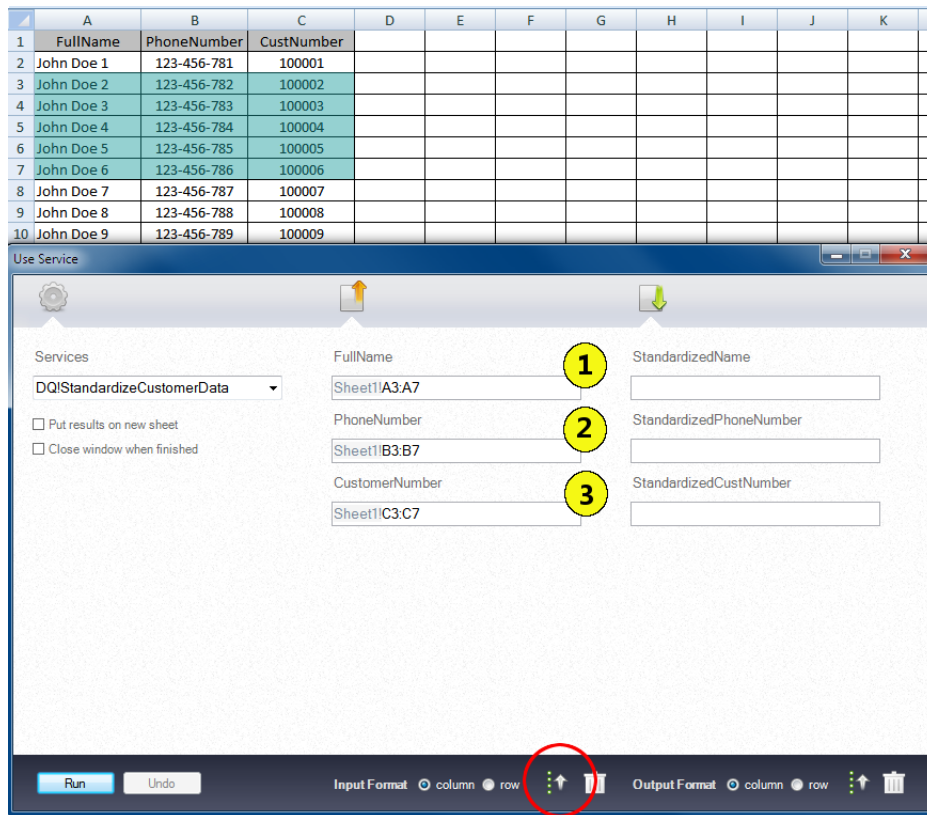
If you click the **Autofill** button in the Input section of the **Use Service** window, the plug-in populates the input fields with the column ranges from your selection. The **Autofill** button is indicated by the red circle in the figure.

When you select a column or row header, Data Quality for Excel uses the full column range. For example, if you select the header for Column C and click **Autofill**, Data Quality for Excel writes `C:C` into the currently selected service field. However, when the Data Quality for Excel plug-in runs a service, the plug-in recalculates the range in service fields that use full columns or rows. The plug-in locates the last populated cell in the row or column and updates the field so that the range uses the last populated cell as an endpoint. For example, if the last populated cell is C28, then the plug-in updates the field to read `C1:C28`.

Multiple Ranges from a Single Range

You can populate input and output fields with multiple ranges that Data Quality for Excel detects from the columns or rows in a single range.

The following figure shows the correlation between a selected range in a worksheet and output fields in the **Use Service** window:



The highlighted blue area in the Excel worksheet indicates the cell range that you select. The numbers in the **Use Service** window indicate the order in which the Data Quality for Excel plug-in populates the input fields.

If you click the **Autofill** button in the Input section of the **Use Service** window, the plug-in populates the input fields with the cell ranges from your selection. The **Autofill** button is indicated by the red circle in the figure.

Running a Data Quality Service

Select a data quality service to process data contained in Microsoft Excel worksheets.

1. Click **Use** in the Informatica ribbon.
2. Select a data quality service.
Data Quality for Microsoft Excel loads input and output fields from the data quality service.
3. Choose input and output formats.
4. Enter cell ranges in the data quality service input and output fields. You can click **Autofill** to fill the fields with cell ranges that you select in the Excel worksheet.
5. Click **Run**.
The data quality service processes the data and writes the results to the Microsoft Excel worksheet.

Data Quality for Microsoft Excel Configuration

You can define settings for a default WSDL directory, the number of Excel inputs to send in one batch, and configure service options. You can also export or import Data Quality for Microsoft Excel properties.

Export and Import Settings

You can export and import Data Quality for Microsoft Excel settings.

You can export the settings in the **Settings** window to an XML file. You can send the XML file to other Excel users so that they can configure the Data Quality for Excel plug-in to use the same settings.

To export an XML settings file, click the **Export Settings** button and choose a location to store the file. To import an XML settings file, click the **Import Settings** button and browse to a saved XML settings file.

Settings Window Properties

You can use the **Settings** window to configure Data Quality for Microsoft Excel properties.

You can configure the following properties in the **Settings** window:

Directory of WSDL files

The directory where Data Quality for Microsoft Excel plug-in searches for WSDL files when you click the **Add** button. A data quality developer provides this location to Excel users.

When you click **Add**, the plug-in reads the WSDL files in the directory and adds service links in the **Use Service** window. You can click **Browse** to select the directory to search.

Batch Size

Sets the number of inputs to send at one time to data quality services. Increasing the batch size can increase the speed at which data quality services process data. You can set the batch size from 1 to 100. Default is 1.

If you configure Data Quality for Excel to create a worksheet for service output, services do not use batch size settings.

Always Put Results on New Sheet

Sets the default value for the **Put results on new sheet** property in the **Use Service** window. Default is disabled.

Always Close Window When Finished

Sets the default value for the **Close window when finished** property in the **Use Service** window. Default is disabled.

Change Service Window Size Dynamically

Resizes the **Service** window to fit inputs and outputs. Default is enabled.

Export Settings

Exports the properties in the **Settings** window to an XML file. You can import this XML file to set the properties in the **Settings** window.

Import Settings

Imports properties from an XML file to the **Settings** window.

Service Review

You can review the service results when you troubleshoot Data Quality for Microsoft Excel.

The **Review** window displays the results of the last service used by Data Quality for Microsoft Excel. You can also save a log file that you can send to a developer.

Review Window Properties

You can use the **Review** window to review service results and export a log file.

You can review or configure the following properties:

Service Name

Name of the last service called.

Actions

Actions that occurred during the last service call, displayed in chronological order. These actions display green icons for good results, amber icons for warnings, and red icons for failures. If the **Review** window displays amber or red icons, you can save a log file to send to a data quality developer for review.

Time

Time the last service was called.

Save Details

Save a log file that you can send to a data quality developer for review.

Troubleshooting

I cannot add a data quality service.

To add a data quality service, the service must be a valid Informatica web service that Data Quality for Excel can run. Contact a developer to verify that Data Quality for Excel can run the service.

A data quality service that performs address validation does not write output data.

When you run an address validation service, the server machine loads address validation reference data. Loading this reference data can take a lengthy amount of time because the a reference data set can often be several gigabytes in size.

If the Data Quality for Microsoft Excel plug-in does not receive a response from a data quality service in a timely manner, the plug-in processes the next batch of records without writing output data. If the service does not respond before all record batches are sent, Data Quality for Microsoft Excel does not write any output data.

If an address validation service does not write output data for one or more records, run the service again for those records.

I cannot run a data quality service.

You cannot run a data quality service if any of the following conditions are true:

- The computer you are using does not have access to the application service where the data quality service is deployed.

- You do not populate a service input field.
- You populate an input or output field with a range is not available in an Excel worksheet.
- The data quality service has multiple input or output fields and you populate the fields with ranges that contain a different number of cells.

CHAPTER 4

Data Quality Service Example

This chapter includes the following topics:

- [Data Quality Service Example Overview, 21](#)
- [Input Data, 22](#)
- [Using the Data Quality Service, 22](#)
- [Output Data, 23](#)

Data Quality Service Example Overview

Your organization needs to generate a validated, formatted mailing address for customer records. You inform a data quality developer about the types of information you require in the input and output records.

You inform the developer that you will provide the following input data:

- Street Address
- Locality
- State
- ZIP Code
- Country

You inform the developer that you require the output data to be validated and formatted in three address lines.

The developer creates a data quality service that analyzes addresses and produces validated and formatted data. The developer provides you with the URL for this service.

You add the service to Data Quality for Microsoft Excel and run the service against customer address data in Excel worksheets.

The data quality service in this example analyzes address data and generates formatted postal addresses that the U.S. Postal Service certifies as deliverable addresses.

Input Data

The input data records are customer records that contain fields for street address, locality, state, ZIP Code, and country.

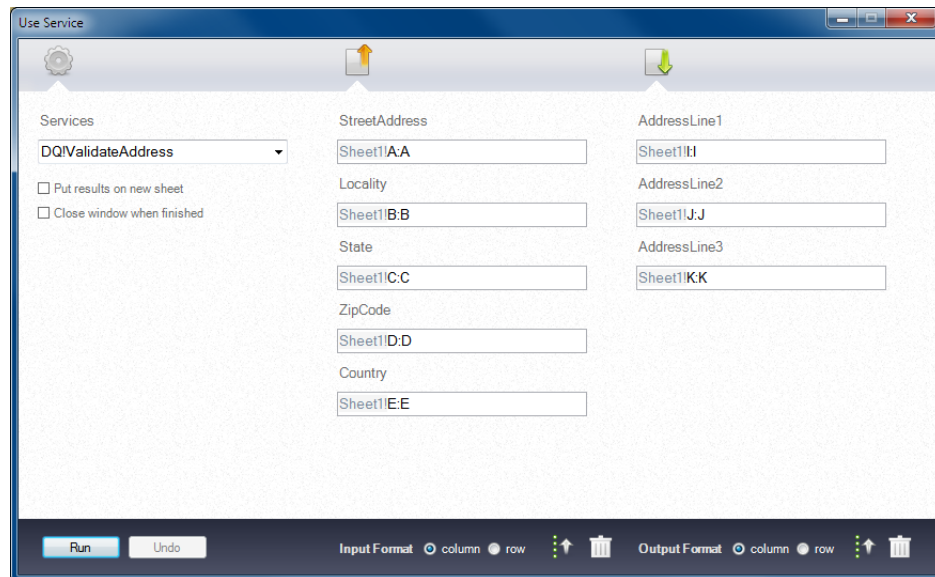
The following table shows the input data records:

Street Address	Locality	State	ZIP Code	Country
E1022 COUNTY RD P	ELMWOOD	WI	54740	USA
2009 N 16TH ST 14	BISMARCK	ND	58501	USA
222 Red School Lane	Phillipsburg	NJ	8865	USA
8 Greenway Plaza Suite 910	Houston	TX	77046	USA
5757 Underwood Rd	Pasadena	TX	77507	USA
25 Atlantic Ave.	Erlanger	KY	41018	USA
PO Box 20	Muncie	IN	61857	USA
R3 BX 112	MADISON	SD	57042	USA
497 E HWY 70	GRANDSBURG	WI	54840	USA
16924 18TH ST SE	GARDNER	ND	58036	USA

Using the Data Quality Service

To validate address data, you add the service URL provided by the developer. You can then run the service against the data in your worksheet.

The following figure shows a Service window that contains the inputs and outputs for an address validation service:



Populate the input fields with the corresponding range of data in the worksheet. When you run the service, the data quality service analyzes the input data and generates validated addresses.

Output Data

The output data records are columns of formatted address data. The U.S. Postal Service certifies these addresses as deliverable postal addresses.

The following table shows the output data records:

Address1	Address2	Address3
E1022 COUNTY RD P	ELMWOOD WI 54740	UNITED STATES
2009 N 16TH ST APT 14	BISMARCK ND 58501-2058	UNITED STATES
222 RED SCHOOL LN	PHILLIPSBURG NJ 08865-2219	UNITED STATES
8 GREENWAY PLZ STE 910	HOUSTON TX 77046-0892	UNITED STATES
5757 UNDERWOOD RD	PASADENA TX 77507-1031	UNITED STATES
25 ATLANTIC AVE	ERLANGER KY 41018-3151	UNITED STATES
PO BOX 20	MUNCIE IL 61857-0020	UNITED STATES
R3 BX 112	MADISON SD 57042	UNITED STATES
497 E HWY 70	GRANDSBURG WI 54840	UNITED STATES
16924 18TH ST SE	GARDNER ND 58036-9737	UNITED STATES

CHAPTER 5

Data Quality Services for Developers

This chapter includes the following topics:

- [Data Quality Services for Developers Overview, 24](#)
- [Deploying Data Quality for Excel in an Organization, 25](#)
- [Data Quality Service Requirements, 25](#)
- [Data Quality Services Optimization, 25](#)
- [Adding a Data Quality Service to a Data Integration Service, 26](#)
- [Adding a Data Quality Service to a Web Services Hub, 26](#)
- [Provide a Data Quality Service to Excel Users, 26](#)

Data Quality Services for Developers Overview

A data quality developer can add data quality services to an Informatica web service or to a PowerCenter Web Services Hub. To begin creating a data quality service, the data quality developer uses the Developer tool to create data quality mappings.

To add a data quality service to a PowerCenter Web Services Hub, the developer exports the mapping to the PowerCenter repository as a maplet. The developer uses the PowerCenter Developer and the PowerCenter Workflow Manager to reuse data quality mapping logic in a web service mapping. The developer creates a workflow for the web service mapping and configures web services for that workflow.

To add a data quality service to an Informatica web service, the developer uses Informatica Developer to reuse data quality mapping logic in a web service. The developer deploys the web service to an application in the Data Integration Service.

Deploying Data Quality for Excel in an Organization

To simplify the task of managing data quality services for organization users, you can export Data Quality for Excel settings to an XML file that users can import. You can also store WSDL files in a network location.

After you install and configure Data Quality for Excel, you can export the plug-in settings to an XML file. You send this file to Excel users in the organization. To import the settings, Excel users open the **Settings** window in Data Quality for Excel and use the **Import Settings** button.

You can save data quality services as WSDL files. You can store those files in a network directory accessible by Excel users, and you can update the files in this directory as needed. You can send the WSDL directory location to Excel users, or you can send Excel users an XML settings file that contains the WSDL directory location.

When you update WSDL files in a network location, you must tell Excel users. To update the data quality services from the network location, Excel users open the **Settings** window in Data Quality for Excel and use the **Add** button. The plug-in updates the services using the files in the WSDL directory specified in the Settings window. You

Data Quality Service Requirements

Data quality services that you run in the Data Quality for Excel plug-in must meet a number of requirements.

When developing services for Data Quality for Excel, a developer should verify that the services meet the following requirements:

- The service runs on an Informatica 9.1.0 or later web service or on a PowerCenter 8.6.1 or later Web Services Hub.
- The service does not require a user name and password.
- The service does not use the HTTPS protocol.
- The service does not use certificates for authentication.
- The service does not generate faults.

Data Quality Services Optimization

When a developer creates a web service for Microsoft Excel, the developer configures the web service to process data from Microsoft Excel more quickly and more efficiently.

Use the following conventions:

- If a web service contains multiple inputs, use a parent element to group the input elements.
- Set "Maximum Occurrences" to "unbounded" to allow the web service to process multiple records in one batch. For web services with multiple inputs, set the "Maximum Occurrences" property for the parent element to "unbounded." For web services with a single input, set the "Maximum Occurrences" property for the input element to "unbounded."

Adding a Data Quality Service to a Data Integration Service

To run a data quality service using the Data Quality Web Services Option, you must verify that a Data Integration service is running in the domain. You use the Developer tool to add a data quality service to the Data Integration Service.

1. Verify that a Data Integration service is running in the domain.
2. In the Developer tool, create a data quality mapping in the Model repository service.
3. Select **File > New > Data Service**.
4. Select the web service option.
5. Click **Next** to configure the web service.
6. Click **Next** to add and configure input and output elements.
7. Click **Finish**.
8. Open the web service in the editor.
9. Click the mapping operation to open the operation in an editor tab.
10. Open the data quality mapping you created earlier.
11. Copy the objects in the mapping editor.
12. Open the mapping operation tab for the data service. Paste the copied objects into the mapping operation.
13. Deploy the service object to an application in the Data Integration Service.

Adding a Data Quality Service to a Web Services Hub

To run a data quality service using PowerCenter, you must verify that a PowerCenter Web Services Hub is running in the domain. You export data quality mappings from the Developer tool to the PowerCenter Client. You copy the data quality mapping objects to web services mappings that you configure as web services.

1. Verify that a PowerCenter Web Services Hub is running in the domain.
2. Export a data quality mapping from the Developer tool to the PowerCenter repository as a mapplet.
3. In the PowerCenter Developer, create a web service source and target.
4. In the PowerCenter Developer, create a web service mapping using the web service source, target, and the contents of the imported data quality mapping.
5. In the PowerCenter Workflow Manager, create a workflow for the web service mapping
6. In the PowerCenter Workflow Manager, configure the web services options for the workflow.

Provide a Data Quality Service to Excel Users

Developers can provide data quality services to Data Quality for Microsoft Excel users through web service URLs or WSDL files in XML format.

Copying Data Service Web Service URLs for Microsoft Excel Users

A developer can copy the URL for the data quality service from the General Properties of the web service in Informatica Administrator. The URL uses the following structure: `http://<<ServerName>>:<<PortName>>/DataIntegrationService/WebService/<<Service_Name>>/`.

1. Select the Data Integration Service in the Administrator tool.
2. Select the Applications tab.
3. Expand the application that contains the web service.
4. Select the web service.
5. In the Properties pane, copy the URL located in the WSDL URL property.
6. Send the URL to the Excel user.

Copying PowerCenter Web Services Hub URLs for Microsoft Excel Users

A developer can copy Web Services Hub URLs from the PowerCenter Web Services Hub. These URLs use the following structure: `http://<<ServerName>>:<<PortName>>/wsh/services/RealTime/<<Service_Name>>?WSDL`.

1. Select the Web Services Hub Service in the Administrator tool.
2. Click the Service URL in the Properties tab to open the PowerCenter Web Services Hub.
3. In the Web Services Hub tree view, select a web service type.
4. In the Web Services tab, select a service.
5. In the Web Services toolbar, click the **WSDL** button.
The WSDL file opens in another browser window.
6. Copy the URL from the browser address bar.
7. Send the URL to the Excel user.

Saving WSDL files for Data Quality for Microsoft Excel

To save a WSDL file for Data Quality for Microsoft Excel, a developer opens a URL for a data quality service in a web browser. The developer uses the browser **File** menu to save the displayed XML page as a WSDL file.

The developer sends the saved WSDL file to an Excel user. Excel users add these files to store service links in Data Quality for Microsoft Excel.

INDEX

A

Autofill function
column ranges [15](#)
multiple ranges [16](#)
selection order [14](#)
unaligned cell ranges [15](#)

C

cell ranges
selecting unaligned [15](#)
column ranges
Autofill [15](#)
commands
Informatica ribbon [11](#)

D

data quality services
adding [12](#), [19](#)
copying web service hub URLs [27](#)
copying web service URLs [27](#)
creating [7](#)
Data Integration service [26](#)
deploying [25](#)
example overview [21](#)
input fields [14](#)
installing [10](#)
optimizing [25](#)
output fields [14](#)
overview [7](#)
removing [12](#)
requirements [25](#)
reusing settings [18](#)
reviewing results [19](#)
running [17](#), [19](#)
saving WSDL files [27](#)
troubleshooting [19](#)
usage example [22](#)
using [7](#)
using PowerCenter [26](#)
Web Service Hub [26](#)
Web Services option [26](#)

I

Informatica ribbon
commands [11](#)
input data records
example [22](#)
input fields
Autofill [14](#)

input fields (*continued*)
column ranges [15](#)
example [22](#)
multiple ranges [16](#)
selection order [14](#)
unaligned cell ranges [15](#)
installing
Data Quality for Excel [10](#)

K

keyboard shortcuts
Use Service window [13](#)

M

missing output
troubleshooting [19](#)

O

output data records
example [23](#)
output fields
Autofill [14](#)
column ranges [15](#)
example [23](#)
multiple ranges [16](#)
unaligned cell ranges [15](#)

P

prerequisites
data quality developers [9](#)
Excel users [9](#)
properties
Review window [19](#)
Settings window [18](#)
Use Service window [12](#)

R

results
reviewing [19](#)
Review window
properties [19](#)

S

- selection order
 - Autofill function [14](#)
 - column ranges [15](#)
 - multiple ranges [16](#)
 - unaligned cell ranges [15](#)
- settings
 - exporting [18](#)
 - importing [18](#)
- Settings window
 - properties [18](#)

U

- unaligned cell ranges
 - Autofill [15](#)

- Use Service window
 - keyboard shortcuts [13](#)
 - properties [12](#)

W

- web service
 - overview [7](#)
- WSDL files
 - saving [27](#)
 - storing [25](#)