informatica Cloud

Data Quality Quick Start Bundle User Guide

Overview

Use the Data Quality Quick Start bundle to use data quality functionality to process company and person data in mapping configuration tasks.

A bundle is a set of prebuilt integration templates that you can run with mapping configuration tasks. Bundles improve development productivity and enhance the quality of data quality projects.

The Data Quality Quick Start bundle includes the following integration templates:

- **DQ_Account_Match.** Matches account data based on the company name and billing address.
- **DQ_Contact_Lead_Match.** Matches personal data based on the name and mailing address.
- **DQ_Account_Std_Validate_NorthAmerica.** Standardizes the company name in account data based on Data Quality dictionary files and validates address information.
- **DQ_Contact_Std_Validate_NorthAmerica.** Parses name information for person data, generates gender information, and validates date of birth, email, and address information.
- **DQ_Lead_Std_Validate_NorthAmerica.** Parses name information for person data, generates gender information, and validates date of birth, email, and address information.

Note: The integration templates are based on Salesforce data object structures, but can be used for any similar data. For example, you can use the DQ_Account_Match integration template to match data for any source data that includes company names and addresses.

For more information about mapplets, integration templates, or other Informatica Cloud functionality, see the Informatica Cloud User Guide or online help.

Installing the Bundle

The Data Quality Quick Start bundle displays as an available bundle in your organization. To view and install the bundle, in your organization, click **Configure** > **Published Bundles**.

After you install the bundle, you can use the integration templates in the bundle.

Prerequisites

Before you use the Data Quality Quick Start bundle, verify the following prerequisites:

- Informatica Cloud Standard Edition.
- DQ Package license.

Bundle Sample Files

The bundle includes sample files that you can use. The sample files provide a basic structure and some sample data to enable working with the integration templates. You might need to edit the files to work in your environment.

You can download the sample files from the following community article: Data Quality Quick Start Bundle. You can download the following sample files: DQ_QuickStart_SampleFiles.zip.

DQ_Account_Match Integration Template

Use the DQ_AccountMatch integration template to match account data based on the company name and billing address. This template was developed with the Salesforce Account object as the source, but you can use any account or company data that includes name and address information.

Matching occurs between a set of records grouped by a user-defined group key field. Group key configuration can affect the accuracy of the matching as well as task performance. Consider the grouping and performance results when you choose the group key. For example, using a country code as the group key field can result in large groups that cause performance issues. However, grouping by postal code can result in restrictively small groups that miss potential matches. Instead, you might group by the first three characters of the company name concatenated with the substring of the first three digits of the zip code

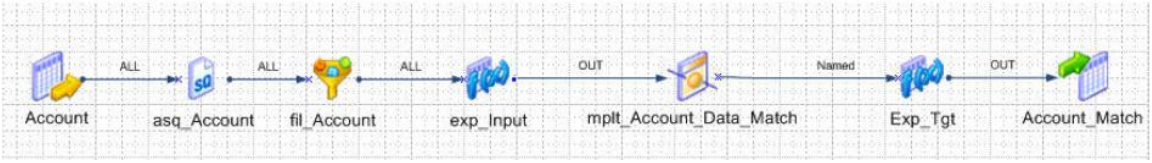
The template creates a cluster ID for the first row in a group, also known as the *master record*. All other rows in the group are then compared to the master record. When a match is found, the row is assigned the same cluster ID as the master record, and the cluster size is increased by 1. A unique row has a cluster size of 1.

Note: For best results, cleanse the billing address data before you use this template.

Template Data Flow

In this template, you can configure a filter to reduce the data passed to the mapplet. The Expression transformations provide the field mapping parameters for the mapplet and the target.

The following image shows the data flow of the template:



Template Parameters

The following table describes the parameters in the template:

Parameter Name	Label in Task	Description
$Account$	Account	Source connection and source object that includes company data.
$Account_Match$	Account_Match	Target connection and target object for matched company data.
$Fil_Condition$	Filter	Filter condition. Excludes records based on the condition. The task performs the filter on source data. By default the filter condition is set to TRUE, which passes all rows.
$Input_FieldMap$	Input_FieldMap	Field mappings from the data flow to the mapplet.
$Tgt_FieldMap$	Target_FieldMap	Field mappings from the mapplet to the target.

Mapplet Input and Output Ports

Connect the data flow to the mapplet input ports and output ports as necessary.

The following table shows the mapplet input ports:

Mapplet Input Port	Description
InRowdID	Required. Record/primary key field.
In_GroupKey	Required. Group key field. Use to group data together for matching.
Company	Required. Company name.
The mapplet includes the following billing address fields: • BillCity • BillState • BillPostcode • BillCountry	Billing address fields. **Note:** Although billing address fields are not mandatory, they are critical for matching. Unmapped fields or missing data can adversely impact matching.

Mapplet Input Port	Description
The mapplet includes the following shipping address fields: • ShipStreetAdrs • ShipCity • ShipState • ShipPostcode • ShipCountry	Shipping address fields.
The mapplet includes the following additional company fields: • Phone • Fax • WebAdrs • TickerSymbol	Additional company fields.

The following table shows the mapplet output ports:

Mapplet Output Field	Description
Out_UID_OUTPUT	Data from the In_RowID with no additional processing. Used for joining data after the match process.
Out_CompanyName_OUTPUT	Company name.
Bill address output fields: • Out_BillingStreetAdrs_OUTPUT • Out_BillingCity_OUTPUT • Out_BillingState_Prov_OUTPUT • Out_BillingPostalcode_OUTPUT • Out_BillingCountry_OUTPUT	Billing output fields.
Shipping address output fields: • Out_ShippingStreetAdrs_OUTPUT • Out_ShippingCity_OUTPUT • Out_ShippingState_Prov_OUTPUT • Out_ShippingPostalcode_OUTPUT • Out_ShippingCountry_OUTPUT	Shipping address output fields.
Other company output fields: • Out_Phone_OUTPUT • Out_Fax_OUTPUT • Out_WebSite_OUTPUT • Out_TickerSymbol_OUTPUT	Other company output fields.
Out_ClusterId_OUTPUT	Cluster ID. Each row matched in a cluster uses the same cluster ID.
Out_GroupKey_OUTPUT	Group key field.
Out_ClusterSize_OUTPUT	Number of rows grouped in the cluster.
Out_RowId_OUTPUT	Row ID used for matching.
Out_DriverId_OUTPUT	Row number of the master row in a cluster.
Out_DriverScore_OUTPUT	Shows the match score when the driver score is used to determine the match.
Out_LinkId_OUTPUT	Row number of other, non-master rows in a cluster.
Out_LinkScore_OUTPUT	Shows the match score when the link score is used to determine the match.

Sample Files

You can use the following sample files with the template:

- Source. Account_data_match_in.csv.
- Target. Out_account_match.csv.

DQ_Contact_Lead_Match Integration Template

Use the DQ_Contact_Lead_Match integration template to match person data based on name and mailing address. This template was developed with the Salesforce Contact and Lead objects as the source, but you can use any person-related data that includes name and address information.

Matching occurs between a set of records grouped by a user-defined group key field. Group key configuration can affect the accuracy of the matching as well as task performance. Consider the grouping and performance results when you choose the group key. For example, using a country code as the group key field can result in large groups that cause performance issues. However, grouping by postal code can result in restrictively small groups that miss potential matches. Instead, you might group by the first three characters of the company name concatenated with the substring of the first three digits of the zip code

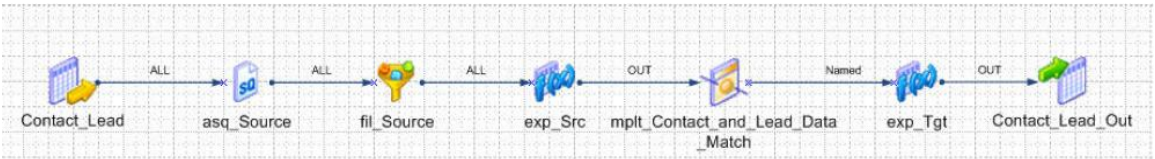
The template creates a cluster ID for the first row in a group, also known as the *master record*. All other rows in the group are then compared to the master record. When a match is found, the row is assigned the same cluster ID as the master record, and the cluster size is increased by 1. A unique row has a cluster size of 1.

For best results, cleanse the billing address data before you use this template.

Template Data Flow

In this template, you can configure a filter to reduce the data passed to the mapplet. The Expression transformations provide the field mapping parameters for the mapplet and the target.

The following image shows the data flow of the template:



Template Parameters

The following table describes the parameters in the template:

Parameters	Label in Task	Description
$Contact_Lead$	Contact_Lead	Source connection and source object for contact lead data.
$Contact_Lead_Out$	Contact_Lead_Out	Target connection and target object.
$fil_Condition$	Filter	Filter condition. Excludes records based on the condition. The task performs the filter on source data. By default the filter condition is set to TRUE, which passes all rows.
$Input_FieldMap$	Input_FieldMap	Field mappings from the data flow to the mapplet.
$Tgt_FieldMap$	Tgt_FieldMap	Field mappings from the mapplet to the target.

Mapplet Input and Output Ports

Connect the data flow to the mapplet input ports and output ports as necessary.

The following table shows the mapplet input ports:

Mapplet Input Port	Description
InRowdID	Required. Record/primary key field.
In_GroupKey	Required. Group key field. Use to group data together for matching.
Name	Name.
Title	Title.
Company	Company name.
The mapplet includes the following mailing address fields: • MailAddress1 • MailCity • MailState • MailPostcode • MailCountry	Billing address fields. **Note:** Although billing address fields are not mandatory, they are critical for matching. Unmapped fields or missing data can adversely impact matching.
The mapplet includes the following secondary address fields: • OtherAddress1 • OtherCity • OtherState • OtherPostcode • OtherCountry	Secondary address fields. .
The mapplet includes the following additional personal information: • Email • Phone • Fax • MobilePhone • HomePhone • OtherPhone • AssistantPhone • DateofBirth • WebAdrs	Additional personal information.

The following table shows the mapplet output ports:

Mapplet Output Field	Description
Out_UID_OUTPUT	Data from the In_RowID with no additional processing.
Out_Name_OUTPUT	Name.
Out_Title_OUTPUT	Title.
Out_Company_OUTPUT	Company.
The mapplet includes the following address fields: • Out_MailingStreetAdrs_OUTPUT • Out_MailingCity_OUTPUT • Out_MailingState_Prov_OUTPUT • Out_MailingPostalcode_OUTPUT • Out_MailingCountry_OUTPUT	Address output fields.
The mapplet includes the following secondary address fields: • Out_OtherStreetAdrs_OUTPUT • Out_OtherCity_OUTPUT • Out_OtherState_Prov_OUTPUT • Out_OtherPostalcode_OUTPUT • Out_OtherCountry_OUTPUT	Secondary address output fields.

Mapplet Output Field	Description
The mapplet includes the following additional personal information fields: • Out_Email_OUTPUT • Out_Phone_OUTPUT • Out_Fax_OUTPUT • Out_MobilePhone_OUTPUT • Out_HomePhone_OUTPUT • Out_OtherPhone_OUTPUT • Out_AssistantPhone_OUTPUT • Out_Date_of_Birth_OUTPUT • Out_WebSite_OUTPUT	Additional personal information fields.
Out_ClusterId_OUTPUT	Cluster ID. Each row matched in a cluster uses the same cluster ID.
Out_GroupKey_OUTPUT	Group key field.
Out_ClusterSize_OUTPUT	Number of rows grouped in the cluster.
Out_RowId_OUTPUT	Row ID used for matching.
Out_DriverId_OUTPUT	Row number of the master row in a cluster.
Out_DriverScore_OUTPUT	Shows the match score when the driver score is used to determine the match.
Out_LinkId_OUTPUT	Row number of other, non-master rows in a cluster.
Out_LinkScore_OUTPUT	Shows the match score when the link score is used to determine the match.

Sample Files

You can use the following sample files with the template:

- Source. Contact_lead_data_match_in.csv.
- Target. Out_contact_lead_match.csv.

DQ_Account_Std_Validate_NorthAmerica Integration Template

Use the DQ_Account_Std_Validate_NorthAmerica integration template to standardize and validate account information. This template was developed with the Salesforce Account object as the source, but you can use any data that includes company name and address information.

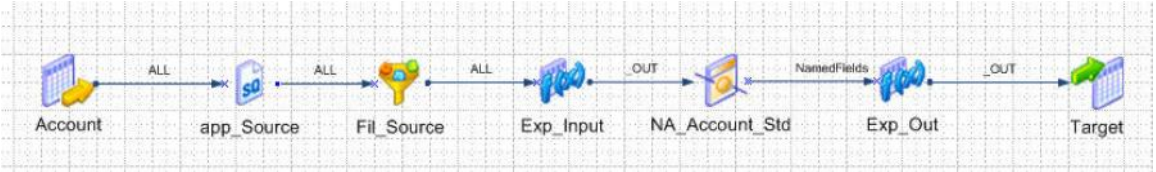
The template validates and standardizes company data as follows:

- Standardizes company name against a static dictionary
- Standardizes phone numbers and fax numbers to the following format (nnn) nnn-nnnn
- Generates the following country information from static dictionaries: full country name, two-character country code, three-character country code.
- Corrects the case for websites and ticker symbols.
- Validates state, country, and postal codes against static dictionaries.

Template Data Flow

In this template, you can configure a filter to reduce the data passed to the mapplet. The Expression transformations provide the field mapping parameters for the mapplet and the target.

The following image shows the data flow of the template:



Template Parameters

The following table describes the parameters in the template:

Parameter Name	Label in Task	Description
$Account$	Account	Source connection and source object that includes company data.
Tgt	Account_Std	Target connection and target object for matched company data.
$Filter$	Filter	Filter condition. Excludes records based on the condition. The task performs the filter on source data. By default the filter condition is set to TRUE, which passes all rows.
$Src_FieldMap$	Source_FieldMap	Field mappings from the data flow to the mapplet.
$TgtFieldMap$	Target_FieldMap	Field mappings from the mapplet to the target.

Mapplet Input and Output Ports

Connect the data flow to the mapplet input ports and output ports as necessary.

The following table shows the mapplet input ports:

Mapplet Input Port	Description
In_UID	Required. Record/primary key field.
In_CompanyName	Required. Company name.
The mapplet includes the following billing address fields: • In_BillingStreetAdrs • In_BillingCity • In_BillingState_Prov • In_BillingPostalcode • In_BillingCountry	Billing address fields.
The mapplet includes the following shipping address fields: • In_ShippingStreetAdrs • In_ShippingCity • In_ShippingState_Prov • In_ShippingPostalcode • In_ShippingCountry	Shipping address fields.
The mapplet includes the following additional company information: • In_Phone • In_Fax • In_Web_URL • In_TickerSymbol	Additional company information.

The following table shows the mapplet output ports:

Mapplet Output Field	Description
Out_UID	Data from the In_RowID with no additional processing.

Mapplet Output Field	Description
Out_CompanyName_Std	Company name standardized based on a static data quality dictionary.
The mapplet includes the following billing address fields: • Out_BillingStreetAdrs • Out_BillingCity • Out_BillingState_Prov • Out_BillingPostalcode • Out_Billing_Country • Out_Full_ISO_BillingCountryName • Out_2Char_ISO_BillingCountryCode • Out_3Char_ISO_BillingCountryCode	Billing address fields.
The mapplet includes the following shipping address fields: • Out_ShippingStreetAdrs • Out_ShippingCity • Out_ShippingState_Prov • Out_ShippingPostalcode • Out_ShippingCountry • Out_Full_ISO_ShippingCountryName • Out_2Char_ISO_ShippingCountryCode • Out_3Char_ISO_ShippingCountryCode	Shipping address fields.
The mapplet includes the following standardized company information: • Out_Website_Std • Out_Phone_Std • Out_Fax_Std • Out_TickerSymbol_Std	Company information placed in a standardized format.
The mapplet includes the following status fields: • Out_Billing_State_Status • Out_Billing_Zipcode_Status • Out_Billing_Country_Status • Out_Shipping_Country_Status • Out_Shipping_Zipcode_Status • Out_Shipping_State_Status • Out_Phone_Status • Out_Fax_Status	Status fields. Indicate the validity of data in the corresponding field.

Sample Files

You can use the following sample files with the template:

- Source. Account_Input.csv.
- Target. Out_Account_Std.csv.

DQ_Contact_Std_Validate_NorthAmerica Integration Template

Use the DQ_Contact_Std_Validate_NorthAmerica integration template to standardize and validate contact information. This template was developed with the Salesforce Contact object as the source, but you can use any data that includes personal data.

The template validates and standardizes personal data as follows:

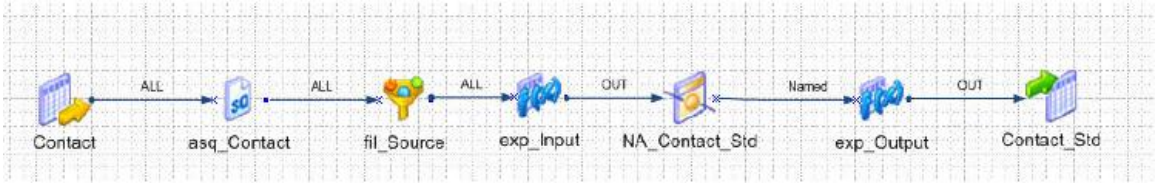
- Standardizes first names against a static dictionary.
- Generates gender based on name.
- Standardizes phone numbers and fax numbers to the following format (nnn) nnn-nnnn.
- Standardizes the date of birth to the following format: MM-DD-YYYY.

- Validates state, country, and postal codes against static dictionaries.
- Generates the following country information from static dictionaries: full country name, two-character country code, three-character country code.

Template Data Flow

In this template, you can configure a filter to reduce the data passed to the mapplet. The Expression transformations provide the field mapping parameters for the mapplet and the target.

The following image shows the data flow of the template:



Template Parameters

The following table describes the parameters in the template:

Parameter Name	Label in Task	Description
$Contact_Input$	Contact	Source connection and source object that includes contact data.
$Contact_Std$	Contact_Std	Target connection and target object for standardized and validated contact data.
$Filter$	Filter	Filter condition. Excludes records based on the condition. The task performs the filter on source data. By default the filter condition is set to TRUE, which passes all rows.
$Input_FieldMap$	Input_FieldMap	Field mappings from the data flow to the mapplet.
$Tgt_FieldMap$	Target_FieldMap	Field mappings from the mapplet to the target.

Mapplet Input and Output Ports

Connect the data flow to the mapplet input ports and output ports as necessary.

The following table shows the mapplet input ports:

Mapplet Input Port	Description
In_UID	Required. Record/primary key field.
In_Name	Required. Name.
In_Title	Title.
The mapplet includes the following mailing address fields: • In_MailingStreetAdrs • In_MailingCity • In_MailingState_Prov • In_MailingPostalcode • In_MailingCountry	Mailing address fields. .

Mapplet Input Port	Description
The mapplet includes the following secondary address fields: • In_OtherStreetAdrs • In_OtherCity • In_OtherState_Prov • In_OtherPostalcode • In_OtherCountry	Secondary address fields. .
The mapplet includes the following additional personal information: • In_Email • In_Phone • In_Fax • In_MobilePhone • In_HomePhone • In_OtherPhone • In_AssistantPhone • In_DateOfBirth	Additional personal information. .

The following table shows the mapplet output ports:

Mapplet Output Field	Description
Out_UID	Data from the In_RowID with no additional processing.
The mapplet includes the following name fields: • Out_FullName • Out_NamePrefix • Out_FirstName • Out_Firstname_Std • Out_MidName • Out_Surname • Out_NameSuffix	Name information.
Out_Title	Title.
Out_Gender	Gender based on name information.
Out_NonNameData	Data passed with name data that cannot be parsed as a name.
The mapplet includes the following mailing address fields: • Out_MailingStreetAdrs • Out_MailingCity • Out_MailingState_Prov • Out_MailingPostalcode • Out_Mailing_Country • Out_Mailing_Full_ISO_CountryName • Out_Mailing_2Char_ISO_CountryCode • Out_Mailing_3Char_ISO_CountryCode	Mailing address fields.
The mapplet includes the following secondary address fields: • Out_OtherStreetAdrs • Out_OtherCity • Out_OtherState_Prov • Out_OtherPostalcode • Out_OtherCountry • Out_Other_Full_ISO_CountryName • Out_Other_2Char_ISO_CountryCode • Out_Other_3Char_ISO_CountryCode	Secondary address fields.

Mapplet Output Field	Description
The mapplet includes the following additional personal information fields: • Out_DOB_Std_Format • Out_Email_Std • Out_Phone_Std • Out_Fax_Std • Out_MobilePhone_Std • Out_HomePhone_Std • Out_OtherPhone_Std • Out_AssisantPhone_Std	Additional personal information fields in standardized formats.
The mapplet includes the following status fields: • Out_NameParse_Status • Out_NameParse_Status_Message • Out_NameParse_Confidence • Out_Mailing_Postcode_Status • Out_Mailing_State_Status • Out_Mailing_Country_Status • Out_Other_State_Status • Out_Other_Postcode_Status • Out_Other_Country_Status • Out_DOB_Status • Out_Email_Status • Out_Phone_Status • Out_Fax_Status • Out_MobilePhone_Status • Out_HomePhone_Status • Out_OtherPhone_Status • Out_AssisantPhone_Status	Status fields. Indicate the validity of data in the corresponding field.

Sample Files

You can use the following sample files with the template:

- Source. Contact_Input.csv.
- Target. Out_Contact_Std.csv.

DQ_Lead_Std_Validate_NorthAmerica Integration Template

Use the DQ_Lead_Std_Validate_NorthAmerica integration template to standardize and validate lead information. This template was developed with the Salesforce Lead object as the source, but you can use any data that includes personal data.

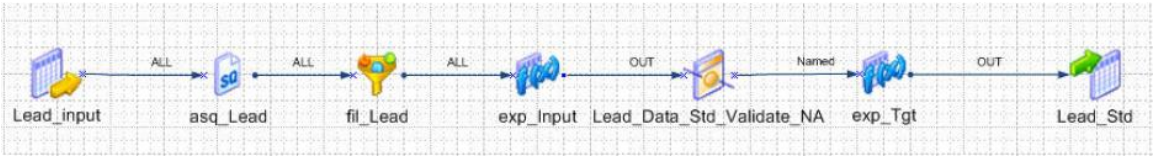
The template validates and standardizes personal data as follows:

- Standardizes first names against a static dictionary.
- Generates gender based on name.
- Standardizes phone numbers and fax numbers to the following format (nnn) nnn-nnnn.
- Standardizes the date of birth to the following format: MM-DD-YYYY.
- Validates state, country, and postal codes against static dictionaries.
- Generates the following country information from static dictionaries: full country name, two-character country code, three-character country code.

Template Data Flow

In this template, you can configure a filter to reduce the data passed to the mapplet. The Expression transformations provide the field mapping parameters for the mapplet and the target.

The following image shows the data flow of the template:



Template Parameters

The following table describes the parameters in the template:

Parameter Name	Label in Task	Description
$Lead_Input$	Lead	Source connection and source object that includes lead data.
$Lead_Std$	Lead_Std	Target connection and target object for standardized and validated lead data.
$fil_Condition$	Filter	Filter condition. Excludes records based on the condition. The task performs the filter on source data. By default the filter condition is set to TRUE, which passes all rows.
$Input_FieldMap$	Input_FieldMap	Field mappings from the data flow to the mapplet.
$Target_FieldMap$	Target_FieldMap	Field mappings from the mapplet to the target.

Mapplet Input and Output Ports

Connect the data flow to the mapplet input ports and output ports as necessary.

The following table shows the mapplet input ports:

Mapplet Input Port	Description
In_UID	Required. Record/primary key field.
In_Name	Required. Name.
In_Title	Title.
The mapplet includes the following mailing address fields: • In_MailingStreetAdrs • In_MailingCity • In_MailingState_Prov • In_MailingPostalcode • In_MailingCountry	Mailing address fields. .
The mapplet includes the following secondary address fields: • In_OtherStreetAdrs • In_OtherCity • In_OtherState_Prov • In_OtherPostalcode • In_OtherCountry	Secondary address fields. .
The mapplet includes the following additional personal information: • In_Email • In_Phone • In_Fax • In_MobilePhone • In_HomePhone • In_OtherPhone • In_AssistantPhone • In_DateOfBirth	Additional personal information. .

The following table shows the mapplet output ports:

Mapplet Output Field	Description
Out_UID	Data from the In_RowID with no additional processing.
The mapplet includes the following name fields: • Out_FullName • Out_NamePrefix • Out_FirstName • Out_Firstname_Std • Out_MidName • Out_Surname • Out_NameSuffix	Name information.
Out_Title	Title.
Out_Gender	Gender based on name information.
Out_NonNameData	Data passed with name data that cannot be parsed as a name.
The mapplet includes the following mailing address fields: • Out_MailingStreetAdrs • Out_MailingCity • Out_MailingState_Prov • Out_MailingPostalcode • Out_Mailing_Country • Out_Mailing_Full_ISO_CountryName • Out_Mailing_2Char_ISO_CountryCode • Out_Mailing_3Char_ISO_CountryCode	Mailing address fields.
The mapplet includes the following secondary address fields: • Out_OtherStreetAdrs • Out_OtherCity • Out_OtherState_Prov • Out_OtherPostalcode • Out_OtherCountry • Out_Other_Full_ISO_CountryName • Out_Other_2Char_ISO_CountryCode • Out_Other_3Char_ISO_CountryCode	Secondary address fields.
The mapplet includes the following additional personal information fields: • Other Out_DOB_Std_Format • Out_Email_Std • Out_Phone_Std • Out_Fax_Std • Out_MobilePhone_Std • Out_HomePhone_Std • Out_OtherPhone_Std • Out_AssisantPhone_Std	Additional personal information fields in standardized format.

Mapplet Output Field	Description
The mapplet includes the following status fields: • Out_NameParse_Status • Out_NameParse_Status_Message • Out_NameParse_Confidence • Out_Mailing_Postcode_Status • Out_Mailing_State_Status • Out_Mailing_Country_Status • Out_Other_State_Status • Out_Other_Postcode_Status • Out_Other_Country_Status • Out_DOB_Status • Out_Email_Status • Out_Phone_Status • Out_Fax_Status • Out_MobilePhone_Status • Out_HomePhone_Status • Out_OtherPhone_Status • Out_AssisantPhone_Status	Status fields. Indicate the validity of data in the corresponding field.

Sample Files

You can use the following sample files with the template:

- Source. Lead_Input.csv.
- Target. Out_Lead_Std.csv.

Configuring the Mapping Configuration Task

After you import the template to your organization, you can use it in a mapping configuration task.

Use the Mapping Configuration Task wizard to create a new task and configure it as follows:

1. On the **Definition** page, select the integration template.
2. On the **Sources** page, select a source connection and source object.
3. On the **Targets** page, select the target connection and target object.
4. On the **Mapplet Input** page, optionally add a filter condition to reduct the source data passed to the mapplet, and then configure the mapplet input field mappings.
5. On the **Mapplet Output** page, configure the mapplet output field mappings.
6. Save and run the task.

Informatica Global Customer Support

You can contact a Customer Support Center online or by telephone.

For online support, click Submit Support Request in the Informatica Cloud application. You can also use Informatica MySupport to log a case. MySupport requires a user name and password. You can request a user name and password at https://mysupport.informatica.com.

The telephone numbers for Informatica Global Customer Support are available from the Informatica web site at http://www.informatica.com/us/services-and-training/support-services/global-support-centers/.