Informatica® Data Quality
10.5

Accelerator Guide

Table of Contents

Chapter 13: Spain Accelerator. 126

Chapter 14: United Kingdom Accelerator. 133

Chapter 15: U.S./Canada Accelerator. 144

Preface

Read the *Data Quality Accelerator Guide* to learn about the accelerators that Informatica prepares for Data Quality users. Accelerators are content bundles that address common data quality issues in a country, a region, or an industry. Accelerators can include prebuilt mapplets, data domains, and reference data objects that you can import to your Informatica environment.

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 Accelerators

This chapter includes the following topics:

Accelerators Overview

Accelerators are content bundles that address common data quality problems in a country, a region, or an industry. An accelerator might contain mapplets or rule specifications that you can use to analyze and enhance the data in an organization. An accelerator might also contain data domains that you can use to discover the types of information that the data contains.

You import the mapplets, rule specifications, and data domains to the Model repository. Informatica configures the objects to respond to the business rules that you might define for the organization data.

Note: The accelerators use the terms *mapplet* and *rule* to identify a mapplet. When you import the mapplets to the Model repository, the Developer tool creates the mapplet objects in a folder named *Rules*.

Informatica Data Quality includes a Core accelerator and a Core Data Domain accelerator. You can buy and download additional accelerators from Informatica.

Accelerator Structure

An accelerator is a compressed file that contains repository metadata files and other files in a directory structure. The directory structure depends on the type of accelerator. General accelerators contain rules, reference data objects, demonstration mappings, and demonstration data sources. Data Domain accelerators contain rules, reference data objects, data domains, and data domain groups.

General Accelerator Structure

General accelerators include the rules that analyze and enhance organization data and the sample mappings that demonstrate the rule operations. General accelerators also contain the reference data files and source data files that the rules and mappings use.

A general accelerator contains the following directories:

- Accelerator_Content
- Accelerator_Sources

Accelerator_Content Directory

The Accelerator_Content directory contains the following components:

Accelerator XML file

Contains metadata for rules, rule specifications, demonstration mappings, reference tables, and data objects.

Reference data file

Contains the reference data that the rules, rule specifications, and mappings use to identify different forms of data values. The reference data file is a compressed file that contains dictionary files in multiple directories. Specify the compressed file when you import the corresponding XML file. The import process copies the reference data to tables in the reference data database.

Note: If you export a mapping that contains a rule to PowerCenter®, copy the dictionary files to a directory that the PowerCenter Integration Service can read.

Accelerator_Sources Directory

The Accelerator_Sources directory contains the demonstration data file. The demonstration data file is a compressed file that contains the source data for the demonstration mappings. Copy the source data file to the file system.

Data Domain Accelerator Structure

Data domain accelerators include the data domains that determine the types of information in a data set and the rules that define the data domain logic. The accelerators also contain the reference data files that the data domains and rules use.

A data domain accelerator contains the following files:

Data domain metadata file

Contains metadata for the data domains and data domain groups that you add to the data domain glossary.

Rule metadata file

Contains metadata for the rules that define the data domain logic and for the reference data objects that the data domains use.

Reference data file for the data domains

Contains the reference data that a data domain uses when you run a profile that contains the data domain. The reference data file is a compressed file that contains dictionary files in multiple directories. Specify the compressed file when you import the corresponding XML file. The import process copies the reference data to tables in the reference data database.

Reference data file for the data domain rules

> Contains the reference data that a rule uses when you run a data domain that contains the rule. The reference data file is a compressed file that contains dictionary files in multiple directories. Specify the compressed file when you import the corresponding XML file. The import process copies the reference data to tables in the reference data database.

Accelerator Installation

To install an accelerator, import the repository object metadata to a Model repository project and copy the demonstration data files to the file system. Use the Developer tool to import the repository objects.

When you import rules and demonstration mappings, select the repository project from the Object Explorer. When you import data domains, select the repository project from the **Preferences** dialog box. In each case, the import operation prompts you to select the compressed file that contains the reference data that the XML file specifies.

General Accelerator Example

You might import the following metadata file for the Core accelerator:

```
Informatica_Core_Accelerator_1011.xml
```

When you import the metadata file, select the following reference data file:

```
Informatica_Core_Accelerator_1011.zip
```

Data Domain Accelerator Example

You might import the following metadata file for the Core Data Domain accelerator:

```
Informatica_IDE_DataDomain_1011.xml
```

When you import the metadata file, select the following reference data file:

```
Informatica_IDE_DataDomain_1011.zip
```

The following image shows the data domains in the **Preferences** dialog box:
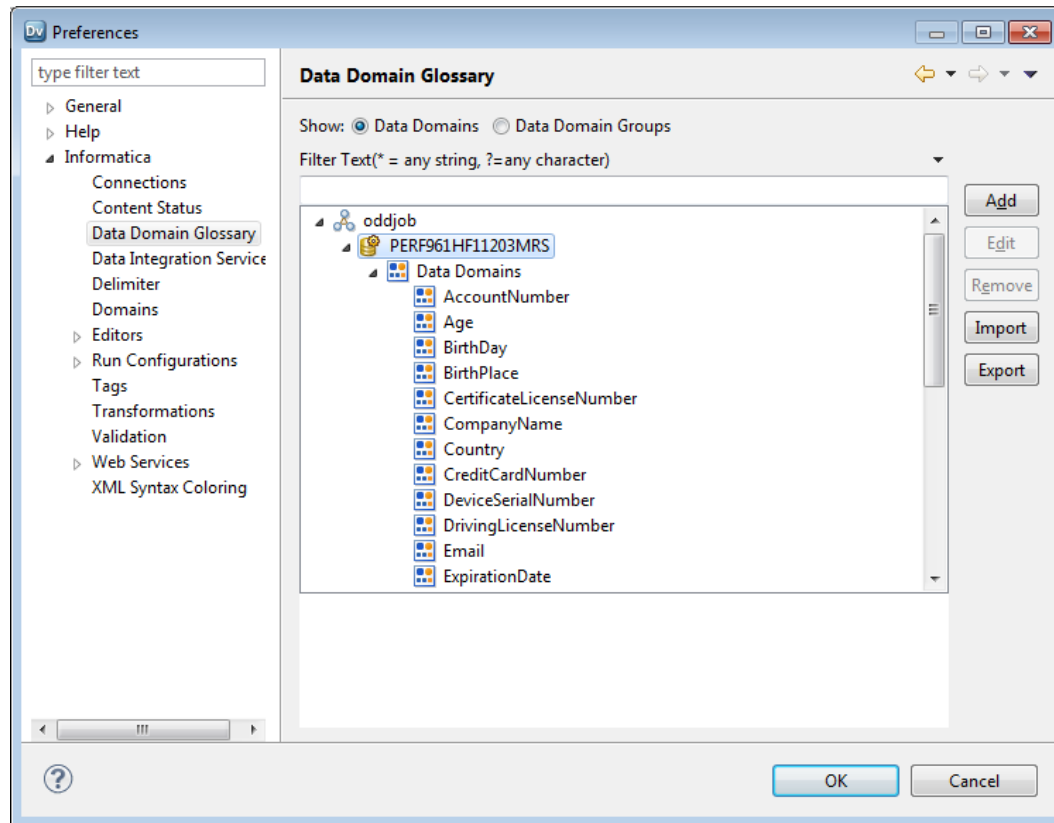
Source Data for Sample Mappings

When you import a general accelerator, copy the demonstration data files to the following directory on the Data Integration Service machine:

```
<Informatica Installation Directory>\services\DQContent\INFA_Content\demos\source_data
```

Rules and Guidelines for Accelerator Installation

The repository objects and data files in an accelerator operate in the same way as other objects and files in the Informatica system. Some rules and guidelines apply to the accelerator contents.

Consider the following rules and guidelines when you install an accelerator:

- Before you import or copy files, verify that you have all privileges on the Data Integration Service, the Content Management Service, and the Analyst Service.
- Import the accelerators to a single Model repository project. Create the project before you import the accelerators.
- Install the Core accelerator before you install another accelerator.
- Install the Core Data Domain accelerator before you install the Data Domain accelerator.
- If you import a metadata file that contains an object in common with an accelerator that you imported earlier, replace the object in the repository.

To use the accelerator rules that perform address validation, download and install the address reference data files for the country that the accelerator specifies. To use the accelerator rules that perform identity match analysis, download and install the identity population files for the country that the accelerator specifies. You buy the address reference data files and identity population files from Informatica.

Importing Rule and Mapping Objects

Use the Developer tool to import metadata for accelerator rules, demonstration mappings, and mapping data sources. During the import operation, select the reference data file that the rules and mappings use.

1. In the Developer tool, connect to the Model repository that contains the destination project for the metadata.

2. In the Object Explorer, select the destination project.

   For example, select the *Informatica_DQ_Content* project. If required, create a project in the Model repository.

3. Select **File** > **Import**.

4. In the **Import** dialog box, select **Informatica** > **Import Object Metadata File (Advanced)**.

5. Click **Next**.

6. Browse to the XML metadata file in the accelerator directory structure, and select the file.

7. Click **Open**, and click **Next**.

8. In the **Source** pane, select the items that appear under the project node.

9. In the **Target** pane, select the destination project.

10. Click **Add to Target**.

    - If the repository project contains an object that you want to add, the Developer tool prompts you to merge the object with the current object. Click **Yes** to merge the objects.

    - If the Developer tool prompts you to rename the objects, click **No**.

    - If any object remains in the **Source** pane, use the pointer to move the object to the target project.

11. Click **Next**.

12. Browse to the compressed reference data file in the accelerator directory structure, and select the file.

13. Click **Open**.

14. Verify that the code page is UTF-8, and click **Next**.

15. In the **Target Connection** field, select the reference data database.

16. Click **Finish**.

Importing Data Domains and Data Domain Groups

Use the Preferences dialog box to import metadata for data domains and data domain groups. During the import operation, select the reference data file that the data domains use.

1. In the Developer tool, connect to the Model repository that contains the destination project for the metadata.

2. Select **Window** > **Preferences.**

3. In the **Preferences** dialog box, expand the Informatica node and select **Data Domain Glossary**.

4. In the repository pane, select the top-level node for the data domains or the data domain groups.

5. Click **Import**.

6. Browse to the XML metadata file in the accelerator directory structure, and select the file.

7. Click **Open**, and click **Next**.

8. In the **Source** pane, select the data domain glossary project.

9. In the **Target** pane, select the destination project.

10. Select the following option in the Resolution field:

    Replace option in target

11. Click **Add Contents to Target**.

    - If the Developer tool prompts you to add the objects, click **Yes**.

    - If the Developer tool prompts you to rename the objects, click **No**.

12. Click **Next**.

13. If the import operation identifies dependencies, copy the dependent objects from the source project to the target project.

14. Click **Next**.

15. Browse to the compressed reference data file in the accelerator directory structure, and select the file.

16. Click **Open**.

17. Verify that the code page is UTF-8, and click **Next**.

18. In the **Target Connection** field, select the reference data database.

19. Click **Finish**.

Accelerator Components

When you import an accelerator, the Developer tool creates folders for the rules, data domains, and other objects that the accelerator specifies. Each folder contains subfolders that organize the objects by country and by the type of data quality operation that they perform.

Use the Core accelerator to create the folders in a repository project. When you import additional accelerators, you add objects and folders to the project.

The following image shows the Informatica_DQ_Content project folder structure when you import multiple accelerators to the project:
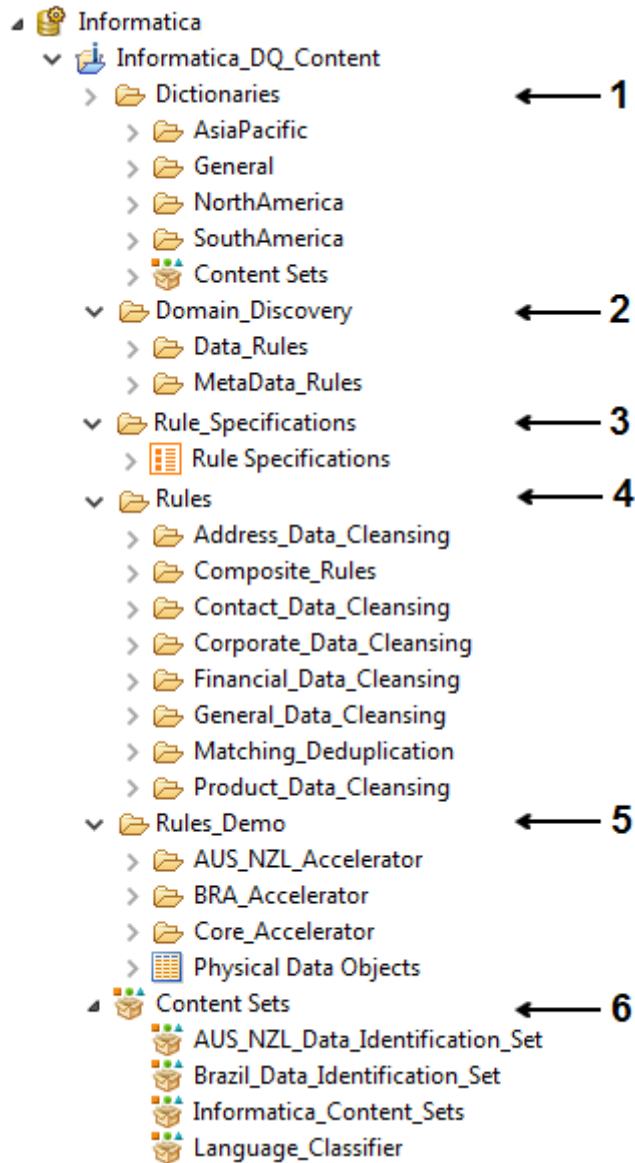
1. Dictionaries folder
2. Domain_Discovery folder
3. Rule_Specifications folder
4. Rules folder
5. Rules_Demo folder
6. Content Sets folder

The project contains the following top-level folders:

Dictionaries

The Dictionaries folder contains reference table objects. Each object refers to a table in the reference data database.

Domain_Discovery

The Domain_Discovery folder contains the rules that define the data domains in the accelerators that you install. The folder contains a Data_Rules folder and a Metadata_Rules folder. The rules in the Data_Rules folder correspond to the data domains that analyze column data values. The rules in the Metadata_Rules folder correspond to the data domains that analyze column names.

Rule_Specifications

The Rule_Specifications folder contains the rule specifications that you use to verify data against business rules.

Rules

The Rules folder contains the rules that you use to analyze and enhance data.

Rules_Demo

The Rules_Demo folder contains the demonstration mappings and demonstration data sources.

Content Sets

The Content Sets folder contains reference data objects that do not specify data in the reference data database.

Content Sets

A content set is a reference data object that does not store data in database tables. Content sets include character sets, pattern sets, regular expressions, token sets, probabilistic models, and classifier models.

The import operation adds the rules to the following repository folder:

```
[Informatica_DQ_Content]\Content Sets
```

Note: To view a list of the elements in a content set, open the content set in the Developer tool and select the **Tags** tab.

Data Domains

A data domain describes the data values that can represent a single type of business information in a column. Use data domains to determine the type of information in a column and to find information of a specified type in a column. The accelerators include data domains for a range of information types, including Social Security numbers, credit card numbers, email addresses, and job titles.

For example, a database table might contain Social Security numbers in a Comments column that any user can read. You must identify the records that contain the Social Security numbers and delete or move the Social Security numbers. You add the *SSN* data domain to a profile, and you run the profile on the Comments column.

You can assign a data domain to one or more data domain groups. Use the data domain groups to organize the data domains based on the type of business analysis that the data domains perform. The data domain glossary lists the data domains and data domain groups that you add to the Model repository. Use the **Preferences** menu in the Developer tool to add data domains to the data domain glossary. To update the data definitions in a data domain, use the rules in the data domain accelerator.

Note: You cannot view the data domain objects in the Object Explorer.

Demonstration Mappings

The demonstration mappings are run-time objects that apply one or more rules to a data source and write the results to another data source. You can use the demonstration mappings as templates for other mappings.

The import operation adds the mappings and data source objects to the following repository folder:

`[Informatica_DQ_Content]\Rules_Demo`

When you import an accelerator, the import operation adds the data source for the demonstration mappings to the `Rules_Demo` folder. Copy the data source files from the Accelerator_Sources directory to the file system.

Reference Tables

A reference table contains standard and alternative versions of a set of data values. Rules use reference tables to verify that data values are accurate and correctly formatted.

The import operation adds the reference tables to the following repository folder:

`[Informatica_DQ_Content]\Dictionaries`

Rule Specifications

A rule specification is an asset that represents the data requirements of a business rule in logical form.

The import operation adds the rule specifications to the following repository folder:

`[Informatica_DQ_Content]\Rule_Specifications`

Rule specifications are intended to answer the questions that commonly arise within business rules in an organization. For example, a rule specification might define one or more conditions that the business data must satisfy. Rule specifications and mapplets can be very similar, and they can use the same background data analysis and data transformation processes. Data analysts can create rule specifications in the Design workspace of the Analyst tool.

Rules

The accelerator rules define a range of data analysis and data transformation operations. You can add a single rule or a series of rules to a mapping.

Use accelerator rules to perform the following data quality tasks:

Address validation

Validate and enhance the data in postal address records. The rules require address reference data files.

Data parsing

Parse information from records. Parsing rules can extract multiple types of information, including person names, organization names, telephone numbers, dates, and identification numbers.

Data standardization

Standardize the spelling and format of data values. Standardization rules can identify and correct multiple types of information, including person names, organization names, telephone numbers, dates, and identification numbers.

Duplicate analysis

Find duplicate records in a data set. Duplicate analysis rules compare the records in a data set and generate a numeric score that represents the degree of similarity between the records.

The duplicate analysis rules can read records that contain general corporate data and records that contain identity data. The identity data rules require identity population data files.

The import operation adds the rules to the following repository folder:

```
[Informatica_DQ_Content]\Rules
```

Find the rules that perform address validation, data parsing, and data standardization operations in the *Data Cleansing* subfolders in the accelerator project. Find the rules that perform duplicate analysis in the *Matching Deduplication* subfolder in the accelerator project.

If you import rules for a country or region, you add a subfolder for composite rules. A composite rule combines multiple rules in a nested format in a single rule.

Tags and Rules

Accelerator rules include tags that indicate the type of data that the rule can read and the type of operation that the rule can perform.

To view the tags that apply to a rule, open the rule in the Developer tool and click the Tags tab. You can use the Search options in the Developer tool to find accelerators that contain a tag that you specify.

Accelerator Use in PowerCenter

You can export rules and mappings from the Model repository to the file system and to the PowerCenter repository. When you export the objects, select the reference tables, data objects, and other dependencies on the objects that you export.

The export operation copies the reference table data to the file system. Copy the files to the PowerCenter Integration Service host machine. The reference data file locations in the PowerCenter directory structure must correspond to the locations of the reference tables in the Model repository folder structure.

The following path describes a sample directory structure for the reference data objects in a PowerCenter installation:

```
<Informatica_installation_directory>\services\<Model_repository_project_name>
\<Model_repository_project_folder_name>
```

Note: If the PowerCenter product version does not match the Developer tool version, verify that the PowerCenter environment includes the Data Quality Integration Plug-in.

For more information about Data Quality integration with PowerCenter, read the Informatica Data Quality Integration for PowerCenter User Guide.

CHAPTER 2

Core Accelerator

This chapter includes the following topics:

Core Accelerator Overview

Use the rules in the Core accelerator to verify and enhance business data in any country or region.

The Core accelerator includes rules that perform the following data quality processes:

- Address data cleansing
- Contact data cleansing
- Corporate data cleansing
- General data cleansing
- Matching and deduplication data cleansing
- Product data cleansing

The Core accelerator contains mapplets and reference data objects that other accelerators can reuse. Install the Core accelerator before you install any other accelerator.

Core Address Data Cleansing Rules

Use the address data cleansing rules to parse, standardize, and validate address data.

Find the address data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
```

The following table describes the address data cleansing rules in the Core accelerator:

Name	Description
mplt_Global_AddressValidation5_v2_Discrete _Webservice	Validates postal addresses from multiple countries. Use the mapplet when you can connect the input address fields to the Discrete input ports on the Address Validator transformation. The mapplet calls an address validation web service. Use the mapplet as an example when you set up other web service mapplets.
mplt_Global_AddressValidation5_v2_Hybrid_ Webservice	Validates postal addresses from multiple countries. Use the mapplet when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation. The mapplet calls an address validation web service. Use the mapplet as an example when you set up other web service mapplets.
mplt_Global_AddressValidation5_v2_Multiline _Webservice	Validates postal addresses from multiple countries. Use the mapplet when you can connect the input address fields to the Multiline input ports on the Address Validator transformation. The mapplet calls an address validation web service. Use the mapplet as an example when you set up other web service mapplets.
rule_Calc_Distance_Between_Geocoordinates	Calculates the distance between two sets of geocoordinates.
rule_Country_Identification	Identifies a country.
rule_Country_Name_Standardization	Standardizes country names. The rule returns a country name, a two-character ISO country code, and a three-character ISO country code.
rule_Geoocordinate_In_Polygon	Verifies the presence of geocordinate points within an area that three or more geocordinate points define.
rule_Global_Address_Parse_Hybrid	Parses unstructured addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_Global_Address_Parse_Multiline	Parses unstructured addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_Global_Address_Validation_Discrete	Validates the deliverability of address records from multiple countries. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_Global_Address_Validation_Discrete_w_ Geocoding	Validates the deliverability of address records from multiple countries and adds latitude and longitude coordinates to each output addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_Global_Address_Validation_Hybrid	Validates the deliverability of address records from multiple countries. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.

Name	Description
rule_Global_Address_Validation_Hybrid_w_Geocoding	Validates the deliverability of address records from multiple countries and adds latitude and longitude coordinates to each output addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_Global_Address_Validation_Multiline	Validates the deliverability of address records from multiple countries. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_Global_Address_Validation_Multiline_w_Geocoding	Validates the deliverability of address records from multiple countries and adds latitude and longitude coordinates to each output addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.

Core Contact Data Cleansing Rules

Use the contact data cleansing rules to parse and validate data about business contacts and individuals.

Find the contact address data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

The following table describes the contact data cleansing rules in the Core accelerator:

Name	Description
rule_Email_Parse	Parses email addresses from data fields.
rule_Email_Parse_and_Validate	Parses email addresses from data fields and validates the format of each email address.
rule_Email_Parse_Into_Mailbox_Domain	Parses email addresses into mailbox, domain, and subdomain ports. For example, the rule parses *info@informatica.com* in the following manner: - Mailbox: info - Subdomain: informatica - Domain: com
rule_Email_Validation	Validates the format of email addresses. The rule does not verify that the email addresses are accurate or active. The rule returns Valid or Invalid.

Name	Description
rule_Identify_Suspect_Names	Identifies names that might not be genuine person names. The rule compares the input values to a reference table of names that are unlikely to be genuine. For example, the reference table includes the names of fictional characters.
rule_Phone_Country_Code_Validation	Verifies that the country code in a telephone number is correct for the given country name. When the country name identifies a United States territory, the rule verifies that the country code is correct and that the three-digit area code is a valid area code for a United States territory. For example, the rule validates 1340 as a country code and area code in the United States Virgin Islands.

Core Corporate Data Cleansing Rules

Use the corporate data cleansing rules in the Core accelerator to standardize corporate data.

Find the corporate data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing

The following table describes the corporate data cleansing rules in the Core accelerator:

Name	Description
rule_Company_Name_Standardization	Uses reference tables to standardize company names.

Core General Data Cleansing Rules

Use the general data cleansing rules to parse, standardize, and validate data.

Find the general data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\General_Data_Cleansing

The following table describes the general data cleansing rules in the Core accelerator:

Name	Description
mplt_Parse_Tokens_Into_Single_Field	Parses each word in a space-delimited string to a separate port.
rule_Add_Leading_Zero	Adds the numeral "0" to the beginning of a string.
rule_Add_Parentheses_At_Start_End_ofLine	Adds parenthetical symbols at the start and end of a string.
rule_Add_Plus_To_Start_of_Line	Adds the plus symbol at the start of a string.

Name	Description
rule_Add_Space_Around_Ampersand	Adds a space before and after all ampersands in a string.
rule_Add_Space_Around_Hyphen	Adds a space before and after all dashes and hyphens in a string.
rule_Add_Space_Between_Number_Letter	Adds a space in between a character pair composed of one numeral and one alphabetic character. Reading from left to right, the mapplet adds a space to the first numeral-alphabetic character pair in the data.
rule_Add_Spaces_Around_Period	Adds a space before and after all periods in a string.
rule_AllTrim	Removes all leading and trailing spaces from the input data fields.
rule_Assign_DQ_AddressResolutionCode_Description	Assigns a description to the Address Resolution Code output from the Address Validator transformation.
rule_Assign_DQ_ElementInputStatus_Description	Assigns a description to the Element Input Status output from the Address Validator transformation. The description corresponds to the output from Data Quality transformations in releases prior to Data Quality 9.0.
rule_Assign_DQ_ElementRelevance_Description	Assigns a description to the Element Relevance output from the Address Validator transformation. The description corresponds to the output from Data Quality transformations in releases prior to Data Quality 9.0.
rule_Assign_DQ_ElementResultStatus_Description	Assigns a description to the Element Result Status output from the Address Validator transformation. The description corresponds to the output from Data Quality transformations in releases prior to Data Quality 9.0.
rule_Assign_DQ_ExtendedElementStatus_Description	Assigns a description to the Extended Element Result Status output from the Address Validator transformation.
rule_Assign_DQ_GeocodingStatus_Description	Assigns a description to the Geocoding Status output from the Address Validator transformation. The description corresponds to the output from Data Quality transformations in releases prior to Data Quality 9.0.
rule_Assign_DQ_Mailability_Score_Description	Assigns a description to the Mailability Score output from the Address Validator transformation. The description corresponds to the output from Data Quality transformations in releases prior to Data Quality 9.0.
rule_Assign_DQ_Match_Code_Description	Assigns a description to the Match Code output from the Address Validator transformation. The description corresponds to the output from Data Quality transformations in releases prior to Data Quality 9.0.
rule_Classify_Language	Classifies a string as one of the following languages: Arabic, Dutch, English, French, German, Italian, Portuguese, Russian, Spanish, or Turkish. The rule uses the Language_Classifier content set to identify the languages. **Note:** The rule returns a language for every string that it analyzes. If a string belongs to a language that the rule does not recognize, the rule returns the language that most closely matches the text in the string.

Name	Description
rule_Compare_Dates	Calculates the difference between two dates. The mapplet uses the following units of measure: - Hours - Days - Months - Years Each output value is exclusive from the other values. The outputs cannot be added to represent the difference between the data values.
rule_Completeness	Checks a single port for NULL values. Returns "Complete" if the port contains data. Returns "Incomplete" if the port is empty or contains a NULL value.
rule_Completeness_Multi_Port	Checks multiple ports for NULL values. Returns "Complete" if all ports contain data. Returns "Incomplete" if any port is empty or contains a NULL value.
rule_Concatenate_Words	Concatenates two fields. Uses a character space as a separator.
rule_Convert_Match_Codes_to_Legacy_Values	Converts the output from the Match Code port in an Address Validator transformation to the equivalent address validation match code in Data Quality 8.6.
rule_CreditCard_Number_Validation	Validates credit card numbers for credit cards that use the Luhn algorithm. Validation includes, but is not limited to, the following credit cards: - American Express - Diners Club Carte Blanche - Diners Club International - Diners Club US & Canada - Discover Card - JCB - Maestro - Master Card - Solo - Switch - Visa - Visa Electron The rule returns "Valid" or "Invalid."
rule_Date_Complete	Verifies that the input string conforms to a date format that the rule recognizes. The rule reads the following reference data object: - user_defined_dates_infa
rule_Date_of_Birth_Validation	Checks the number of years between a date of birth and the current date. Returns "Adult" or "Minor" in addition to "Valid" if the number of years 120 or lower. Returns "Invalid" if the number of years is greater than 120.
rule_Date_Parse	Parses date data from a string to a port that the rule specifies. The rule recognizes dates in the following formats: - dd/mm/yyyy - mm/dd/yyyy - yyyy/dd/mm The rule returns a date and also returns a string that contains the input text without the date.

Name	Description
rule_Date_Standardization	Standardizes date strings to an output format that you specify. To set the output format, open the dq_FormatDate Expression transformation in the rule and update the Output_Date_Format expression variable and the Delimiter expression variable. If the input data does not describe a valid date, the rule returns the digit 0 for each input character.
rule_Date_Validation	Validates date strings that appear in a single format in a data column. To configure the date format that the rule uses for validation, open the dq_ValidateDate Expression transformation in the rule and update the In_Date_Format expression variable. The default format is "MM/DD/YYYY." The rule returns "Valid" or "Invalid."
rule_Date_Validation_Variable_Format	Validates date strings that appear in multiple formats in a data column. Use the rule when a data source includes the following columns: - A column that contains date values in multiple formats. - A column that identifies the format of the date value in each row. If the column does not identify a date format for a row, the rule applies the format "MM/DD/YYYY" to the date value. The rule reads all data values that the *is_date()* function recognizes. The rule returns "Valid" or "Invalid."
rule_Days_Between_Dates	Calculates the number of days between two dates.
rule_Days_From_Current_Date	Calculates the number of days between a specified date and the current date.
rule_EAN13_Algorithm	Validates an International Article Number. The rule returns "Valid" if the check digit is correct for the number and "Invalid" if the check digit is incorrect.
rule_GTIN_Validation	Validates a Global Trade Item Number (GTIN). The rule validates eight-dight, twelve-digit, thirteen-digit, and fourteen-digit numbers. The rule returns "Valid" if the check digit is correct for the number and "Invalid" if the check digit is incorrect.
rule_IsNumeric	Verifies that the input data is numeric. The rule returns "True" or "False."
rule_LowerCase	Returns all alphabetic characters in lower case.
rule_Luhn_Algorithm	Applies the Luhn algorithm to a numeric string. The rule can validate numeric strings, such as credit card numbers.
rule_Mask_Profanity	Checks input data for profanity. Masks profanity as "CENSORED" in the output data.
rule_Negative_Number_Validation	Validates that the input data is a negative number.
rule_Numeric_Completeness	Checks for NULL values in numeric inputs.

Name	Description

rule_Parse_Alpha_Chars_from_Non_Alpha_Chars	Identifies the alphabetic characters and the non-alphabetic characters in an input string and writes each set of characters to different output ports. For example, the rule parses the following values from the input string `teststring_123`:
	
	`testrtring`
	`_123`

rule_Parse_First_Word	Parses the first word in an input string to a port that the rule specifies.
rule_Parse_Number_At_End_Of_Line	Parses any number that occurs at the end of an input string to a port that the rule specifies. The rule reads strings from left to right.
rule_Parse_Number_At_Start_Of_Line	Parses any number that occurs at the start of an input string to a port that the rule specifies. The rule reads strings from left to right.
rule_Parse_Profanity	Compares strings to a reference table of profane terms and parses any term that matches a reference table value to a port that the rule specifies.
rule_Parse_Text_Between_Parentheses	Parses strings that are enclosed in parentheses to a port that the rule specifies. The rule contains an output port for the parsed strings and an output port for the input text without the parsed strings.
rule_Parse_Text_in_Single_Quotes	Parses strings that are enclosed in quotation marks to a port that the rule specifies. When the input data contains multiple quoted elements, the rule parses the final element. The rule reads the input strings from left to right. The rule contains an output port for the parsed strings and an output port for the input text without the parsed strings.
rule_Past_Date_Label	Determines whether an input date is earlier than the system date or later than the system date.
rule_Personal_Company_Identification	Parses person names and company names to different ports that the rule specifies. The rule has the following outputs: - Person name - Company name - Data category, such as person name or company name - Data that the rule cannot parse
rule_Postive_Number_Validation	Verifies that the input data is a positive number.
rule_Prepend_Zero_to_Single_Digit	Prepends the numeral "0" to single numeric characters.
rule_Remove_All_Leading_Zeros	Removes all instances of the numeric character "0" from the beginning of a string.
rule_Remove_Apostrophe	Removes apostrophes. The rule merges the text strings on either side of the apostrophe.
rule_Remove_Control_Characters	Removes control characters from text strings. The rule returns a string that contains the control characters and a string that contains the input text without the control characters.

Name	Description
rule_Remove_Extra_Spaces	Replaces all consecutive spaces with a single space and trims leading and trailing spaces.
rule_Remove_Hyphen	Removes hyphens.
rule_Remove_Leading_Zero	Removes a single instance of the numeric character "0" from the beginning of a string.
rule_Remove_Limited_Punctuation	Removes extraneous characters. Extraneous characters include slashes, back slashes, periods, exclamation marks, underscores, and multiple consecutive spaces.
rule_Remove_Non_Numbers	Removes all characters that are not numeric.
rule_Remove_Parentheses	Removes right and left parenthesis symbols.
rule_Remove_Period	Removes periods.
rule_Remove_Period_Parentheses	Removes the following characters: - Left and right parentheses - Periods
rule_Remove_Punctuation	Removes punctuation symbols.
rule_Remove_Punctuation_and_Space	Removes all punctuation and all space characters.
rule_Remove_Quotation	Removes quotation marks.
rule_Remove_Slashes	Removes forward slashes and back slashes.
rule_Remove_Space	Removes all character spaces.
rule_Replace_Ampersand_With_Space	Replaces ampersands with spaces.
rule_Replace_Hyphen_Underscore_with_Space	Replaces hyphens and underscores with spaces.
rule_Replace_Hyphen_with_Space	Replaces hyphens with spaces.
rule_Replace_Limited_Punct_with_Space	Replaces the following punctuation characters with a single space: dash, back slash, period, exclamation mark, and underscore. The rule also replaces two, three, and four consecutive spaces with a single space.
rule_Replace_Non_Alphabetic_with_Space	Replaces numerals and punctuation characters with a single space.
rule_Replace_Period_With_Space	Replaces periods with a single space.
rule_Replace_Punctuation_with_Space	Replaces all punctuation with spaces.
rule_Replace_Slashes_With_Space	Replaces forward slashes and back slashes with spaces.
rule_Reverse_String_Input	Reverses the order of characters in input strings.

Name	Description
rule_String_Completeness	Checks a string for completeness. The rule also searches the input strings for values in the reference table string_default_values_infa. The reference table contains values such as NA, DEFAULT, and XX. If an input string contains a value in the reference table, the rule identifies the string as incomplete.
rule_TitleCase	Converts strings to title case. In title case strings, the first letter of each word is capitalized.
rule_Translate_Diacritic_Characters	Replaces diacritic characters with ASCII equivalents. For example, the rule converts "ã" to "a".
rule_UpperCase	Returns all alphabetic characters in upper case.
rule_URL_Validation	Validates the format and structure of a URL.
rule_Years_Since_Date_of_Birth	Calculates the number of years since the input date.

Core Matching and Deduplication Rules

Use the matching and deduplication rules to identify duplicate records.

Find the matching and deduplication rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Matching_Deduplication

The following table describes the matching and deduplication rules in the Core accelerator:

Name	Description
mplt_Consolidate_and_Remove_Duplicate_Rows	Consolidates clusters of duplicate records into a single record and removes the redundant duplicate records.

Core Product Data Cleansing Rules

Use the product data cleansing rules to parse, standardize, and validate product data.

Find the product data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Product_Data_Cleansing

The following table describes the product data cleansing rules in the Core accelerator:

Name	Description
rule_Color_Parse	Parses color values to a port that the rule specifies.
rule_Parse_Quantity_And_UOM	Parses the first instance of a quantity and a unit of measure from a string to a port that the rule specifies. The rule reads the string from left to right and returns the following data: - Quantity. - Unit of measure. - The input string without the quantity and unit of measure values.
rule_UOM_Standardization	Standardizes a unit of measure. The rule returns standardized and unstandardized values for quantity and unit of measure. It also returns a string that contains the input text with a standardized unit of measure.
rule_UPC_Validation	Validates a Universal Product Code and returns a standardized Universal Product code.

Core Demonstration Mappings

The demonstration mappings in the Core accelerator use multiple rules to demonstrate data quality processes.

Find the demonstration mappings in the following repository location:

    [Informatica_DQ_Content]\Rules_Demo\Core_Accelerator

The accelerator contains the following demonstration mappings:

m_customer_data_demo

Parses, standardizes, and validates United States and Canadian data.

m_product_demo

Parses product descriptions and validates the quality of the descriptions.

CHAPTER 3

Data Domains Accelerator

This chapter includes the following topics:

Data Domains Accelerator Overview

A data domain is a predefined or user-defined Model repository object that uses rules to discover the functional meaning of column data or column names. The data domain rules define data patterns and column name patterns that match source data and metadata. You can use the data domain rules to update the data domain logic.

Use the data domains in the Data Domains accelerator to discover the functional meaning of source data based on column names or column data values.

The Data Domains accelerator includes the following types of rule:

- Data rule. Finds columns with data that matches the logic that the rule defines.
- Column name rule. Finds columns with column names that match column-name logic that the rule defines.

The data domain rules return Boolean values that indicate whether the column data or column name meets the rule criteria. The data domain rules use regular expressions or reference tables to look for specific values or patterns. For example, you can use a nine-digit rule expression to find data values in the Social Security number format.

When you use expressions in data domain rules, some unrelated data values might also meet the rule expression criteria. For example, United States ZIP codes in the source data might meet the Social Security number format. To make the data domain inference effective, review the data domain discovery results for discrepancies. After you review and verify the data domain discovery results, you can decide to associate a data domain with a data column.

Data Domains in the Data Domains Accelerator

Use the predefined data domains in profiles to perform data domain discovery and identify critical data characteristics within an enterprise.

Note: In the table, the asterisk (*) symbol is a wildcard character.

The following table describes the data domains available in the Data Domains accelerator:

Name	Description	Dependent Rule Type	Data Domain Group
Account_Status	Discovers columns that contain data that matches the account status values in a reference table.	Data rule	Account_Bank
AccountNumber	Discovers column names that contain the "a*c*num" or "acc" string.	Metadata rule	Account_Bank, PCI, PHI
Address	Discovers columns that contain address data values, such as cities, counties, countries, country codes, address prefixes, or address suffixes. Supports addresses from Canada, France, Germany, Great Britain, New Zealand, Portugal, and the United States. Discovers column names that can indicate or match terms for address data.	Data rule Metadata rule	General, PHI, PII
Admission_date	Discovers date values in the column data and standardizes the date values to a single format. Discovers column names that can indicate or match terms for admission date. See also Discharge_date.	Data rule Metadata rule	PHI
Age	Discovers column data with values from 1 through 120. Discovers column names that contain the "age" string or "dob" string.	Data rule Metadata rule	PHI, PII
AlphaNumeric_SpecialCharacter	Discovers column data that contains unformatted alphanumeric data and special-character data.	Data rule	General
Austria_NationalID	Discovers column data that matches the format of national identity numbers in Austria. Discovers column names that can indicate or match terms for national identification data.	Data rule Metadata rule	NationalID

Name	Description	Dependent Rule Type	Data Domain Group
Bank_Routing_Number_ABA_Number	Discovers numbers in column data that match the format of the American Banking Association routing number. Discovers column names that match the indicators that the metadata rule defines for American Banking Association routing numbers.	Data rule Metadata rule	General
BIC_SwiftCode	Discovers column data that matches Bank Identifier Code (BIC) or Society for Worldwide Interbank Financial Telecommunication (SWIFT) code data. The data rule uses pattern recognition and country code validation to find the data values. Discovers column names that match the indicators that the metadata rule defines for BIC data or SWIFT code data.	Data rule Metadata rule	Account_Bank
BinaryValue	Discovers column data that contains binary values.	Data rule	General
BirthDay	Discovers column data that matches valid birth dates. A birth date is valid for an age up to 120 years. Discovers column names that contain the "dob" string, "date*of*bir*" string, or "birth*da*" string.	Data rule Metadata rule	PHI, PII
BirthPlace	Discovers column data that matches country names in the reference data. Discovers column names that contain the "birth*place" string or "location*birth" string.	Data rule Metadata rule	PHI, PII
Brazil_IDDoc	Discovers column data that matches the format of the national identity number in Brazil.	Data rule	NationalID
Brazil_NationalID_CPF	Discovers column data that matches the format of the individual taxpayer registry identification number in Brazil. Discovers column names that can indicate or match terms for national identification data.	Data rule Metadata rule	NationalID
Brazil_NationalID_RG	Discovers column data that matches the format of national identity numbers (*Registro Geral* or *Carteira de Identidade*) in Brazil. Discovers column names that can indicate or match terms for national identification data.	Data rule Metadata rule	The rule does not belong to a data domain group.

Name	Description	Dependent Rule Type	Data Domain Group
Bulgaria_NationalID	Discovers column data that matches the format of national identity numbers in Bulgaria. Discovers column names that can indicate or match terms for national identification data.	Data rule Metadata rule	NationalID
Canada_Driver_License_Narrow	Discovers column data that matches the format of driver license numbers in Canada, with the following exceptions: - License numbers from British Columbia, Quebec, Manitoba, and Prince Edward Island. - License numbers with four, five, six, seven, or eight digits. Discovers column names that can indicate or match terms for driver license data.	Data rule Metadata rule	The rule does not belong to a data domain group.
Canada_SIN	Discovers column data that matches the Social Insurance number format in Canada. Discovers column names that can indicate or match terms for Social Insurance number data.	Data rule Metadata rule	NationalID
CertificateLicenseNumber	Discovers column names that contain the "cert*lic*number" string, "cert*lic*no*" string, "lic* nu*" string, or "lic*no*" string.	Metadata rule	PHI
China_NationalID	Discovers column data that matches the format of national identity numbers in China. Discovers column names that can indicate or match terms for national identity numbers.	Data rule Metadata rule	NationalID
City	Discovers column data that matches city names around the world. Discovers column names that can indicate or match terms for city names.	Data rule Metadata rule	PHI
CompanyName	Discovers column data that matches the organization name values in a reference table. Discovers column names that contain the "company" string.	Data rule Metadata rule	Contact, PII
ComputerAddress	Discovers column data that matches the format of IP addresses or MAC addresses. Discovers column names that can indicate or match terms for IP addresses or MAC addresses.	Data rule Metadata rule	General

Name	Description	Dependent Rule Type	Data Domain Group
Country	Discovers column data that matches country names in the reference data. Discovers column names that contain the "iso*countr*code" string, "iso*country" string, or "countr*" string.	Data rule Metadata rule	Address, PHI, PII
CountryCode_Phone	Discovers telephone numbers in column data based on international dialing codes. Discovers column names that can indicate or match terms for telephone numbers.	Data rule Metadata rule	Contact
CreditCard_AMEX	Discovers column data that matches the American Express credit card number format.	Data rule	Account_Bank
CreditCard_DinersCard	Discovers column data that matches the Diners Club International credit card number format.	Data rule	Account_Bank
CreditCard_DiscoverCard	Discovers column data that matches the Discover credit card number format.	Data rule	Account_Bank
CreditCard_JCB	Discovers column data that matches the JCB International credit card number format.	Data rule	Account_Bank
CreditCard_MasterCard	Discovers column data that matches the MasterCard credit card number format.	Data rule	Account_Bank
CreditCard_Visa	Discovers column data that matches the Visa credit card number format.	Data rule	Account_Bank
CreditCardNumber	Discovers column data that matches the format of credit card numbers from major credit card organizations. Discovers column names that contain the "ccn" string, "cr*ca*nu" string, or "credit*no*" string.	Data rule Metadata rule	Account_Bank, PCI, PII
Croatia_NationalID	Discovers column data that matches the format of national identity numbers in Croatia. Discovers column names that can indicate or match terms for national identity numbers.	Data rule Metadata rule	NationalID
Currency	Discovers column data that contains a currency indicator. Discovers column names that can indicate or match terms for currency data.	Data rule Metadata rule	Not applicable
Date_AllFormats	Discovers date values in the column data.	Data rule	General

Name	Description	Dependent Rule Type	Data Domain Group
Date_MM_DD_YYYY	Discovers date values in the column data that appear in a single format. The default format is MM/DD/YYYY.	Data rule	General
Denmark_NationalID	Discovers column data that matches the format of national identity numbers in Denmark. Discovers column names that can indicate or match terms for national identity numbers.	Data rule Metadata rule	NationalID
DeviceSerialNumber	Discovers column names that contain the "device*number" string, "device*no*" string, "serial*number" string, "serial*no*" string, or "device*identi*" string.	Metadata rule	PHI
Discharge_date	Discovers date values in the column data and standardizes the date values to a single format. Discovers column names that can indicate or match terms for discharge dates. See also Admission_date.	Data rule Metadata rule	PHI
DriverLicense_Canada	Discovers column data that matches the format of driver license numbers in Canada, with the exception of licenses from British Columbia, Quebec, Manitoba, and Prince Edward Island. Discovers column names that can indicate or match terms for driver license data.	Data rule Metadata rule	NationalID
DriverLicense_US_Canada_GBR	Discovers column data that matches the format of driver license numbers from the United Kingdom and from many states and provinces in Canada and the United States. The data domain does not discover the following licenses: - License numbers from British Columbia, Quebec, Manitoba, and Prince Edward Island. - License numbers with four, five, six, seven, or eight digits. Discovers column names that can indicate or match terms for driver license data.	Data rule Metadata rule	Not applicable
DriversLicense_GBR	Discovers column data that matches the format of United Kingdom driver license numbers. Discovers column names that can indicate or match terms for driver license data.	Data rule Metadata rule	NationalID

Name	Description	Dependent Rule Type	Data Domain Group
DriversLicense_USA	Discovers column data that matches the driver license number format in most of the states in the United States. Discovers column names that can indicate or match terms for driver license data.	Data rule Metadata rule	NationalID, PHI
DriversLicense_USA _Narrow	Discovers column data that matches the driver license number format in a majority of the states in the United States. In order to reduce false positives, the data domain does not cover states in which the license format overlaps with other common formats, such as eight-digit date formats. Discovers column names that can indicate or match terms for driver license data.	Data rule Metadata rule	PHI
DrivingLicenseNumber	Discovers column data that matches driver license numbers in the United Kingdom, in most states of the Unites States, and in Canadian provinces with the exception of British Columbia, Manitoba, Prince Edward Island, and Quebec. Discovers column names that can indicate or match terms for driver license data.	Data rule Metadata rule	PII
Email	Discovers column data that matches a predefined email identification format. Discovers column names that contain the "email" string.	Data rule Metadata rule	Contact, PHI, PII
ExpirationDate	Discovers column data that matches expired credit card dates. Discovers column names that contain the "exp*da*" string or "cr*exp*" string.	Data rule Metadata rule	PCI
Finland_NationalID	Discovers column data that matches the format of national identity numbers in Finland. Discovers column names that can indicate or match terms for national identity numbers.	Data rule Metadata rule	NationalID
FirstName	Discovers column data that matches values in a reference table of first names. Discovers column names that contain the "f*nam*" string.	Data rule Metadata rule	Contact, PCI, PHI, PII
France_INSEE	Discovers column data that matches the French Institut National de la Statistique et des Études Économiques (INSEE) number format.	Data rule	NationalID

Name	Description	Dependent Rule Type	Data Domain Group
FullName	Discovers strings in column data that match first, middle, and last names in the reference data. Discovers column names that can indicate or match terms for full names.	Data rule Metadata rule	Contact, PCI, PHI, PII
Gender	Discovers column data that matches the gender values in a reference table. Discovers column names that contain the "gender" string or strings such as "female" and "male".	Data rule Metadata rule	Contact, PHI, PII
Geocode_Latitude	Discovers column data that matches valid latitude coordinates. Discovers column names that contain the "latitude" string.	Data rule Metadata rule	Address, General
Geocode_LatitudeLongitude	Discovers column data that matches valid latitude or longitude coordinates. Discovers column names that contain strings such as "latitude," "longitude," and "geocode".	Data rule Metadata rule	Address, General
Geocode_Longitude	Discovers column data that matches valid longitude coordinates. Discovers column names that contain the "longitude" string.	Data rule Metadata rule	Address, General
Grade	Discovers column names that contain the "grade" string.	Metadata rule	PII
GreatBritain_NINO	Discovers column data that matches the Great Britain National Insurance number format. Discovers column names that can indicate or match terms for National Insurance numbers.	Data rule Metadata rule	NationalID
Health_Plan_Beneficiary_Number	Discovers column names that can indicate or match terms for health plan beneficiary numbers.	Metadata rule	PHI
Height	Discovers column data with values in a range from 1 through 8, where 8 represents feet in height. Discovers column names that can indicate or match terms for height information.	Data rule Metadata rule	PHI
Hostname	Discovers column data that matches valid host names. Discovers column names that can indicate or match terms for host names.	Data rule Metadata rule	General

Name	Description	Dependent Rule Type	Data Domain Group
IBAN	Discovers column data that matches the International Bank Account Number (IBAN) format of multiple European countries. Discovers column names that can indicate or match terms for IBAN values.	Data rule Metadata rule	Account_Bank
ICD_10_Codes	Discovers column data that matches the codes for medical conditions in the tenth revision of the International Statistical Classification of Diseases and Related Health Problems (ICD). Discovers column names that can indicate or match terms for the ICD codes.	Data rule Metadata rule	PHI
ICD_9_Codes	Discovers column data that matches the codes for medical conditions in the ninth revision of the International Statistical Classification of Diseases and Related Health Problems (ICD). Discovers column names that can indicate or match terms for the ICD codes.	Data rule Metadata rule	PHI
India_NationalID	Discovers column data that matches the format of the Indian Permanent Account Number. Discovers column names that can indicate or match terms for national identity numbers.	Data rule Metadata rule	NationalID
IPAddress	Discovers column data that matches a predefined IP address format. Discovers column names that contain the "ip" string or "inter*port*add" string.	Data rule Metadata rule	General, PII
ISBN	Discovers column data that matches the International Standard Book Number format. Discovers column names that can indicate or match terms for International Standard Book Numbers.	Data rule Metadata rule	General
Italy_FiscalCode	Discovers column data that matches the format of the *Codice Fiscale* or Fiscal Code number in Italy.	Data rule	NationalID
JobPosition	Discovers column data that includes job position descriptors or job titles. Discovers column names that contain the "title" string, "position" string, or "designation" string.	Data rule Metadata rule	PHI, PII

Name	Description	Dependent Rule Type	Data Domain Group
Korea_NationalID	Discovers column data that matches the format of national identity numbers in South Korea. Discovers column names that can indicate or match terms for national identity numbers.	Data rule Metadata rule	NationalID
LastName	Discovers column data that matches values in a reference table of last names. Discovers column names that can indicate or match terms for last name information.	Data rule Metadata rule	Contact, PCI, PHI, PII
MaidenName	Discovers strings in column data that match last names in the reference data. Discovers column names that can indicate or match terms for maiden name information.	Data rule Metadata rule	Contact, PHI, PII
MiddleName	Discovers strings in column data that match first names in the reference data. Discovers column names that can indicate or match terms for a column of middle name information.	Data rule Metadata rule	Contact, PHI, PII
Norway_NationalID	Discovers column data that matches the format of national identity numbers in Norway. Discovers column names that can indicate or match terms for national identity numbers.	Data rule Metadata rule	NationalID
Passport_DEU_MR	Discovers column data that matches the machine-readable German passport number format.	Data rule	NationalID
Passport_GBR	Discovers column data that matches the United Kingdom passport number format.	Data rule	NationalID
Passport_India	Discovers column data that matches the India passport number format.	Data rule	NationalID
Passport_MachineReadable	Discovers column data that matches machine-readable passport numbers of all countries.	Data rule	NationalID
Passport_USA_MR	Discovers column data that matches machine-readable United States passport number format. Discovers column names that can indicate or match terms for a column of passport numbers.	Data rule Metadata rule	NationalID, PHI

Name	Description	Dependent Rule Type	Data Domain Group
PhoneNumber	Discovers column data that matches the United States phone number format. Discovers column names that contain the "phone" string or "fax" string.	Data rule Metadata rule	Contact, PHI
Postcode	Discovers column data that matches the postal codes of multiple countries. Discovers column names that can describe a column of post codes.	Data rule Metadata rule	Address, PCI
Race	Discovers column data that matches the name of a race of people in the reference data. Discovers column names that can indicate or match terms for a race of people.	Data rule Metadata rule	PHI
Religion	Discovers column data that matches the name of a religion in the reference data. Discovers column names that can indicate or match terms for religion or religious affiliation.	Data rule Metadata rule	PHI
Romania_NationalID	Discovers column data that matches the format of national identity numbers in Romania. Discovers column names that can indicate or match terms for national identity numbers.	Data rule Metadata rule	NationalID
Salary	Discovers column data that can represent an amount of money. Discovers column names that can indicate or match terms for compensation, salary, or wages.	Data rule Metadata rule	PII
SouthAfrica_NationalID	Discovers column data that matches the format of national identity numbers in South Africa. Discovers column names that can indicate or match terms for national identity numbers.	Data rule Metadata rule	NationalID
SSN	Discovers column data that matches the United States Social Security number format. Discovers column names that contain the "SSN" string, "social*sec*no" string, or "social* sec*num*" string.	Data rule Metadata rule	NationalID, PHI

Name	Description	Dependent Rule Type	Data Domain Group
SSN_General	Discovers column data that matches the Social Security number format. Based on the SSN Randomization initiative of June 2011, the data domain does not verify the group and area number combination in a column data value. Discovers column names that contain the "SSN" string, "social*sec*no" string, or "social* sec*num*" string.	Data rule Metadata rule	NationalID, PII
State	Discovers column data that matches a state name in the United States. Discovers column names that contain the "add*sta" string, "state" string, or "us*sta*" string.	Data rule Metadata rule	PII
Street	Discovers column data that includes any street descriptor in a set of over twenty descriptors that the data rule specifies. The rule includes the following descriptors: - avenue - boulevard - junction - lane - road - street - valley - way Discovers column names that can indicate or match terms for a column of street address information.	Data rule Metadata rule	PII
Sweden_NationalID	Discovers column data that matches the format of national identity numbers in Sweden. Discovers column names that can indicate or match terms for national identity numbers.	Data rule Metadata rule	NationalID
Taiwan_NationalID	Discovers column data that matches the format of national identity numbers in Taiwan. Discovers column names that can indicate or match terms for national identity numbers.	Data rule Metadata rule	NationalID
Track1_Format_B	Discovers column data that matches Track 1 Format B credit card data. Discovers column names that can indicate or match terms for Track 1 Format B information.	Data rule Metadata rule	PCI

Name	Description	Dependent Rule Type	Data Domain Group
UniqueIdentifyingNumber	Discovers column names that contain the "unique*iden*number" string or "iden*num" string.	Metadata rule	PHI
Unit_of_Measure	Discovers column data that includes units of measurement. Discovers column names that can indicate or match terms for quantity or units of measurement.	Data rule Metadata rule	Not applicable
UPC	Discovers column data that matches the format of a valid Universal Product Code or European Article Number. Discovers column names that can indicate or match terms for Universal Product Codes or European Article Numbers.	Data rule Metadata rule	General
URL	Discovers column data that matches predefined URL formats. Discovers column names that contain the "uni*res*loc" string, "URL" string, or "web" string.	Data rule Metadata rule	PHI
USA_Bank_Account	Discovers column data that matches a bank account number format in the United States. Discovers column names that can indicate or match terms for bank account information.	Data rule Metadata rule	PHI, PII
USA_County	Discovers column data that matches a county name in the United States. Discovers column names that can indicate or match terms for county name information.	Data rule Metadata rule	Address, PHI, PII
USA_National_Drug_Code_NDC	Discovers column data that matches a National Drug Code (NDC) value in the National Drug Code directory in the United States. Discovers column names that can indicate or match terms for National Drug Code or NDC information.	Data rule Metadata rule	PHI
USA_National_Provider_Identifier_Standard_NPI	Discovers column data that matches a National Provider Identifier (NPI) number in the United States. Discovers column names that can indicate or match terms for National Provider Identifier or NPI information.	Data rule Metadata rule	PHI

Name	Description	Dependent Rule Type	Data Domain Group
USA_Taxpayer_Identification_Number	Discovers column data that matches the format of an Individual Taxpayer Identification Number (ITIN) in the United States. Discovers column names that can indicate or match terms for Individual Taxpayer Identification Number or ITIN information.	Data rule Metadata rule	NationalID, PHI
USZip_5digit	Discovers column data that matches United States ZIP Code values. Discovers column names that can indicate or match terms for ZIP or post codes.	Data rule Metadata rule	Address, PHI, PII
VehicleRegPlateNumber	Discovers column names that contain the "registration" string, "number*plate" string, "license*plate" string, or "vehicle*registration" string.	Metadata rule	PHI, PII
Weight	Discovers column data with values in a range from 1 through 500, where 500 represents a weight value. Discovers column names that can indicate or match terms for weight information.	Data rule Metadata rule	PHI
ZipCode	Discovers column data that matches United States ZIP codes. Discovers column names that can indicate or match terms for ZIP or post codes.	Data rule Metadata rule	PCI, PHI, PII

A Note on Data Domain Group Name Abbreviations

The data domain group name PCI is an abbreviation for Payment Card Industry Information.

The data domain group name PHI is an abbreviation for Protected Health Information.

The data domain group name PII is an abbreviation for Personally Identifiable Information.

Column Name Rules in the Data Domains Accelerator

Use the data domain column name rules to identify data columns with names that match the column name logic that the rules define. Each rule uses one or more regular expressions to search for common strings that the column name might include.

For example, the rule *dataDomain_MetaDataRule_BIC_SWIFTCode* contains a Labeler transformation that searches with the following regular expressions:

```
^*[iI][sS][oO].*.[9][3][6][2].*$
^*[sS][wW][iI][fF][tT]*[bB][iI][cC]$
^*{bB][iI][cC].*[cC][oO][dD][eE].*$
```

The column name rules analyze the characters in the column names. The column name rules do not analyze the data values in the columns.

You can find the column name rules in the following repository location:

    [Informatica_DQ_Content]\Domain_Discovery\MetaData_Rules

The following table describes the column name rules in the Data Domains accelerator:

Name	Description
dataDomain_MetaDataRule_ABARoutingNumber	Searches for names that can describe a column of American Banking Association routing numbers.
dataDomain_MetaDataRule_AccountNumber	Searches for names that can describe a column of account numbers.
dataDomain_MetaDataRule_AccountStatus	Searches for names that can describe a column of account status information.
dataDomain_MetaDataRule_Address	Searches for names that can describe a column of address information.
dataDomain_MetaDataRule_AdmissionDate	Searches for names that can describe a column of admission date information.
dataDomain_MetaDataRule_Age	Searches for names that can describe a column of age or date of birth information.
dataDomain_MetaDataRule_BankAccount	Searches for names that can describe a column of bank account information.
dataDomain_MetaDataRule_BIC_SwiftCode	Searches for names that can describe a column of Business Identifier Codes. Business Identifier Codes are also called SWIFT codes and ISO 9362 codes.
dataDomain_MetaDataRule_BirthDay	Searches for names that can describe a column of date of birth or birthday information.
dataDomain_MetaDataRule_BirthPlace	Searches for names that can describe a column of place or location of birth information.
dataDomain_MetaDataRule_CertificateLicenseNumber	Searches for names that can describe a column of certificate license number information.
dataDomain_MetaDataRule_City	Searches for names that can describe a column of city information.
dataDomain_MetaDataRule_CompanyName	Searches for names that can describe a column of company name information.
dataDomain_MetaDataRule_Computer_Address	Searches for names that can describe a column of machine or MAC address data.
dataDomain_MetaDataRule_Country	Searches for names that can describe a column of country information, including ISO country code information.
dataDomain_MetaDataRule_CountryCode_Phone	Searches for names that can describe a column of telephone country codes.
dataDomain_MetaDataRule_County	Searches for names that can describe a column of county information.

Name	Description
dataDomain_MetaDataRule_CreditCardNumber	Searches for names that can describe a column of credit card numbers.
dataDomain_MetaDataRule_CreditCardTrack1FormatB	Searches for names that can describe a column of track 1 format B information from a credit card.
dataDomain_MetaDataRule_Currency	Searches for names that can describe a column of currency information.
dataDomain_MetaDataRule_DeviceSerialNumber	Searches for names that can describe a column of device number or serial number information.
dataDomain_MetaDataRule_DischargeDate	Searches for names that can describe a column of discharge date information.
dataDomain_MetaDataRule_DrivingLicenseNumber	Searches for names that can describe a column of drivers license information.
dataDomain_MetaDataRule_Email	Searches for names that can describe a column of email information.
dataDomain_MetaDataRule_ExpirationDate	Searches for names that can describe a column of expiration date information, for example expiration date information for credit cards.
dataDomain_MetaDataRule_FirstName	Searches for names that can describe a column of first name information.
dataDomain_MetaDataRule_FullName	Searches for names that can describe a column of full name information.
dataDomain_MetaDataRule_Gender	Searches for names that can describe a column of gender information.
dataDomain_MetaDataRule_Grade	Searches for names that can describe a column of grade information.
dataDomain_MetaDataRule_HealthCareBeneficiaryNumber	Searches for names that can describe a column of health care beneficiary numbers.
dataDomain_MetaDataRule_Height	Searches for names that can describe a column of height information.
dataDomain_MetaDataRule_Hostname	Searches for names that can describe a column of computer host name information.
dataDomain_MetaDataRule_IBAN	Searches for names that can describe a column of International Bank Account Numbers.
dataDomain_MetaDataRule_ICD_10	Searches for names that can describe a column of values from the tenth revision of the International Statistical Classification of Diseases and Related Health Problems.
dataDomain_MetaDataRule_ICD_9	Searches for names that can describe a column of values from the ninth revision of the International Statistical Classification of Diseases and Related Health Problems.
dataDomain_MetaDataRule_IPAddress	Searches for names that can describe a column of computer IP address information.

Name	Description
dataDomain_MetaDataRule_ISBN	Searches for names that can describe a column of International Standard Book Numbers.
dataDomain_MetaDataRule_ITIN_USA	Searches for names that can describe a column of Individual Taxpayer Identification Numbers.
dataDomain_MetaDataRule_JobPosition	Searches for names that can describe a column of job title, position, or designation information.
dataDomain_MetaDataRule_LastName	Searches for names that can describe a column of last name information.
dataDomain_MetaDataRule_Latitude	Searches for names that can describe a column of latitude information.
dataDomain_MetaDataRule_LatitudeLongitude	Searches for names that can describe a column of latitude, longitude, or geocoordinate information.
dataDomain_MetaDataRule_Longitude	Searches for names that can describe a column of longitude information.
dataDomain_MetaDataRule_MaidenName	Searches for names that can describe a column of maiden name information.
dataDomain_MetaDataRule_MiddleName	Searches for names that can describe a column of middle name information.
dataDomain_MetaDataRule_NationalId	Searches for names that can describe a column of national identity numbers.
dataDomain_MetaDataRule_NDC_USA	Searches for names that can describe a column of National Drug Code information
dataDomain_MetaDataRule_NPI_USA	Searches for names that can describe a column of National Provider Identifier numbers.
dataDomain_MetaDataRule_Passport	Searches for names that can describe a column of passport information.
dataDomain_MetaDataRule_PhoneNumber	Searches for names that can describe a column of telephone numbers or fax numbers.
dataDomain_MetaDataRule_Quantity	Searches for names that can describe a column of quantity information.
dataDomain_MetaDataRule_Race	Searches for names that can describe a column of race or color information.
dataDomain_MetaDataRule_Religion	Searches for names that can describe a column of information about religion, faith, or belief.
dataDomain_MetaDataRule_Salary	Searches for names that can describe a column of salary, wages, or compensation information.

Name	Description
dataDomain_MetaDataRule_SSN	Searches for names that can describe a column of Social Security numbers.
dataDomain_MetaDataRule_State	Searches for names that can describe a column of United States state information.
dataDomain_MetaDataRule_Street	Searches for names that can describe a column of street address information.
dataDomain_MetaDataRule_UniqueIdentifying Number	Searches for names that can describe a column of unique identification numbers.
dataDomain_MetaDataRule_UPC_EAN	Searches for names that can describe a column of Universal Product Codes or European Article Numbers.
dataDomain_MetaDataRule_URL	Searches for names that can describe a column of Uniform Resource Locator or web address information.
dataDomain_MetaDataRule_VehicleRegPlate Number	Searches for names that can describe a column of vehicle registration or vehicle license plate numbers.
dataDomain_MetaDataRule_Weight	Searches for names that can describe a column of weight information.
dataDomain_MetaDataRule_ZipCode	Searches for names that can describe a column of ZIP Codes.

Data Rules in the Data Domains Accelerator

Use the data domain data rules to identify columns that contain data that matches the rule criteria.

Find the data rules in the following repository location:

    [Informatica_DQ_Content]\Domain_Discovery\Data_Rules

The following table describes the data rules in the Data Domains accelerator:

Name	Description
dataDomain_DataRule_ABARoutingNumber	Identifies column data that matches the format of an American Banking Association routing number. The routing number identifies a financial institution in a financial transaction.
dataDomain_DataRule_Account_Status	Identifies column data that matches account status values in the reference data.
dataDomain_DataRule_Address_Data	Identifies column data that represents address information. The rule recognizes address data from multiple countries globally.
dataDomain_DataRule_Age	Identifies column data with values from 1 through 120.

Name	Description
dataDomain_DataRule_Alphanumeric_Special Character	Identifies column data that contains unformatted alphanumeric data and special-character data.
dataDomain_DataRule_Amount	Identifies column data that represents a physical quantity.
dataDomain_DataRule_AUT_NATID	Identifies column data that matches the Austrian national ID format.
dataDomain_DataRule_BankAccount_USA	Identifies column data that matches a bank account number format in the United States.
dataDomain_DataRule_BGR_NATID	Identifies column data that matches the Bulgarian national ID format.
dataDomain_DataRule_BIC_SWIFTCode	Identifies column data that matches Bank Identifier Code (BIC) or Society for Worldwide Interbank Financial Telecommunication (SWIFT) code by pattern recognition and country code.
dataDomain_DataRule_BinaryValues	Identifies column data that contains binary values.
dataDomain_DataRule_BirthDay	Identifies column data that matches valid birth dates. The rule verifies the number of years between the input date and current date. The rule returns "Adult," "Minor," or "Valid" based on the values from 1 through 120. The rule returns "Invalid" for all other values.
dataDomain_DataRule_BRA_IDDoc	Identifies column data that matches the number format of the *Registro Geral* ID card in Brazil.
dataDomain_DataRule_BRA_Personal_ID	Identifies column data that matches the Brazilian personal ID format.
dataDomain_DataRule_CAN_SIN	Identifies column data that matches the Social Insurance number format in Canada.
dataDomain_DataRule_CHN_NATID	Identifies column data that matches the Chinese national ID format.
dataDomain_DataRule_City	Identifies column data that contains a valid city name. The rule reads reference data that contains international city names.
dataDomain_DataRule_CompanyName	Identifies column data that matches the organization-name values in the reference data.
dataDomain_DataRule_Computer_Address	Identifies column data that matches the format of IP addresses and MAC addresses.
dataDomain_DataRule_Country	Identifies column data that matches an ISO country name.
dataDomain_DataRule_CountryCode_Phone	Identifies column data that matches phone numbers based on international dialing codes.
dataDomain_DataRule_County	Identifies column data that matches a United States county name.
dataDomain_DataRule_CreditCard_AMEX	Identifies column data that matches the American Express credit card number format.
dataDomain_DataRule_CreditCard_DinersCard	Identifies column data that matches the Diners Club International credit card number format.

Name	Description
dataDomain_DataRule_CreditCard_DiscoverCard	Identifies column data that matches the Discover credit card number format.
dataDomain_DataRule_CreditCard_JCB	Identifies column data that matches the JCB International credit card number format.
dataDomain_DataRule_CreditCard_MasterCard	Identifies column data that matches the MasterCard credit card number format.
dataDomain_DataRule_CreditCard_Visa	Identifies column data that matches the Visa credit card number format.
dataDomain_DataRule_CreditCardNumber	Identifies column data that matches the credit card number format of major credit card organizations, such as American Express, Diners Club International, and Maestro.
dataDomain_DataRule_CreditCardTrack1FormatB	Identifies column data that matches Track 1 Format B credit card information.
dataDomain_DataRule_Currency	Identifies column data that matches a currency term in the reference data.
dataDomain_DataRule_Date_Validation	Identifies the date strings in the source data that appear in a single format in a date column. To configure the date format that the rule uses for validation, open the dq_ValidateDate Expression transformation in the rule and update the In_Date_Format expression variable. The default format is "MM/DD/YYYY." The rule returns "Valid" or "Invalid."
dataDomain_DataRule_Date_Validation_All_Formats	Identifies the date values in the column data and standardizes the column data to a single date format.
dataDomain_DataRule_DEU_Machine_Readable_Passport	Identifies column data that matches the machine-readable German passport number format.
dataDomain_DataRule_DNK_NATID	Identifies column data that matches the Danish national ID format.
dataDomain_DataRule_DriversLicense	Identifies column data that matches Canada, United Kingdom, and Unites States driver license numbers based on the length and pattern of the data values.
dataDomain_DataRule_DriversLicense_Canada	Identifies column data that matches Canada driver license numbers except for numbers from the provinces of British Columbia, Quebec, Manitoba, and Prince Edward Island.
dataDomain_DataRule_DriversLicense_Canada_narrow	Identifies column data that matches Canada driver license numbers except for numbers from the provinces of British Columbia, Quebec, Manitoba, and Prince Edward Island. The rule is similar to the dataDomain_DataRule_DriversLicense_Canada rule. However, dataDomain_DataRule_DriversLicense_Canada_narrow performs a more narrow analysis to reduce the likelihood of false positives.
dataDomain_DataRule_DriversLicense_GBR	Identifies column data that matches United Kingdom driver license numbers.

Name	Description
dataDomain_DataRule_DriversLicense_narrow	Identifies column data that matches driver license numbers from the United Kingdom and from many states and provinces in Canada and the United States. The rule does not validate numbers from the provinces of British Columbia, Quebec, Manitoba, and Prince Edward Island. To reduce the likelihood of false positives, the rule does not validate numbers that contain between four and eight digits.
dataDomain_DataRule_DriversLicense_USA	Identifies column data that matches the driver license numbers of most of the states in the United States.
dataDomain_DataRule_DriversLicense_USA_narrow	Identifies column data that matches the driver license numbers of most of the states in the United States. To reduce the likelihood of false positives, the rule excludes data values that comprise between six and eight digits. For example, the rule excludes a value such as 01012017.
dataDomain_DataRule_Email	Identifies column data that matches a predefined email ID format.
dataDomain_DataRule_ExpirationDate	Identifies column data that matches expired credit card dates. The rule compares the input date to the system date for validation.
dataDomain_DataRule_FIN_NATID	Identifies column data that matches the Finnish national ID format.
dataDomain_DataRule_FirstName	Identifies column data that matches values in a reference data set of first names.
dataDomain_DataRule_FRA_INSEE	Identifies column data that matches the French Institut National de la Statistique et des Études Économiques (INSEE) number format.
dataDomain_DataRule_FullName	Identifies the strings in a column of data that contain first, middle, and last names. The rule compares the words in each string to the reference data.
dataDomain_DataRule_GBR_NINO	Identifies column data that matches the United Kingdom National Insurance number format.
dataDomain_DataRule_GBR_Passport_Number	Identifies column data that matches the United Kingdom passport number format.
dataDomain_DataRule_Gender	Identifies column data that matches the gender values in the reference data.
dataDomain_DataRule_Height	Identifies column data with values 1 through 8, where 8 represents feet in height.
dataDomain_DataRule_HostName	Identifies column data that matches valid host names.
dataDomain_DataRule_HRV_NATID	Identifies column data that matches the Croatian national ID format.
dataDomain_DataRule_IBAN	Identifies column data that matches the International Bank Account Number format for multiple European countries.

Name	Description
dataDomain_DataRule_ICD_10	Identifies column data that matches the names of conditions in the tenth revision of the International Statistical Classification of Diseases and Related Health Problems (ICD). The World Health Organization (WHO) maintains the classification.
dataDomain_DataRule_ICD_9	Identifies column data that matches the names of conditions in the ninth revision of the International Statistical Classification of Diseases and Related Health Problems (ICD). The World Health Organization (WHO) maintains the classification.
dataDomain_DataRule_IND_NATID	Identifies column data that matches the Indian Permanent Account Number format.
dataDomain_DataRule_IND_Passport	Identifies column data that matches the Indian passport number format.
dataDomain_DataRule_IPAddress	Identifies column data that matches a predefined IP address format.
dataDomain_DataRule_ISBN	Identifies column data that matches the International Standard Book Number format.
dataDomain_DataRule_ISIN	Identifies column data that matches the international securities identification number (ISIN) format. An ISIN uniquely identifies a security such as a stock or a bond.
dataDomain_DataRule_ItalyFiscalCode	Identifies column data that matches the Italian national ID format.
dataDomain_DataRule_ITIN_USA	Identifies column data that matches the format of an Individual Taxpayer Identification Number (ITIN) in the United States. The Internal Revenue Service issues the identification numbers.
dataDomain_DataRule_JobPosition	Identifies column data that matches the job position names in the reference data.
dataDomain_DataRule_KOR_NATID	Identifies column data that matches the Korean national ID format.
dataDomain_DataRule_LastName	Identifies column data that matches values in a reference data set of last names.
dataDomain_DataRule_Latitude	Identifies column data that matches valid latitude coordinates.
dataDomain_DataRule_LatitudeLongitude	Identifies column data that matches valid pairs of latitude and longitude coordinates, where each pair is separated by a semicolon.
dataDomain_DataRule_Longitude	Identifies column data that matches valid longitude coordinates.
dataDomain_DataRule_Machine_Readable_Passport	Identifies column data that matches machine-readable passport numbers from all countries.
dataDomain_DataRule_NDC_USA	Identifies column data that matches a National Drug Code (NDC) value in the National Drug Code directory in the United States. Each code uniquely identifies a drug that a manufacturer developed for human use.
dataDomain_DataRule_NOR_NATID	Identifies column data that matches the Norwegian national ID format.

Name	Description
dataDomain_DataRule_NPI_USA	Identifies column data that matches a National Provider Identifier (NPI) number in the United States. The Centers for Medicare and Medicaid Services issue the numbers to healthcare providers.
dataDomain_DataRule_PhoneNumber	Identifies column data that matches the United States phone number format.
dataDomain_DataRule_PostCode	Identifies column data that matches the postal codes of multiple countries.
dataDomain_DataRule_Quantity	Identifies column data that describes a physical quantity and includes units of measurement.
dataDomain_DataRule_Race	Identifies column data that matches the name of a race of people in the reference data.
dataDomain_DataRule_Religion	Identifies column data that matches the name of a religion in the reference data.
dataDomain_DataRule_ROU_NATID	Identifies column data that matches the Romanian national ID format.
dataDomain_DataRule_SouthAfrica_NATID	Identifies column data that matches the South African national ID format.
dataDomain_DataRule_Spanish_NIF	Identifies column data that matches the format of the fiscal identification number (NIF) in Spain.
dataDomain_DataRule_SSN	Identifies column data that matches the United States Social Security number format.
dataDomain_DataRule_State	Identifies column data that matches the state names in the United States.
dataDomain_DataRule_Street	Identifies the strings in the column data that describe street address information, for example street, road, avenue. The rule uses a regular expression to find street descriptors in the column data.
dataDomain_DataRule_SWE_NATID	Identifies column data that matches the Swedish national ID format.
dataDomain_DataRule_TWN_NATID	Identifies column data that matches the Taiwanese national ID format.
dataDomain_DataRule_UPC	Identifies column data that matches a valid Universal Product Code. A Universal Product Code is a type of barcode.
dataDomain_DataRule_UPC_EAN	Identifies column data that matches a valid Universal Product Code or European Article Number. Universal Product Codes and European Article Numbers are types of barcode.
dataDomain_DataRule_URL	Identifies column data that matches predefined URL formats.
dataDomain_DataRule_US_Zip5	Identifies column data that matches United States ZIP Codes.
dataDomain_DataRule_USA_Machine_Readable_Passport	Identifies column data that matches a machine-readable United States passport number format.

Name	Description
dataDomain_DataRule_USA_SSN_post_2011June	Identifies column data that matches the Social Security number format in length, numeric values, and minimum and maximum values of the area, group, and serial number sections. Based on the SSN Randomization initiative, effective June 25, 2011, the rule does not verify the issuance of a Social Security number and the group and area number combination.
dataDomain_DataRule_Weight	Identifies column data that describes a weight value. The rule checks for a number between 0 and 500.
dataDomain_DataRule_ZipCode	Identifies column data that matches United States ZIP Codes.

CHAPTER 4

Australia/New Zealand Accelerator

This chapter includes the following topics:

Australia/New Zealand Accelerator Overview

Use the rules in the Australia/New Zealand Accelerator to verify and enhance data from organizations in Australia and New Zealand.

The Australia/New Zealand accelerator includes rules that perform the following data quality operations:

- Address data cleansing
- Contact data cleansing
- Corporate data cleansing
- General data cleansing
- Matching and deduplication

The Australia/New Zealand accelerator also includes composite rules. A composite rule combines multiple rules into a single object.

The accelerator depends on rules that the Core accelerator installs.

Australia/New Zealand Address Data Cleansing Rules

Use the address data cleansing rules to parse, standardize, and validate address data.

Find the address data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
```

The following table describes the address data cleansing rules in the Australia/New Zealand accelerator:

Name	Description
rule_AUS_Address_Parse_Hybrid	Parses unstructured Australian addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_AUS_Address_Parse_Multiline	Parses unstructured Australian addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_AUS_Address_Validation_Discrete	Validates the deliverability of Australian addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_AUS_Address_Validation_Discrete_w_Geocoding	Validates the deliverability of Australian addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_AUS_Address_Validation_Hybrid	Validates the deliverability of Australian addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_AUS_Address_Validation_Hybrid_w_Geocoding	Validates the deliverability of Australian addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_AUS_Address_Validation_Multiline	Validates the deliverability of Australian addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_AUS_Address_Validation_Multiline_w_Geocoding	Validates the deliverability of Australian addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_NZL_Address_Parse_Hybrid	Parses unstructured New Zealand addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_NZL_Address_Parse_Multiline	Parses unstructured New Zealand addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.

Name	Description
rule_NZL_Address_Validation_Discrete	Validates the deliverability of New Zealand addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_NZL_Address_Validation_Discrete_w_Geocoding	Validates the deliverability of New Zealand addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_NZL_Address_Validation_Hybrid	Validates the deliverability of New Zealand addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_NZL_Address_Validation_Hybrid_w_Geocoding	Validates the deliverability of New Zealand addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_NZL_Address_Validation_Multiline	Validates the deliverability of New Zealand addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_NZL_Address_Validation_Multiline_w_Geocoding	Validates the deliverability of New Zealand addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.

Australia/New Zealand Composite Rules

Use the composite rules in the Australia/New Zealand accelerator to add a set of rules to a mapping as a single object. A composite rule is a rule that makes use of the logic of other accelerator rules.

Find the composite rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Composite_Rules
```

The following table describes the composite rules in the Australia/New Zealand accelerator:

Name	Description
rule_AUS_Contact_Data	Parses, standardizes, and validates Australian contact data, such as addresses, telephone numbers, and Tax File Numbers.
rule_NZL_Contact_Data	Parses, standardizes, and validates New Zealand contact data, such as addresses, telephone numbers, and Inland Revenue Department (IRD) numbers.

Composite Rule for Australian Contact Data

The rule `rule_AUS_Contact_Data` reads mapplets from multiple folders in the repository. The rule also includes a nonreusable transformation.

The following table lists the names and the repository locations of the rules and the transformation in `rule_AUS_Contact_Data`:

Rule	Location
Case_Converter	Nonreusable transformation
rule_Assign_DQ_Mailability_Score_Description	[Informatica_DQ_Content]\Rules\General_Data_Cleansing
rule_AUS_Address_Validation_Hybrid	[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
rule_AUS_Company_Name_Standardization	[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing
rule_AUS_Gender_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_AUS_Multi_Person_Name_Parse	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_AUS_Phone_Number_Standardization	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_AUS_Phone_Number_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_AUS_Tax_File_Number_Standardization	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_AUS_Tax_File_Number_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_Email_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_Prename_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_Salutation_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

Composite Rule for New Zealand Contact Data

The rule `rule_NZL_Contact_Data` reads mapplets from multiple folders in the repository. The rule also includes a nonreusable transformation.

The following table lists the names and the repository locations of the rules and the transformation in `rule_NZL_Contact_Data`:

Rule	Location
Case_Converter	Nonreusable transformation
rule_Assign_DQ_Mailability_Score_Description	[Informatica_DQ_Content]\Rules\General_Data_Cleansing
rule_AUS_Company_Name_Standardization	[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing
rule_AUS_Multi_Person_Name_Parse	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_Email_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_NZL_Address_Validation_Hybrid	[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
rule_NZL_Gender_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_NZL_IRD_Number_Standardization	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

Rule	Location
rule_NZL_IRD_Number_Validate	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_NZL_Phone_Number_Standardization	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_NZL_Phone_Number_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_Prename_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_Salutation_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

Australia/New Zealand Contact Data Cleansing Rules

Use the contact data cleansing rules to parse, standardize, and validate data about business contacts and individuals.

Find the contact data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

The following table describes the contact data cleansing rules in the Australia/New Zealand accelerator:

Name	Description
rule_AUS_Driver_Licence_Number_Validation	Validates Australian driver's license numbers based on length and pattern requirements.
rule_AUS_Gender_Assignment	Assigns gender according to first names. The rule returns "M" for male names, "F" for female names, and "U" if the gender is unknown. For example, the rule assigns the name "John Smith" a gender of "M" for male.
rule_AUS_Given_Name_Standard	Generate given names from Australian nicknames.
rule_AUS_Multi_Person_Name_Parse	Parses person name values into separate ports. The rule creates ports for values such as title, first name, middle name, and surname. The rule output includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping. When the name data identifies more than one person, the rule creates an output port for each full name. For example, the rule can read the name "John and Jane Smith" and create output ports for "John Smith" and "Jane Smith."
rule_AUS_Personal_Name_Parsing_FML	Parses the values in a person name into separate ports. The rule creates the ports in the following sequence: - First name, middle name, last name The rule output also includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping.

Name	Description
rule_AUS_Personal_Name_Parsing_LFM	Parses the values in a person name into separate ports. The rule creates the ports in the following sequence: - Last name, first name, middle name The rule output also includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping.
rule_AUS_Phone_Number_Parse	Parses a Australian telephone number from a string. The rule parses the first telephone number in the data, reading from right to left. The rule recognizes telephone numbers that use leading zeros, international dialing codes, or extensions that begin with the hash symbol. The rule processes the following punctuation symbols: the plus sign, parentheses, and the hash symbol. Before you run the rule, remove all other punctuation, including double spaces. The rule returns a telephone number and also returns a string that contains the input text with the telephone number removed.
rule_AUS_Phone_Number_Standardization	Standardizes Australian telephone numbers to international and local dialing formats. The rule recognizes telephone numbers that use leading zeros, international dialing codes, or extensions that begin with the hash symbol.
rule_AUS_Phone_Number_Validation	Validates the area code and length of Australian telephone numbers. The rule returns the region of the telephone number, as well as codes that indicate if the area code and length of a telephone number are valid.
rule_AUS_Tax_File_Number_Parse	Parses Australian Tax File Numbers (TFN).
rule_AUS_Tax_File_Number_Standardization	Standardizes Australian Tax File Numbers (TFN). To configure the standardized format, edit the TFN_Format expression variable in the dq_Format_TFN Expression transformation. Default is "No_punctuation."
rule_AUS_Tax_File_Number_Validation	Validates Australian Tax File Numbers (TFN) based on the check digit in each number.
rule_NZL_Gender_Assignment	Assigns gender according to New Zealand first names. The rule returns "M" for male names, "F" for female names, and "U" if the gender is unknown. For example, the rule assigns the name "John Smith" a gender of "M" for male.
rule_NZL_Given_Name_Standard	Generate given names from New Zealand nicknames.
rule_NZL_IRD_Number_Parse	Parses nine-digit numeric strings as New Zealand Inland Revenue Department numbers (IRD).
rule_NZL_IRD_Number_Standardization	Standardizes New Zealand Inland Revenue Department numbers (IRD). To configure the standardized format, edit the IRD_Format expression variable in the dq_Format_IRD Expression transformation. Default is "No_punctuation." The rule requires that the input is a nine-digit string.
rule_NZL_IRD_Number_Validate	Validates New Zealand Inland Revenue Department numbers (IRD) based on the check digit in each number.

Name	Description
rule_NZL_Phone_Number_Parse	Parses a New Zealand telephone number from a string. The rule parses the first telephone number in the data, reading from right to left.
	The rule recognizes telephone numbers that use leading zeros, international dialing codes, or extensions that begin with the hash symbol. The rule processes the following punctuation symbols: the plus sign, parentheses, and the hash symbol. Before you run the rule, remove all other punctuation, including double spaces.
	The rule returns a telephone number and also returns a string that contains the input text with the telephone number removed.
rule_NZL_Phone_Number_Standardization	Standardizes New Zealand telephone numbers to international and local dialing formats. The rule recognizes telephone numbers that use leading zeros, international dialing codes, or extensions that begin with the hash symbol.
rule_NZL_Phone_Number_Validation	Validates the area code and length of New Zealand telephone numbers. The rule returns the region of the telephone number, as well as codes that indicate if the area code and length of a telephone number are valid.
rule_Prename_Assignment	Generates an honorific according to the gender. You can change the female_prename expression variable from Ms. to Mrs.
rule_Salutation_Assignment	Generates formal and casual greetings from prenames and name tokens. For example, when input data contains "Mr. John Smith," the rule generates the formal greeting "Dear Mr. Smith," and the casual greeting "Dear John,". You can change the prefix and punctuation by editing the variables in the dq_Generate_Salutation Expression transformation.

Dependencies on Core Contact Data Cleansing Rules

The Australia/New Zealand accelerator depends on the following contact data cleansing rules from the Core accelerator:

- rule_Email_Validation

For more information about these rules, see "Core Contact Data Cleansing Rules" on page 21.

Australia/New Zealand Corporate Data Cleansing Rules

Use the corporate data cleansing rules to parse, standardize, and validate corporate data.

Find the corporate data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing
```

The following table describes the corporate data cleansing rules in the Australia/New Zealand accelerator:

Name	Description
rule_AUS_Business_Number_Parse	Parses 11-digit numeric strings as Australian Business Numbers (ABN).
rule_AUS_Business_Number_Standardize	Standardizes Australian Business Numbers (ABN) to the NN NNN NNN NNN format. The rule requires that the input is a 11-digit string.
rule_AUS_Business_Number_Validation	Validates Australian Business Numbers (ABN) based on the check digit in each number.
rule_AUS_Company_Name_Standardization	Standardizes company names to Australian reference table values.

Australia/New Zealand General Data Cleansing Rules

Use the general data cleansing rules to identify the type of information contained within input fields.

Find the general data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\General_Data_Cleansing

The following table describes the general data cleansing rules in the Australia/New Zealand accelerator:

Name	Description
rule_AUS_NZL_NER_Field_Identification	Identifies the type of information contained in an input field. The rule can identify names, Personal IDs, company names, dates, and address data from Australia and New Zealand. The rule returns a label that describes the type of input data. The rule uses probabilistic matching techniques to identify the types of information.

Dependencies on Core General Data Cleansing Rules

The Australia/New Zealand accelerator depends on the following general data cleansing rules from the Core accelerator:

- rule_Assign_DQ_GeocodingStatus_Description
- rule_Assign_DQ_Mailability_Score_Description
- rule_Assign_DQ_Match_Code_Description
- rule_Remove_Extra_Spaces
- rule_Remove_Hyphen
- rule_Remove_Leading_Zero
- rule_Remove_Period_Parentheses
- rule_Remove_Punctuation
- rule_Remove_Punctuation_and_Space
- rule_Remove_Space
- rule_Replace_Limited_Punct_with_Space

- rule_UpperCase

For more information about these rules, see ["Core General Data Cleansing Rules" on page 22](#).

Australia/New Zealand Matching and Deduplication Rules

Use the matching and deduplication rules in the Australia/New Zealand accelerator to measure the levels of similarity between the records in data sets.

Find the matching and deduplication rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Matching_Deduplication
```

The following table describes the matching and deduplication rules in the Australia/New Zealand accelerator:

Name	Description
mplt_AUS_Firstname_and_TFN_Match	Uses field match strategies to identify duplicate rows in Australian data based on Tax File Numbers (TFN) and first names. The mapplet generates group keys from the TFN data.
mplt_AUS_IMO_Company_Name_and_Address_Match	Uses identity match strategies to identify duplicate rows for Australian data based on company names and addresses. The mapplet generates group keys from the postal code data.
mplt_AUS_IMO_Familyname_and_Address_Match	Uses identity match strategies to identify duplicate rows in Australian data based on family names and addresses. The mapplet generates group keys from the postal code data.
mplt_AUS_IMO_Individual_Name_and_Address_Match	Uses identity match strategies to identify duplicate rows in Australian data based on person names and addresses. The mapplet generates group keys from the postal code data.
mplt_AUS_IMO_Personal_Name_and_Data_Match	Uses identity match strategies to identify duplicate rows in Australian data based on person names and personal data. The fields in the personal data column must contain a single type of data, such as a telephone number, email, or Tax File Number. The mapplet generates group keys from the personal data.
mplt_AUS_Individual_Name_and_Address_Match	Uses field match strategies to identify duplicate rows based on person names and Australia address data. The mapplet uses a combination of characters from the surname values and the postal code values to generate group keys.
mplt_AUS_Individual_Name_and_Date_Match	Uses field match strategies to identify duplicate rows based on Australian person names and dates. The mapplet generates group keys from the date data.
mplt_AUS_Individual_Name_and_Email_Match	Uses field match strategies to identify duplicate rows based on email addresses and Australian person names. The mapplet generates group keys from the email address data.
mplt_AUS_Individual_Name_and_Phone_Match	Uses field match strategies to identify duplicate rows based on Australian person names and telephone numbers. The mapplet generates group keys from the telephone number data.

Name	Description
mplt_AUS_Individual_Name_and_TFN_Match	Uses field match strategies to identify duplicate rows for Australian data based on Tax File Numbers (TFN) and person names. The mapplet generates group keys from the TFN data.
mplt_AUS_Individual_Name_Match	Uses field match strategies to identify duplicate rows based on Australian person names. The mapplet generates NYSIIS codes from the surname values and uses the NYSIIS codes as group keys.
mplt_AUS_NZL_Company_Name_and_Address_Match	Uses field match strategies to identify duplicate rows based on company name and address data from Australia and New Zealand. The mapplet uses a combination of characters from the company name values and the postal code values to generate group keys.
mplt_AUS_NZL_Familyname_and_Address_Match	Uses field match strategies to identify duplicate rows based on surname and address data from Australia and New Zealand. The mapplet uses a combination of characters from the surname values and the postal code values to generate group keys.
mplt_Company_Name_Match	Uses field match strategies to identify duplicate rows based on company name. The mapplet generates Soundex codes from the company name values and uses the Soundex codes as group keys.
mplt_NZL_Firstname_and_IRD_Match	Uses field match strategies to identify duplicate rows for New Zealand data based on Inland Revenue Department (IRD) numbers and first names. The mapplet generates group keys from the IRD number.
mplt_NZL_IMO_Company_Name_and_Address_Match	Uses identity match strategies to identify duplicate rows in New Zealand data based on company names and addresses. The mapplet generates group keys from the postal code data.
mplt_NZL_IMO_Familyname_and_Address_Match	Uses identity match strategies to identify duplicate rows in New Zealand data based on family names and addresses. The mapplet generates group keys from the postal code data.
mplt_NZL_IMO_Individual_Name_and_Address_Match	Uses identity match strategies to identify duplicate rows in New Zealand data based on person names and addresses. The mapplet generates group keys from the postal code data.
mplt_NZL_IMO_Personal_Name_and_Data_Match	Uses identity match strategies to identify duplicate rows in New Zealand data based on person names and personal data. The fields in the personal data column must contain a single type of data, such as telephone number, email, or Inland Revenue Department number. The mapplet generates group keys from the personal data.
mplt_NZL_Individual_Name_and_Address_Match	Uses field match strategies to identify duplicate rows based on person names and New Zealand address data. The mapplet uses a combination of characters from the surname values and the postal code values to generate group keys.
mplt_NZL_Individual_Name_and_Date_Match	Uses field match strategies to identify duplicate rows based on New Zealand person names and dates. The mapplet generates group keys from the date data.
mplt_NZL_Individual_Name_and_Email_Match	Uses field match strategies to identify duplicate rows based on email addresses and New Zealand person names. The mapplet generates group keys from the email address data.

Name	Description
mplt_NZL_Individual_Name_and_IRD_Match	Uses field match strategies to identify duplicate rows based on New Zealand person names and Inland Revenue Department (IRD) numbers. The mapplet generates group keys from the IRD number.
mplt_NZL_Individual_Name_and_Phone_Match	Uses field match strategies to identify duplicate rows based on New Zealand person names and telephone numbers. The mapplet generates group keys from the telephone number data.
mplt_NZL_Individual_Name_Match	Uses field match strategies to identify duplicate rows based on New Zealand person names. The mapplet generates NYSIIS codes from the surname values and uses the NYSIIS codes as group keys.
rule_AUS_NZL_Company_Name_and_Address_MatchScore	Generates a match score based on company names and addresses from Australia and New Zealand.
rule_AUS_NZL_Familyname_and_Address_MatchScore	Generates a match score based on surnames and addresses from Australia and New Zealand.
rule_AUS_NZL_Firstname_and_PID_MatchScore	Generates a match score based on first names and personal identification numbers.
rule_AUS_NZL_Individual_Name_and_Address_MatchScore	Generates a match score based on person names and addresses from Australia and New Zealand.
rule_AUS_NZL_Individual_Name_and_PID_MatchScore	Generates a match score based on person names and personal identification numbers.
rule_Company_Name_MatchScore	Generates a match score based on company names.
rule_Individual_Name_and_Date_MatchScore	Generates a match score based on person names and dates.
rule_Individual_Name_and_Email_MatchScore	Generates a match score based on person names and email addresses.
rule_Individual_Name_and_Phone_MatchScore	Generates a match score based on person names and telephone numbers.
rule_Individual_Name_MatchScore	Generates a match score based on person names.

Australia/New Zealand Demonstration Mappings

The demonstration mappings in the Australia/New Zealand accelerator use multiple rules to demonstrate data quality processes.

Find the demonstration mappings in the following repository location:

        [Informatica_DQ_Content]\Rules_Demo\AUS_NZL_Accelerator

The accelerator contains the following demonstration mappings:

m_AUS_customer_data_demo

Parses, standardizes, and validates Australia and New Zealand data.

m_AUS_customer_matching_demo

Parses and standardizes identity data from Australia and New Zealand and performs identity match analysis on the data.

The mapping analyzes the following data combinations and generates match clusters for each combination:

- Person name and address data
- Person name and telephone number

CHAPTER 5

BCBS 239/CCAR Accelerator

This chapter includes the following topics:

BCBS 239/CCAR Accelerator Overview

Use the BCBS 239/CCAR accelerator to verify data records against key international banking requirements.

The accelerator contains rule specifications. A rule specification is repository object that represents the data requirements of a business rule in a logical form. You can add a rule specification to a mapping in the same way that you add a mapplet to a mapping.

The rule specifications test the data sources that you select against the following regulatory requirements:

- The principles of risk measurement and reportage that the Basel Committee on Banking Supervision define in standard 239
- The Comprehensive Capital Analysis and Review (CCAR) of the United States Federal Reserve

The accelerator installs with demonstration mappings that show you how you can connect rule specifications to data sources and targets. Some of the rule specifications use mapplet rules that install with the accelerator.

BCBS 239/CCAR Rule Specifications

Use the rule specifications to verify the input data against different aspects of BCBS 239 and CCAR.

Find the rule specifications in the following repository location:

```
[Informatica_DQ_Content]\Rule Specifications
```

The following table describes the rule specifications in the BCBS 239/CCAR accelerator:

Name	Description
rlsp_ARM_Initial_Rate_must_be_able_to_fit_a_decimal_7_5_datatype_field_or_an_empty_string	The initial rate of an adjustable rate mortgage must fit a decimal (7,5) datatype field or must be an empty string.
rlsp_ARM_Initial_Rate_Period_must_be_a_small_int	The initial rate period of an adjustable rate mortgage must be a smallint value.
rlsp_ARM_Initial_Rate_Period_should_not_be_populated_for_Fixed_Rate_Loans	The initial rate period of an adjustable rate mortgage should not be populated for fixed-rate loans.
rlsp_ARM_Initial_Rate_Should_Not_Be_Greater_Then_Or_Equal_to_18_or_less_than_1	The initial rate of an adjustable rate mortgage should not be greater than or equal to 18 or less than 1.
rlsp_ARM_Lifetime_Rate_Cap_must_be_able_to_fit_a_decimal_7_5_datatype_field_or_an_empty_string	The lifetime rate cap on an adjustable rate mortgage must fit a decimal (7,5) datatype field or must be an empty string.
rlsp_ARM_Lifetime_Rate_Floor_must_be_able_to_fit_a_decimal_7_5_datatype_field_or_an_empty_string	The lifetime rate floor of an adjustable rate mortgage must fit a decimal (7,5) datatype field or must be an empty string.
rlsp_ARM_loans_should_have_ARM_Initial_Rate_Period	An adjustable rate mortgage loan should have an initial rate period.
rlsp_ARM_Margin_at_Origination_must_be_able_to_fit_a_decimal_7_5_datatype_field_or_an_empty_string	The margin at origination of an adjustable rate mortgage must fit a decimal (7,5) datatype field or must be an empty string.
rlsp_ARM_Margin_at_Origination_should_not_change_unless_it_was_not_provided	The margin at origination of an adjustable rate mortgage should not change unless it is not provided.
rlsp_ARM_Negative_Amortization_Limit_must_be_able_to_fit_a_decimal_7_5_datatype_field_or_an_empty_string	The negative amortization limit of an adjustable rate mortgage must fit a decimal (7,5) datatype field or must be an empty string.
rlsp_ARM_Negative_Amortization_Limit_should_be_an_empty_string_if_Negative_Amortization_not_allowed	The negative amortization limit of an adjustable rate mortgage should be an empty string if negative amortization is not allowed.
rlsp_ARM_Negative_Amortization_Limit_Should_Not_Be_Greater_Than_125_Pct	The negative amortization limit of an adjustable rate mortgage should not be greater than 125 percent.
rlsp_ARM_Negative_Amortization_Limit_should_not_be_populated_for_Fixed_Rate_Loans	The negative amortization limit of an adjustable rate mortgage should not be populated for fixed-rate loans.
rlsp_ARM_Periodic_Pay_Cap_must_be_able_to_fit_a_decimal_7_5_datatype_field_or_an_empty_string	The periodic pay cap on an adjustable rate mortgage must fit a decimal (7,5) datatype field or must be an empty string.
rlsp_ARM_Periodic_Pay_Floor_must_be_able_to_fit_a_decimal_7_5_datatype_field_or_an_empty_string	The periodic pay floor on an adjustable rate mortgage must fit a decimal (7,5) datatype field or must be an empty string.
rlsp_ARM_Periodic_Rate_Cap_must_be_able_to_fit_a_decimal_7_5_datatype_field_or_an_empty_string	The periodic rate cap on an adjustable rate mortgage must fit a decimal (7,5) datatype field or must be an empty string.

Name	Description
rlsp_ARM_Periodic_Rate_Floor_must_be_able_to_fit_a_decimal_7_5_datatype_field_or_an_empty_string	The periodic rate floor on an adjustable rate mortgage must fit a decimal (7,5) datatype field or must be an empty string.
rlsp_Balance_LT_125_pct_LoanAmountDisp	The balance should be less than 125 percent of the original loan amount disbursed.
rlsp_Balloon_Term_must_be_a_number_greater_than_equal_to_12_and_less_or_equal_to_600_when_Balloon_Flag_is_Y	The balloon term must be a number greater than or equal to 12 and less than or equal to 600 when the balloon flag is set.
rlsp_Balloon_Term_must_be_a_smallint	The balloon term must be a smallint value.
rlsp_Balloon_Term_should_not_change_unless_it_was_not_provided	The balloon term should not change unless it is not provided.
rlsp_BalloonFlag_Is_Valid_Value	The balloon flag must indicate yes, no, or unknown.
rlsp_BalloonFlag_Unchanged_unless_Unknown_or_Empty	The balloon flag should not change unless it indicates an unknown status or it is not provided.
rlsp_Buydown_Flag_is_Unknown	The buydown flag indicates an unknown status.
rlsp_Buydown_Flag_Must_Be_Valid_Value	The buydown flag must indicate yes, no, or unknown.
rlsp_Buydown_Flag_should_not_change_unless_it_was_Unknown_or_not_provided	The buydown flag should not change unless it indicates an unknown status or it is not provided.
rlsp_Calc_LTV_Has_GT_20pct_Change_from_Original_LTV	The calculated loan-to-value ratio differs from the original loan-to-value ratio by more than 20 percent.
rlsp_Closing_Date_Check	The closing date must be a valid date with the format YYYY-MM-DD, must not contain spaces, and must not be null.
rlsp_ClosingDate_Plus_Term_and_3Months_Greater_Than_Current_Month	The closing date plus the loan term plus three months should not indicate a date that is earlier than the current month.
rlsp_CLTV_Is_Number_GT_10pct_and_LT_150pct	The combined loan-to-value ratio must be a number greater than 10 percent and less than 150 percent.
rlsp_Credit_Bureau_Score_is_Whole_Number_Between_300_and_899	The origination credit bureau score must be a valid whole number between 300 and 899.
rlsp_Credit_Class_Must_Be_Specific_Value	The credit class value must be within a set of defined values.
rlsp_Credit_Class_must_be_Specific_Value_or_an_empty_string	The credit class value must be within a set of defined values or must be an empty string.
rlsp_Credit_class_should_not_change	The current credit class value should match the credit class value from the previous month.
rlsp_FirstPayDate_Before_NextDueDate	The first payment date should not be later than the next payment date.

Name	Description
rlsp_FirstPayDate_LessThanOrEqual_3Months_After_and_LessThan_40_Years_Passed_CurrentDate	The first payment date must not be more than three months later than the current month or more than 40 years ago.
rlsp_FirstPayDate_LessThanOrEqual_ClosingDate	The closing date must be earlier than or equal to the first payment date.
rlsp_FirstPayDate_Valid_Date	The input string must be a valid date value in the format YYYYMMDD.
rlsp_Fixed_loans_should_not_have_an_ARM_Initial_Rate_Adjustment_Period	A fixed loan should not have an adjustable rate mortgage initial rate adjustment period.
rlsp_Fixed_Loans_should_not_have_ARM_Initial_Rate_Period	A fixed loan should not have an adjustable rate mortgage initial rate period.
rlsp_For_Non_Commercial_Loans_Original_Property_Value_must_be_a_valid_positive_whole_number	For a non-commercial loan, the original property value must be a valid positive whole number that is greater than or equal to 1,000 and less than or equal to 10,000,000.
rlsp_Frontend_DTI_is_Positive_Whole_Number	The original front-end debt to income value must be a valid positive whole number.
rlsp_If_not_a_Balloon_should_not_have_a_Balloon_Term	A loan that is not a balloon loan should not have a balloon term.
rlsp_Interest_Only_At_Origination_cannot_change_unless_it_was_Unknown_or_not_provided	The interest-only at origination value cannot change unless it is unknown or it is not provided.
rlsp_Interest_Only_At_Origination_is_Unknown	The interest-only at origination value is unknown.
rlsp_Lien_Position_At_Origin_Must_Be_1	The lien position at origination must be 1.
rlsp_Lien_Position_at_Origination_cannot_change_unless_it_was_Unknown_or_not_provided	The lien position at origination cannot change unless it is unknown or it is not provided.
rlsp_Lien_Position_at_Origination_Must_Be_Specific_Value	The lien position at origination must be within a set of defined values.
rlsp_Loan_Closing_Date_Less_Than_or_Equal_40_Years_Ago	The year of the loan closing date should be no more than 40 years ago.
rlsp_Loan_Closing_Month_Not_Greater_Than_Current_Month	The month of the loan closing date is earlier than the current month.
rlsp_Loan_Number_Length_Check	The loan number is not empty or null and contains no more than 32 characters.
rlsp_Loan_Type_is_Other	The loan is a commercial loan.
rlsp_Loan_Type_is_Unknown_or_Other	The loan is a commercial loan or the loan type is unknown.
rlsp_LoanAmount_Unchanged_if_NotEmpty	The original loan amount disbursed should not change unless the amount is not provided.
rlsp_LoanPurpose_is_Other_or_Unknown	The loan purpose is not defined or is unknown.

Name	Description
rlsp_LoanPurpose_Unchanged_unless_Unknown_or_not_provided	The loan purpose coding value should not change unless it is unknown or it is not provided.
rlsp_Loans_with_ARM_Initial_Rate_Period_should_be_ARMs	A loan with an adjustable rate mortgage initial rate period should be an adjustable rate mortgage.
rlsp_LoanSource_is_Unknown	The loan source is unknown.
rlsp_LoanSource_Unchanged_unless_Unknown_or_not_provided	The loan source must not change unless it is unknown or it is not provided.
rlsp_LoanType_Unchanged_Unless_Unknown_or_not_Provided	The loan type must not change unless it is unknown or it is not provided.
rlsp_LVT_GT_10pct_and_LT_125pct_for_Non_HARP_and_150pct_for_HARP	The original loan-to-value ratio must be a number greater than ten percent and less than 150 percent for HARP (Home Affordable Refinance Program) loans or a number greater than ten percent and less than 125 percent for non-HARP loans.
rlsp_MI_Coverage_Pct_At_Origination_should_not_change_unless_it_was_not_provided	Mortgage insurance cover percentage at origination should not change unless the percentage is not provided.
rlsp_MI_Coverage_Percent_At_Origination_must_be_able_to_fit_a_decimal_4_2_datatype_field_or_an_empty_string	Mortgage insurance cover percentage at origination must fit a decimal (4, 2) datatype field or must be an empty string.
rlsp_MI_Coverage_Percent_At_Origination_should_be_between_1_and_50_pct	Mortgage insurance cover percentage at origination should be a value between one percent and 50 percent.
rlsp_Negative_Amortization_Flag_must_be_Specifc_Value_or_an_empty_string	The negative amortization flag value must be within a set of defined values or must be an empty string.
rlsp_Negative_Amortization_Flag_must_be_Y_or_N	The negative amortization flag value must indicate yes or no.
rlsp_Negative_Amortization_Flag_should_not_change_unless_it_was_not_provided	The negative amortization flag value should not change unless it is not provided.
rlsp_NonHARP_Loan_and_LTV_GTE_125pct	A loan is not a HARP loan and the calculated original loan-to-value ratio is greater than or equal to 125 percent.
rlsp_Number_of_Units_is_Unknown_or_Other	The loan is a commercial loan and the number of units is not specified.
rlsp_Number_Of_Units_should_not_change_unless_it_was_U_or_not_provided	The number of units should not change unless it is unknown or it is not provided.
rlsp_NumberOfUnits_is_Specific_Value_or_Unknow	The number of units must be within a set of defined values.
rlsp_Occupancy_cannot_change_unless_it_was_Unknown_or_not_provided	The occupancy value should not change unless it is unknown or it is not provided.
rlsp_Occupancy_Code_Must_be_Specific_Value_or_Unknown	The occupancy code must be within a set of defined values.

Name	Description
rlsp_Occupancy_is_Unknown	The occupancy value is unknown.
rlsp_Option_ARM_At_Origination_must_be_Specific_Value	The option adjustable rate mortgage at origination value must be Y or N.
rlsp_Option_ARM_At_Origination_must_be_Specific_Value_or_an_empty_string	The option adjustable rate mortgage at origination value must be Y or N or an empty string.
rlsp_Option_ARM_At_Origination_must_be_the_same_as_prior_month	The option adjustable rate mortgage at origination flag must be the same as the previous month.
rlsp_Origin_Credit_Bureau_Score_should_not_change_unless_it_was_not_provided	The origination credit bureau score should not change unless it is not provided.
rlsp_Original_Backend_DTI_is_Positive_Whole_Number	The original back-end debt-to-income ratio must be a valid positive whole number and must be less than 100.
rlsp_Original_Backend_DTI_is_SmallInt	The original back-end debt-to-income ratio must be a smallint value.
rlsp_Original_Backend_DTI_Unchanged_If_Provided	The original back-end debt-to-income ratio should not change unless it is not provided.
rlsp_Original_CLTV_Is_4_2_format_or_Is_Empty	The original combined loan-to-value ratio must fit a decimal (4, 2) datatype field or must be an empty string.
rlsp_Original_CLTV_Not_Provided	The original combined loan-to-value ratio is not provided.
rlsp_Original_CLTV_Unchanged_If_Not_Empty	The original combined loan-to-value ratio should not change unless it is not provided.
rlsp_Original_Frontend_DTI_GT_Backend_DTI	The original front-end debt-to-income ratio is greater than the back-end debt-to-income ratio.
rlsp_Original_Frontend_DTI_is_Smallint	The original front-end debt-to-income ratio must be a smallint value.
rlsp_Original_Frontend_DTI_Unchanged_If_Provided	The original Frontend DTI should not change unless it is not provided
rlsp_Original_Loan_Amount_Disbursed_must_be_a_valid_positive_whole_number	The original loan amount disbursed must be a valid positive whole number.
rlsp_Original_Loan_Amount_Disbursed_must_be_an_integer	The original loan amount disbursed must be an integer value.
rlsp_Original_Loan_Term_should_be_a_number_between_12_and_600	The original loan term should be a number between 12 and 600.
rlsp_Original_LVT_Is_Number_with_4_2_format	The original loan-to-value ratio must fit a decimal (4, 2) datatype field or must be an empty string.
rlsp_Original_LVT_Unchanged_If_Not_Empty	The original loan-to-value ratio should not change unless it is not provided.

Name	Description
rlsp_Original_Property_Value_must_be_an_Integer	The original property value must be an integer.
rlsp_OriginalPropertyValue_Unchanged_if_Not_Empty	The original property value should not change unless it is not provided.
rlsp_Origination_Credit_Bureau_Score_is_not_between_300_and_899	The origination credit bureau score is a value between 300 and 899.
rlsp_Origination_Credit_Bureau_Score_must_be_a_smallint	The origination credit bureau score must be a smallint value.
rlsp_Origination_Credit_Bureau_Score_should_not_change_unless_it_was_not_provided	The origination credit bureau score should not change unless it is not provided.
rlsp_Property_type_cannot_change_unless_it_was_Unknown_or_not_provided	The property type cannot change unless it is unknown or it is not provided.
rlsp_Property_ZIP_Code_Sould_Be_Less_than_or_equal_to_5_characters_or_an_empty_string	The property ZIP Code must contain five or more characters or must be an empty string.
rlsp_PropertyState_Length_Not_GreaterThan_2	The state code for the property contains no more than two characters.
rlsp_PropertyType_is_Unknown	The property type is unknown.
rlsp_Recourse_Flag_is_Unknown	The recourse flag indicates an unknown status.
rlsp_RecourseFlag_Unchanged_unless_Unknown_or_Empty	The recourse flag should not change unless it is unknown or it is not provided.
rlsp_State_Abbreviation_Valid_for_Non_Commercial_Loan	For non-commercial loans, the property state must be a valid postal abbreviation of one of the 50 States, or DC, or a valid United States possession.
rlsp_Three_Digit_Zip_Code_must_be_Consistent_with_State	A three-digit ZIP Code must be consistent with a United States state.
rlsp_ZipCode_Length_of_5_for_NonCommercial_Loan	For non-commercial loans, the property ZIP Code should contain five digits.
rlsp_Zipcode_Unchanged_and_NotEmpty	The property ZIP Code should not change unless it is not provided

Dependent Rules

A rule specification can read a mapplet rule. The BCBS 239 CCAR accelerator depends on a set of mapplet rules that install to the accelerator folder in the Model repository.

The rule specifications in the BCBS 239 CCAR accelerator read the following rules:

- rule_Convert_to_Integer. Converts a string value to an integer.
- rule_Convert_to_Number. Converts a string value to a number.
- rule_Decimal_Length. Returns the decimal length of a string.
- rule_Is_Date. Verifies that a string is a valid date.

- rule_Is_Numeric. Verifies that a string is a valid number.

BCBS 239/CCAR Demonstration Mappings

The demonstration mappings in the BCBS 239/CCAR accelerator use multiple rule specifications to demonstrate data quality processes.

Find the demonstration mappings in the following repository location:

```
[Informatica_DQ_Content]\Rules_Demo\Informatica_CCAR_BCBS239_Accelerator
```

The BCBS 239/CCAR accelerator includes the following mappings:

m_CCAR_balloon_term_demo

Validates the input data against the conditions that the rule specifications define for balloon loans.

m_CCAR_fixed_demo

Validates the input data against the conditions that the rule specifications define for fixed-term loans.

CHAPTER 6

Brazil Accelerator

This chapter includes the following topics:

Brazil Accelerator Overview

Use the rules in the Brazil accelerator to verify and enhance data from organizations in Brazil.

The Brazil accelerator includes rules that perform the following data quality operations:

- Address data cleansing
- Contact data cleansing
- Corporate data cleansing
- General data cleansing
- Matching and deduplication

The Brazil accelerator also includes a composite rule. A composite rule combines multiple rules into a single object.

The accelerator depends on rules that the Core accelerator installs.

Brazil Address Data Cleansing Rules

Use the address data cleansing rules to parse, standardize, and validate address data.

Find the address data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
```

The following table describes the address data cleansing rules in the Brazil accelerator:

Name	Description
rule_BRA_Address_Parse_Hybrid	Parses unstructured Brazilian addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_BRA_Address_Parse_Multiline	Parses unstructured Brazilian addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_BRA_Address_Validation_Discrete	Validates the deliverability of Brazilian addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_BRA_Address_Validation_Discrete_w_Geocoding	Validates the deliverability of Brazilian addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_BRA_Address_Validation_Hybrid	Validates the deliverability of Brazilian addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_BRA_Address_Validation_Hybrid_w_Geocoding	Validates the deliverability of Brazilian addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_BRA_Address_Validation_Multiline	Validates the deliverability of Brazilian addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_BRA_Address_Validation_Multiline_w_Geocoding	Validates the deliverability of Brazilian addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.

Brazil Composite Rules

Use the composite rules in the Brazil accelerator to add a set of rules to a mapping as a single object. A composite rule is a rule that makes use of the logic of other accelerator rules.

Find the composite rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Composite_Rules
```

The following table describes the composite rule in the Brazil accelerator:

Name	Description
rule_BRA_Contact_Data	Parses, standardizes, and validates Brazilian contact data, such as addresses, telephone numbers, and Cadastro de Pessoas Físicas (CPF) numbers.

The rule `rule_BRA_Contact_Data` reads mapplets from multiple folders in the repository. The rule also includes a nonreusable transformation.

The following table lists the names and the repository locations of the rules and the transformation in `rule_BRA_Contact_Data`:

Rule	Location
Case_Converter	Nonreusuable transformation
rule_Assign_DQ_Mailability_Score_Description	[Informatica_DQ_Content]\Rules\General_Data_Cleansing
rule_BRA_Address_Validation_Hybrid	[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
rule_BRA_Company_Suffix_Standardization	[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing
rule_BRA_Personal_CPF_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_BRA_Personal_Name_Parse_Validate	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_BRA_Phone_Number_Standardization	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_BRA_Phone_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_BRA_Prename_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_BRA_Salutation_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_Email_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

Brazil Contact Data Cleansing Rules

Use the contact data cleansing rules to parse, standardize, and validate data about business contacts and individuals.

Find the contact data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
```

The following table describes the contact data cleansing rules in the Brazil accelerator:

Name	Description
rule_BRA_Gender_Assignment	Assigns gender according to first name. The rule returns "M" for male names, "F" for female names, and "U" if the gender is unknown. For example, the rule assigns the name "Joao Coelho" a gender of "M" for male.
rule_BRA_Given_Name_Standard	Generate given names from Brazilian nicknames.
rule_BRA_Personal_CPF_Validation	Validates check digits for Cadastro de Pessoas Físicas (CPF) numbers.
rule_BRA_Personal_Name_Parse _Validate	Parses person name values into separate ports. The rule creates ports for values such as title, first name, middle name, and surname. The rule also indicates if the name might be a company name and validates the spelling of the name. The rule output includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping.
rule_BRA_Personal_PIS_PASEP_ Validation	Validates Brazilian social insurance numbers.
rule_BRA_Personal_Voter_Registration_Validation	Validate the check digits in Brazil voter registration numbers.
rule_BRA_Phone_Number_Parse	Parses a Brazilian telephone number from a string. The rule parses the first telephone number in the data, reading from left to right. The rule returns a telephone number and also returns a string that contains the input text with the telephone number removed.
rule_BRA_Phone_Number_Standardization	Standardizes Brazilian telephone numbers. The rule returns the telephone number in the following formats: - Standard - nn nnnn nnnn - Dashes - nn-nnnn-nnnn - No Spaces - nnnnnnnnnn
rule_BRA_Phone_Validatation	Validates the area code and length of Brazilian telephone numbers. The rule returns codes that indicate if the area code and length of a telephone number are valid.
rule_BRA_Prename_Assignment	Generates an honorific according to the gender. You can change the female_prename expression variable from "Sra" to "Sta".
rule_BRA_Salutation_Assignment	Generates formal and casual greetings from prenames and name tokens. For example, when input data contains "Sr. Joao Coelho," the rule generates the formal greeting "Prezado Sr. Coelho," and the casual greeting "Prezado Joao,". You can change the prefix and punctuation by editing the variables in the dq_Generate_Salutation Expression transformation.

Dependencies on Core Contact Data Cleansing Rules

The Brazil accelerator depends on the following contact data cleansing rules from the Core accelerator:

- rule_Email_Parse_Into_Mailbox_Domain
- rule_Email_Validation

For more information about these rules, see "Core Contact Data Cleansing Rules" on page 21.

Brazil Corporate Data Cleansing Rules

Use the corporate data cleansing rules in the Brazil accelerator to standardize and validate corporate data.

Find the corporate data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing

The following table describes the corporate data cleansing rules in the Brazil accelerator:

Name	Description
rule_BRA_Company_CNPJ_Validation	Validates Cadastro Nacional da Pessoa Jurídica (CPNJ) numbers. CPNJ numbers identify Brazilian companies.
rule_BRA_Company_Suffix_Standardization	Standardizes Brazilian company suffixes.

Brazil General Data Cleansing Rules

Use the general data cleansing rules to identify the type of information contained in input fields.

Find the general data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\General_Data_Cleansing

The following table describes the general data cleansing rules in the Brazil accelerator:

Name	Description
rule_BRA_NER_Field_Identification	Identifies the type of information contained in an input field. The rule can identify names, Personal IDs, company names, dates, and Brazilian address data. The rule returns a label that describes the type of input data. The rule uses reference data and probabilistic matching techniques to identify the types of information.

Dependencies on Core General Data Cleansing Rules

The Brazil accelerator depends on the following general data cleansing rules from the Core accelerator:

- rule_Assign_DQ_GeocodingStatus_Description
- rule_Assign_DQ_Mailability_Score_Description
- rule_Assign_DQ_Match_Code_Descriptions
- rule_Remove_Extra_Spaces
- rule_Remove_Non_Numbers
- rule_Remove_Punctuation
- rule_Remove_Punctuation_and_Space
- rule_Replace_Limited_Punct_with_Space
- rule_TitleCase
- rule_UpperCase

For more information about these rules, see "Core General Data Cleansing Rules" on page 22.

Brazil Matching and Deduplication Rules

Use the matching and deduplication rules to measure the levels of similarity between the records in data sets.

Find the matching and deduplication rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Matching_Deduplication

The following table describes the matching and deduplication rules in the Brazil accelerator:

Name	Description
mplt_BRA_Company_Name_and_ Address_Match	Uses field match strategies to identify duplicate rows based on company name and Brazilian address data. The mapplet uses a combination of characters from the company name values and the postal code values to generate group keys.
mplt_BRA_Familyname_and_Addr ess_Match	Identifies duplicate rows in Brazilian data based on family names and addresses. The mapplet uses a combination of characters from the surname values and the postal code values to generate group keys.
mplt_BRA_Firstname_and_CPF_ Match	Uses field match strategies to identify duplicate rows based on first name and Cadastro de Pessoas Físicas (CPF) number. The mapplet generates group keys from the CPF number.
mplt_BRA_IMO_Company_Name_ and_Address_Match	Uses identity match strategies to identify duplicate rows in Brazilian data based on company names and addresses. The mapplet generates group keys from the postal code data.
mplt_BRA_IMO_Familyname_and _Address_Match	Uses identity match strategies to identify duplicate rows in Brazilian data based on family names and addresses. The mapplet generates group keys from the postal code data.
mplt_BRA_IMO_Individual_Name_ and_Address_Match	Uses identity match strategies to identify duplicate rows in Brazilian data based on person names and addresses. The mapplet generates group keys from the postal code data.
mplt_BRA_IMO_Personal_Name_ and_Data_Match	Uses identity match strategies to identify duplicate rows in Brazilian data based on person names and personal data. The fields in the personal data column must contain a single type of data, such as telephone number, email, or Cadastro de Pessoas Físicas number. The mapplet generates group keys from the personal data.
mplt_BRA_Individual_Name_and_ Address_Match	Uses field match strategies to identify duplicate rows based on person names and Brazilian address data. The mapplet uses a combination of characters from the surname values and the postal code values to generate group keys.
mplt_BRA_Individual_Name_and_ CPF_Match	Uses field match strategies to identify duplicate rows based on Brazilian person names and Cadastro de Pessoas Físicas (CPF) numbers. The mapplet generates group keys from the CPF number.
mplt_BRA_Individual_Name_and_ Date_Match	Uses field match strategies to identify duplicate rows based on Brazilian person names and date data. The mapplet generates group keys from the date data.
mplt_BRA_Individual_Name_and_ Email_Match	Uses field match strategies to identify duplicate rows based on Brazilian person names and email addresses. The mapplet generates group keys from email address data.

Name	Description
mplt_BRA_Individual_Name_and_Phone_Match	Uses field match strategies to identify duplicate rows based on Brazilian person names and telephone numbers. The mapplet generates group keys from telephone numbers.
mplt_Company_Name_Match	Uses field match strategies to identify duplicate rows based on company name. The mapplet generates Soundex codes from the company name values and uses the Soundex codes as group keys.
rule_BRA_Company_Name_and_Address_MatchScore	Generates a match score based on company names and Brazilian address data.
rule_BRA_Familyname_and_Address_MatchScore	Generates a match score based on surnames and Brazilian address data.
rule_BRA_Firstname_and_CPF_MatchScore	Generates a match score based on first name and Cadastro de Pessoas Físicas (CPF) number.
rule_BRA_Individual_Name_and_Address_MatchScore	Generates a match score based on person names and Brazilian address data.
rule_BRA_Individual_Name_and_CPF_MatchScore	Generates a match score based on person names and Brazilian address data.
rule_BRA_Individual_Name_and_Phone_MatchScore	Generates a match score based on person names and telephone numbers.
rule_Company_Name_MatchScore	Generates a match score based on company names.
rule_Individual_Name_and_Date_MatchScore	Generates a match score based on person names and dates.
rule_Individual_Name_and_Email_MatchScore	Generates a match score based on person names and email addresses.

Brazil Demonstration Mappings

The demonstration mappings in the Brazil accelerator use multiple rules to demonstrate data quality processes.

Find the demonstration mappings in the following repository location:

    [Informatica_DQ_Content]\Rules_Demo\BRA_Accelerator

The accelerator contains the following demonstration mappings:

m_BRA_customer_data_demo

Parses, standardizes, and validates Brazilian data.

m_BRA_customer_matching_demo

Parses and standardizes identity data from Brazil and performs identity match analysis on the data.

The mapping analyzes the following data combinations and generates match clusters for each combination:

- Person name and address data

- Person name and telephone number

CHAPTER 7

Financial Services Accelerator

This chapter includes the following topics:

Financial Services Accelerator Overview

Use the Financial Services accelerator to verify and enhance data from organizations in the financial services sector.

The Financial Services accelerator includes rules that perform the following data quality processes:

- Contact data cleansing
- Financial data cleansing
- General data cleansing
- Matching and deduplication

The accelerator depends on data cleansing rules that the Core accelerator installs.

Financial Services Contact Data Cleansing Rules

Use the contact data cleansing rules to standardize contact data about business contacts and individuals.

Find the contact data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
```

The following table describes the contact data cleansing rule in the Financial Services accelerator:

Name	Description
rule_USA_Given_Name_Standard	Generates given names from United States nicknames. For example, the rule standardizes the nickname "Bob" to the given name "Robert."

Financial Services Financial Data Cleansing Rules

Use the financial data cleansing rules to parse, standardize, and validate financial data.

Find the financial data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Financial_Data_Cleansing

The following table describes the financial data cleansing rules in the Financial Services accelerator:

Name	Description
rule_Account_Status_Validation	Validates the account status. The rule requires account status reference data.
rule_Accrual_Period_Validation	Validates that the start date is earlier than the end date.
rule_Age_For_Account_Validation	Validates the customer age for the account type. The rule uses the age_per_account_infa reference table. You must update the reference table with your own data.
rule_Beta_Coefficient_Validation	Validates that the Beta coefficient string is a number. The rule indicates that the string is a positive number, negative number, zero, or not a number.
rule_BIC_SWIFT_Code_Validation	Validates a Bank Identifier Code (BIC) or Society for Worldwide Interbank Financial Telecommunication (SWIFT) code by pattern recognition and country code validation.
rule_CAN_Transit_Number_Validation	Uses paper and electronic fund transactions to validate the format of a Canadian transit number.
rule_Credit_Card_Expiry_Check	Validates a credit card expiration date. The rule compares the credit card expiration date to the system date and identifies expired dates. The rule accepts a seven character string in the format MM/YYYY.
rule_Credit_Card_Security_Code_Validation	Validates that the credit card security code is a whole number that contains three or four digits.
rule_Currency_Code_Country_Validation	Validates that the currency code is valid for the ISO three-character country code.
rule_Currency_Code_Validation	Validates the currency code. The rule returns "Valid" or "Invalid."
rule_CUSIP_Validation	Validates the format and length of the check digit value. The rule returns a status that describes the validity of the check digit value and a message that explains the status.
rule_Delta_Validation	Validates that the delta value is positive, negative, or zero.

Name	Description
rule_Dividend_Yield_Validation	Validates that the dividend yield string is a number greater than or equal to zero. The rule returns whether the string is a positive number, negative number, zero, or not a number.
rule_EAD_Drawn_Balance_Validation	Validates that the amount listed in the exposure at default (EAD) is not less than the drawn balance. The rule follows the guidelines for EAD calculation by the Financial Services Authority in the United Kingdom.
rule_EAD_Validation	Validates that the exposure at default (EAD) string is a number. The rule returns whether the string is a positive number, negative number, zero, or not a number.
rule_EPS_Validation	Validates that the input is a number greater than or equal to zero.
rule_Ex_Dividend_Date_Validation	Validates that the ex-dividend date and the record date are valid dates and that the ex-dividend date is earlier than the record date. The rule identifies dates with a difference of more than 15 days as not valid. The rule returns the difference in days between the record date and the ex-dividend date.
rule_Gamma_Validation	Validates that the Gamma string is a number. The rule returns whether the string is a positive number, negative number, zero, or not a number.
rule_GBR_Bank_Account_Parse	Parses eight-digit numeric strings as United Kingdom bank account numbers.
rule_GBR_Bank_Account_Validation	Validates United Kingdom bank account numbers. The rule returns codes that indicate whether the input is numeric and whether it is the correct number of digits.
rule_GBR_Bank_Sort_Code_Parse	Parses six-digit numeric strings as United Kingdom bank sort codes. The rule parses strings of numbers in the following formats: - Consecutive numbers (999999) - Numbers delimited with a dash (99-99-99)
rule_GBR_Bank_Sort_Code_Standardize	Standardizes a United Kingdom bank sort code to the format "NN-NN-NN."
rule_GBR_Bank_Sort_Code_Validation	Validates the format and length of United Kingdom bank sort codes that are standardized to the dash-delimited format (99-99-99). The rule returns a Status port that describes the validity of the sort code and a Validation Note port that explains the status. If the sort code prefix matches a known assignment for a United Kingdom bank, the Validation Note port includes the bank name.
rule_Interest_Rate_Within_Range	Validates if the decimal interest rate value is within the specified range. The range is set by the two variable ports in the Expression transformation. The rule returns "True" or "False."
rule_ISIN_Code_Validation	Validates that the input value is an International Securities Identification Number (ISIN). The rule checks the structure of the value and verifies the check digit.
rule_Loan_to_Value_Ratio	Calculates the loan to value ratio, which is the loan amount divided by the property value.
rule_Loss_Given_Default_Validation	Validates that the string is numeric and a positive, negative, or zero value.
rule_Market_Cap_Validation	Validates that the input is a number greater than or equal to zero.

Name	Description
rule_Maturity_Date_Validation	Validates that the maturity date is greater than the system date.
rule_Positive_Close_Price_Value _Validation	Validates that the input is a number greater than zero.
rule_Positive_Coupon_Percent_V alidation	Validates that the input is a number greater than zero.
rule_Positive_Last_Price_Value_ Validation	Validates that the input is a number greater than zero.
rule_Positive_Open_Price_Valida tion	Validates that the input is a number greater than zero.
rule_Positive_Volume_Validation	Validates that the input is a number greater than zero.
rule_Price_Earnings_Ratio_Valid ation	Validates that the price-to-earnings ratio is a positive number in the range of 0 - 100.
rule_Probability_of_Default_Vali dation	Validates that the probability of default value is numeric and indicates if it is a positive, negative, or zero value. If positive, The rule returns status messages for values in the following ranges: - < = .1 - > .1 and < = .5 - > .5 and < = 1 - > 1
rule_Rating_Code_Validation	Validates that a rating is in the Standard & Poor's ratings scale, the Moody's ratings scale, or in a user-defined list.
rule_Rating_Date_Validation	Validates that the rating date is one year greater than the system date.
rule_Risk_Weighted_Asset_Valid ation	Validates that a risk weighted asset is a positive number.
rule_SEDOL_Validation	Validates a Stock Exchange Daily Official List (SEDOL) code by checking its format and check digit.
rule_Stock_Exchange_Validation	Validates most stock exchanges world wide by name and symbol.
rule_USA_Routing_Number_Valid ation	Validates a standard magnetic ink character recognition line (MICR) formatted routing number. Validates the Associated Federal Reserve Bank, the structure of the input, and the checksum calculation.
rule_Volatility_Validation	Validates that the volatility value is a number greater than or equal to zero.

Financial Services General Data Cleansing Rules

Use the general data cleansing rules to identify the type of information contained in input fields.

Find the general data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\General _Data_Cleansing
```

The following table describes the general data cleansing rule in the Financial Services accelerator:

Name	Description
rule_Postive_Number_Validation	Validates that the number is greater than zero.

Dependencies on Core General Data Cleansing Rules

The Financial Services accelerator depends on the following general data cleansing rules from the Core accelerator:

- rule_Remove_Punctuation
- rule_Remove_Punctuation_and_Space
- rule_Remove_Space
- rule_UpperCase

For more information about these rules, see "Core General Data Cleansing Rules" on page 22.

Financial Services Matching and Deduplication Rules

Use the matching and deduplication rules to generate match scores and identify duplicate records.

Find the matching and deduplication rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Matching_Deduplication

The following table describes the matching and deduplication rules in the Financial Services accelerator:

Name	Description
mplt_Company_Name_and_Address_Match	Identifies duplicate rows based on company name and United States address data. The mapplet uses a combination of characters from the company name values and the postal code values to generate group keys.
mplt_Company_Name_Match	Identifies duplicate rows based on company name. The mapplet generates Soundex codes from the company name values and uses the Soundex codes as group keys.
mplt_Familyname_and_Address_Match	Identifies duplicate rows based on surname and United States address data. The mapplet uses a combination of characters from the surname values and the postal code values to generate group keys.
mplt_Individual_Name_and_Address_Match	Identifies duplicate rows based on person names and United States address data. The mapplet generates NYSIIS codes from the surname values and uses the NYSIIS codes as group keys.
mplt_Individual_Name_and_Date_Match	Identifies duplicate rows based on person names and date data. The mapplet generates group keys generated from the date data.
mplt_Individual_Name_and_Email_Match	Identifies duplicate rows based on person names and email addresses. The mapplet matches generates keys generated from the email address data.
mplt_Individual_Name_and_Phone_Match	Identifies duplicate rows based on person names and telephone numbers. The mapplet generates group keys from telephone numbers.

Name	Description
mplt_Individual_Name_Match	Identifies duplicate rows based on person names. The mapplet generates NYSIIS codes from the surname values and uses the NYSIIS codes as group keys.
rule_Company_Name_and_Address_MatchScore	Generates a match score based on company names and United States addresses.
rule_Company_Name_MatchScore	Generates a match score based on company names.
rule_Familyname_and_Address_MatchScore	Generates a match score based on surnames and United States addresses.
rule_Individual_Name_and_Address_MatchScore	Generates a match score based on person names and United States addresses.
rule_Individual_Name_and_Date_MatchScore	Generates a match score based on person names and dates.
rule_Individual_Name_and_Email_MatchScore	Generates a match score based on person names and email addresses.
rule_Individual_Name_and_Phone_MatchScore	Generates a match score based on person names and telephone numbers.
rule_Individual_Name_MatchScore	Generates a match score based on person names.

CHAPTER 8

France Accelerator

This chapter includes the following topics:

France Accelerator Overview

Use the rules in the France accelerator to verify and enhance data from organizations in France.

The France accelerator includes rules that perform the following data quality processes:

- Address data cleansing
- Contact data cleansing
- Corporate data cleansing
- General data cleansing
- Matching and deduplication

The accelerator depends on data cleansing rules that the Core accelerator installs.

France Address Data Cleansing Rules

Use the address data cleansing rules to parse, standardize, and validate address data.

Find the address data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
```

The following table describes the address data cleansing rules in the France accelerator:

Name	Description
rule_FRA_Address_Parse_Hybrid	Parses unstructured French addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_FRA_Address_Parse_Multiline	Parses unstructured French addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_FRA_Address_Validation_Discrete_w_Geocoding	Validates the deliverability of French addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_FRA_Address_Validation_Discrete	Validates the deliverability of French addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_FRA_Address_Validation_Hybrid_w_Geocoding	Validates the deliverability of French addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_FRA_Address_Validation_Hybrid	Validates the deliverability of French addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_FRA_Address_Validation_Multiline _w_Geocoding	Validates the deliverability of French addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_FRA_Address_Validation_Multiline	Validates the deliverability of French addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.

France Composite Rules

Use the composite rule in the France accelerator to add a set of rules to a mapping as a single object. A composite rule is a rule that makes use of the logic of other accelerator rules.

The composite rules in the France accelerator install to the following repository location:

```
[Informatica_DQ_Content]\Rules\Composite_Rules
```

Composite Rule for French Contact Data

The following table describes the composite rule for contact data in the France accelerator:

Name	Description
rule_FRA_Contact_Data	Parses, standardizes, and validates French contact data, such as addresses and telephone numbers.

The rule `rule_FRA_Contact_Data` reads mapplets from multiple folders in the repository. The rule also includes a nonreusable transformation.

The following table lists the names and the repository locations of the rules and the transformation in `rule_FRA_Contact_Data`:

Rule	Location
Case_Converter	Nonreusable transformation
rule_Assign_DQ_Mailability_Score_Description	[Informatica_DQ_Content]\Rules\General_Data_Cleansing
rule_Email_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_FRA_Address_Validation_Hybrid	[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
rule_FRA_Company_Name_Standardization	[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing
rule_FRA_Gender_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_FRA_Multi_Person_Name_Parse	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_FRA_Phone_Number_Standardize	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_FRA_Phone_Number_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_FRA_Prename_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_FRA_Salutation_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

France Contact Data Cleansing Rules

Use the contact data cleansing rules to parse, standardize, and validate data about business contacts and individuals.

Find the contact data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
```

The following table describes the contact data cleansing rules in the France accelerator:

Name	Description
rule_FRA_Gender_Assignment	Assigns gender according to first names. The rule returns "M" for male names, "F" for female names, and "U" if the gender is unknown. For example, the rule assigns the name "Jean Leclerc" a gender of "M" for male.
rule_FRA_Given_Name_Standard	Generates given names from French nicknames.
rule_FRA_INSEE_Parse	Parses the French Institut National de la Statistique et des Études Économiques (INSEE) number from a string.
rule_FRA_INSEE_Standardization	Standardizes the French INSEE number to a string of 13 digits or to 13 digits followed by a space and the two-digit control key.
rule_FRA_INSEE_Validation	Validates the INSEE number based on the gender, date, and Code Officiel Géographique (COG) values.
rule_FRA_Multi_Person_Name_Parse	Parses person name values into separate ports. The rule creates ports for values such as title, first name, middle name, and surname. The rule output includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping. When the name data identifies more than one person, the rule creates an output port for each full name. For example, the rule can read the name "Jean et Marianne Leclerc" and create output ports for "Jean Leclerc" and "Marianne Leclerc."
rule_FRA_Personal_Name_Parsing_FML	Parses the values in a person name into separate ports. The rule creates the ports in the following sequence: - First name, middle name, last name The rule output also includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping.
rule_FRA_Personal_Name_Parsing_LFM	Parses the values in a person name into separate ports. The rule creates the ports in the following sequence: - Last name, first name, middle name The rule output also includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping.
rule_FRA_Phone_Number_Parse	Parses a French telephone number from a string. The rule parses the first telephone number in the data, reading from right to left. The rule recognizes telephone numbers that use leading zeros, international dialing codes, or extensions that begin with the hash symbol. The rule processes the following punctuation symbols: the plus sign, parentheses, and the hash symbol. Before you run the rule, remove all other punctuation, including double spaces. The rule returns a telephone number and also returns a string that contains the input text with the telephone number removed.
rule_FRA_Phone_Number_Standardize	Standardizes French telephone numbers to international and local dialing formats. The rule recognizes telephone numbers that use leading zeros, international dialing codes, or extensions that begin with the hash symbol.

Name	Description
rule_FRA_Phone_Number_Validation	Validates the area code and length of French telephone numbers. The rule returns the region of the telephone number, as well as codes that indicate if the area code and length of a telephone number are valid.
rule_FRA_Prename_Assignment	Generates an honorific according to the gender.
rule_FRA_Salutation_Assignment	Generates formal and casual greetings from prenames and name tokens. For example, when input data contains "M. Jean Leclerc," the rule generates the formal greeting "Monsieur Leclerc," and the casual greeting "Cher Jean,". You can change the prefix and punctuation by editing the variables in the dq_Generate_Salutation Expression transformation.

Dependencies on Core Contact Data Cleansing Rules

The France accelerator depends on the following contact data cleansing rules from the Core accelerator:

- rule_Email_Validation

For more information about these rules, see "Core Contact Data Cleansing Rules" on page 21.

France Corporate Data Cleansing Rules

Use the corporate data cleansing rules to standardize corporate data.

Find the corporate data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing

The following table describes the corporate data cleansing rules from the France accelerator:

Name	Description
rule_FRA_Company_Name_Standardization	Standardizes the company names to reference table values.
rule_FRA_SIRET_Number_Parse	Parses the French système d'identification du répertoire des établissements (SIRET) number from a string.
rule_FRA_SIRET_Number_Standardize	Standardizes a 14-digit number to the NNN NNN NNN NNNNN format regardless of the spacing or punctuation in the string. There is no standardization for a string with less than 14 digits.
rule_FRA_SIRET_Number_Validation	Validates a système d'identification du répertoire des établissements (SIRET) number. Rule assumes the number is in the standard format after applying the rule_FRA_SIRET_Number_Standardization rule.

France General Data Cleansing Rules

Use the general data cleansing rules to identify the type of information contained in input fields.

Find the general data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\General_Data_Cleansing

The following table describes the general data cleansing rules in the France accelerator:

Name	Description
rule_FRA_NER_Field_Identification	Identifies the type of information contained in an input field. The rule can identify names, Personal IDs, company names, dates, and French address data. The rule returns a label that describes the type of input data. The rule uses reference data to identify the types of information. The rule uses probabilistic matching techniques to identify the types of information.

The France accelerator depends on the following general data cleansing rules from the Core accelerator:

- rule_Assign_DQ_GeocodingStatus_Description
- rule_Assign_DQ_Mailability_Score_Description
- rule_Assign_DQ_Match_Code_Description
- rule_Luhn_Algorithm
- rule_Remove_Extra_Spaces
- rule_Remove_Parentheses
- rule_Remove_Punctuation
- rule_Remove_Punctuation_and_Space
- rule_Replace_Limited_Punct_with_Space
- rule_UpperCase

For more information about these rules, see .

France Matching and Deduplication Rules

Use the matching and deduplication to generate match scores and identify duplicate records.

The matching and deduplication rules in the France accelerator install to the following repository location:

    [Informatica_DQ_Content]\Rules\Matching_Deduplication

The following table describes the matching and deduplication rules in the France accelerator:

Name	Description
mplt_Company_Name_Match	Uses field match strategies to identify duplicate rows based on company names. The mapplet generates Soundex codes from the company name values and uses the Soundex codes as group keys.
mplt_FRA_Company_Name_and_Address_Match	Uses field match strategies to identify duplicate rows based on company names and addresses. The mapplet uses a combination of characters from the company name values and the postal code values to generate group keys.
mplt_FRA_Familyname_and_Address_Match	Uses field match strategies to identify duplicate rows based on family names and addresses. The mapplet uses a combination of characters from the surname values and the postal code values to generate group keys.
mplt_FRA_Firstname_and_INSEE_Match	Uses field match strategies to identify duplicate rows based on the French Institut National de la Statistique et des Études Économiques (INSEE) number. The mapplet generates group keys from the INSEE number data.
mplt_FRA_Firstname_Surname_DOB_and_Postcode_Match	Uses field match strategies to identify duplicate rows of personal names, date of birth, and postal codes. The mapplet generates group keys from the postal code data.
mplt_FRA_IMO_Company_Name_and_Address_Match	Uses identity match strategies to identify duplicate rows in French data based on company names and addresses. The mapplet generates group keys from the postal code data.
mplt_FRA_IMO_Familyname_and_Address_Match	Uses identity match strategies to identify duplicate rows in French data based on family names and addresses. The mapplet generates group keys from the postal code data.
mplt_FRA_IMO_Individual_Name_and_Address_Match	Uses identity match strategies to identify duplicate rows in French data based on person names and addresses. The mapplet generates group keys from the postal code data.
mplt_FRA_IMO_Personal_Name_and_Data_Match	Uses identity match strategies to identify duplicate rows in French data based on person names and personal data. The fields in the personal data column must contain a single type of data, such as telephone number, email, or Institut National de la Statistique et des Études Économiques (INSEE) number. The mapplet generates group keys generated from personal data.
mplt_FRA_Individual_Name_and_Date_Match	Uses field match strategies to identify duplicate rows based on French person names and date data. The mapplet generates group keys from the dates.
mplt_FRA_Individual_Name_and_Email_Match	Uses field match strategies to identify duplicate rows based on French person names and email addresses. The mapplet generates group keys from the email address data.
mplt_FRA_Individual_Name_and_INSEE_Match	Uses field match strategies to identify duplicate rows based on French person names and the INSEE numbers. The mapplet generates group keys generated from the INSEE number data.
mplt_FRA_Individual_Name_Match	Uses field match strategies to identify duplicate rows based on French person names. The mapplet generates NYSIIS codes from the surname values and uses the NYSIIS codes as group keys.
rule_Company_Name_MatchScore	Generates a match score based on company names.

Name	Description
rule_FRA_Company_Name_and_Address_MatchScore	Generates a match score based on company names and French addresses.
rule_FRA_Familyname_and_Address_MatchScore	Generates a match score based on surnames and French addresses.
rule_FRA_Firstname_and_INSEE_MatchScore	Generates a match score based on first names and any data in the personal data column such as telephone number, email, or the INSEE number.
rule_FRA_Firstname_Surname_DOB_and_Postcode_MatchScore	Generates a match score based on the surnames, dates of birth, and postal codes.
rule_FRA_Individual_Name_and_INSEE_MatchScore	Generates a match score based on the person names and the INSEE numbers.
rule_Individual_Name_and_Date_MatchScore	Generates a match score based on person names and dates.
rule_Individual_Name_and_Email_MatchScore	Generates a match score based on person names and email addresses.
rule_Individual_Name_MatchScore	Generates a match score based on person names.

France Demonstration Mappings

The demonstration mappings in the France accelerator use multiple rules to demonstrate data quality processes.

Find the demonstration mappings in the following repository location:

```
[Informatica_DQ_Content]\Rules_Demo\FRA_Accelerator
```

The accelerator contains the following demonstration mappings:

m_FRA_customer_data_demo

Parses, standardizes, and validates French data.

m_FRA_customer_matching_demo

Parses and standardizes identity data from France and performs identity match analysis on the data.

The mapping analyzes the following data combinations and generates match clusters for each combination:

- Person name and address data
- Person name and telephone number

CHAPTER 9

Germany Accelerator

This chapter includes the following topics:

Germany Accelerator Overview

Use the rules in the Germany accelerator to verify and enhance data from organizations in Germany.

The Germany accelerator includes rules that perform the following data quality processes:

- Address data cleansing
- Contact data cleansing
- Corporate data cleansing
- General data cleansing
- Matching and deduplication

The accelerator depends on data cleansing rules that the Core accelerator installs.

Germany Address Data Cleansing Rules

Use the address data cleansing rules to parse, standardize, and validate address data.

Find the address data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
```

The following table describes the address data cleansing rules in the Germany accelerator:

Name	Description
rule_DEU_Address_Parse_Hybrid	Parses unstructured German addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_DEU_Address_Parse_Multiline	Parses unstructured German addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_DEU_Address_Validation_Discrete_w_Geocoding	Validates the deliverability of German addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_DEU_Address_Validation_Discrete	Validates the deliverability of German addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_DEU_Address_Validation_Hybrid_w_Geocoding	Validates the deliverability of German addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_DEU_Address_Validation_Hybrid	Validates the deliverability of German addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_DEU_Address_Validation_Multiline_w_Geocoding	Validates the deliverability of German addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_DEU_Address_Validation_Multiline	Validates the deliverability of German addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.

Germany Composite Rules

Use the composite rules in the Germany accelerator to add a set of rules to a mapping as a single object. A composite rule is a rule that makes use of the logic of other accelerator rules.

Find the composite rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Composite_Rules
```

Composite Rule for German Contact Data

The following table describes the composite rule for contact data in the Germany accelerator:

Name	Description
rule_DEU_Contact_Data	Parses, standardizes, and validates German contact data, such as addresses and telephone numbers.

The rule `rule_DEU_Contact_Data` reads mapplets from multiple folders in the repository. The rule also includes a nonreusable transformation.

The following table lists the names and the repository locations of the rules and the transformation in `rule_DEU_Contact_Data`:

Rule	Location
Case_Converter	Nonreusable transformation
rule_Assign_DQ_Mailability_Score_Description	[Informatica_DQ_Content]\Rules\General_Data_Cleansing
rule_DEU_Address_Validation_Hybrid	[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
rule_DEU_Company_Name_Standardization	[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing
rule_DEU_Gender_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_DEU_Multi_Person_Name_Parse	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_DEU_Phone_Number_Standardization	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_DEU_Phone_Number_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_DEU_Prename_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_DEU_Salutation Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_Email_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

Germany Contact Data Cleansing Rules

Use the contact data cleansing rules to parse, standardize, and validate data about business contacts and individuals.

Find the contact data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
```

The following table describes the contact data cleansing rules in the Germany accelerator:

Name	Description
rule_DEU_Gender_Assignment	Assigns gender according to first names. The rule returns "M" for male names, "F" for female names, and "U" if the gender is unknown. For example, the rule assigns the name "Hans Adler" a gender of "M" for male.
rule_DEU_Given_Name_Standard	Generates given names from German nicknames.
rule_DEU_Multi_Person_Name_Parse	Parses person name values into separate ports. The rule creates ports for values such as title, first name, middle name, and surname. The rule output includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping. When the name data identifies more than one person, the rule creates an output port for each full name. For example, the rule can read the name "Hans und Maria Adler" and create output ports for "Hans Adler" and "Maria Adler."
rule_DEU_Personal_Name_Parsing_FML	Parses the values in a person name into separate ports. The rule creates the ports in the following sequence: - First name, middle name, last name The rule output also includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping.
rule_DEU_Personal_Name_Parsing_LFM	Parses the values in a person name into separate ports. The rule creates the ports in the following sequence: - Last name, first name, middle name The rule output also includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping.
rule_DEU_Phone_Number_Parse	Parses a German telephone number from a string. The rule parses the first telephone number in the data, reading from right to left. The rule recognizes telephone numbers that use leading zeros, international dialing codes, or extensions that begin with the hash symbol. The rule processes the following punctuation symbols: the plus sign, parentheses, and the hash symbol. Before you run the rule, remove all other punctuation, including double spaces. The rule returns a telephone number and also returns a string that contains the input text with the telephone number removed.
rule_DEU_Phone_Number_Standardization	Standardizes German telephone numbers to international and local dialing formats. The rule recognizes telephone numbers that use leading zeros, international dialing codes, or extensions that begin with the hash symbol.
rule_DEU_Phone_Number_Validation	Validates the area code and length of German telephone numbers. The rule returns the region of the telephone number, as well as codes that indicate if the area code and length of a telephone number are valid.

Name	Description
rule_DEU_Prename_Assignment	Generates an honorific according to the gender.
rule_DEU_Salutation_Assignment	Generates formal and casual greetings from prenames and name tokens. For example, when input data contains "Herr Hans Adler," the rule generates the formal greeting "Sehr geehrter Herr Adler," and the casual greeting "Lieber Hans,". You can change the prefix and punctuation by editing the variables in the dq_Generate_Salutation Expression transformation.

Dependencies on Core Contact Data Cleansing Rules

The Germany accelerator depends on the following contact data cleansing rules from the Core accelerator:

- rule_Email_Validation

For more information about these rules, see "Core Contact Data Cleansing Rules" on page 21.

Germany Corporate Data Cleansing Rules

Use the corporate data cleansing rules to standardize corporate data.

Find the corporate data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing

The following table describes the corporate data cleansing rule in the Germany accelerator:

Name	Description
rule_DEU_Company_Name_Standardization	Uses reference tables to standardize company names.

Germany General Data Cleansing Rules

Use the general data cleansing rules to identify the type of information contained in input fields.

Find the general data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\General_Data_Cleansing

The following table describes the general data cleansing rules in the Germany accelerator:

Name	Description
rule_DEU_NER_Field_Identification	Identifies the type of information contained in an input field. The rule can identify names, Personal IDs, company names, dates, and German address data. The rule returns a label that describes the type of input data. The rule uses reference data to identify the types of information. The rule uses probabilistic matching techniques to identify the types of information.

Dependencies on Core General Data Cleansing Rules

The Germany accelerator depends on the following general data cleansing rules from the Core accelerator:

- rule_Assign_DQ_GeocodingStatus_Description
- rule_Assign_DQ_Mailability_Score_Description
- rule_Assign_DQ_Match_Code_Descriptions
- rule_Remove_Extra_Spaces
- rule_Remove_Hyphen
- rule_Remove_Leading_Zero
- rule_Remove_Parentheses
- rule_Remove_Period_Parentheses
- rule_Remove_Punctuation
- rule_Remove_Punctuation_and_Space
- rule_Remove_Space
- rule_Replace_Limited_Punct_with_Space
- rule_UpperCase

For more information about these rules, see "Core General Data Cleansing Rules" on page 22.

Germany Matching and Deduplication Rules

Use the matching and deduplication rules to generate match scores and identify duplicate records.

Find the matching and deduplication rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Matching_Deduplication

The following table describes the matching and deduplication rules in the Germany accelerator:

Name	Description
mplt_Company_Name_Match	Uses field match strategies to identify duplicate rows based on company names. The mapplet generates Soundex codes from the company name values and uses the Soundex codes as group keys.
mplt_DEU_Company_Name_and_Address_Match	Uses field match strategies to identify duplicate rows in German data based on company name and address data. The mapplet uses a combination of characters from the company name values and the postal code values to generate group keys.
mplt_DEU_Familyname_and_Address_Match	Uses field match strategies to identify duplicate rows in German data based on surname and address data. The mapplet uses a combination of characters from the surname values and the postal code values to generate group keys.
mplt_DEU_Firstname_3CharsSurname_DOB_and_Postcode_Match	Uses field match strategies to identify duplicate rows in German data based on personal names, first three characters of the family names, date of birth, and postal codes. The mapplet generates group keys from the postal code data.

Name	Description
mplt_DEU_Firstname_and_PID_Match	Uses field match strategies to identify duplicate rows in German data based on personal names and personal IDs grouped. The mapplet generates group keys from the personal ID data.
mplt_DEU_Firstname_Surname_2ElementsDOB_and_Postcode_Match	Uses field match strategies to identify duplicate rows in German data based on personal names, two elements of the date of birth, and postal codes. The mapplet generates group keys from the postal code data.
mplt_DEU_Firstname_Surname_DOB_and_Postcode_Match	Uses field match strategies to identify duplicate rows in German data based on personal names, date of birth, and postal codes. The mapplet generates group keys from the postal code data.
mplt_DEU_IMO_Company_Name_and_Address_Match	Uses identity match strategies to identify duplicate rows in German data based on company names and addresses. The mapplet generates group keys from the postal code data.
mplt_DEU_IMO_Familyname_and_Address_Match	Uses identity match strategies to identify duplicate rows in German data based on surnames and addresses. The mapplet generates group keys from the postal code data.
mplt_DEU_IMO_Individual_Name_and_Address_Match	Uses identity match strategies to identify duplicate rows in German data based on person names and addresses. The mapplet generates group keys from the postal code data.
mplt_DEU_IMO_Personal_Name_and_Data_Match	Uses identity match strategies to identify duplicate rows in German data based on person names and personal data. The fields in the personal data column must contain a single type of data, such as telephone number, email, or personal ID. The mapplet generates group keys from the personal data.
mplt_DEU_Individual_Name_and_Date_Match	Uses field match strategies to identify duplicate rows based on person names and date data grouped by date. The mapplet generates group keys from the date data.
mplt_DEU_Individual_Name_and_Email_Match	Uses field match strategies to identify duplicate rows in German data based on person names and email addresses. The mapplet generates group keys from the email address data.
mplt_DEU_Individual_Name_and_Phone_Match	Uses field match strategies to identify duplicate rows in German data based on person names and telephone numbers. The mapplet generates group keys from the telephone number data.
mplt_DEU_Individual_Name_and_PID_Match	Uses field match strategies to identify duplicate rows in German data based on person names and the personal IDs. The mapplet generates group keys from the personal ID data.
mplt_DEU_Individual_Name_Match	Uses field match strategies to identify duplicate rows in German data based on person names. The mapplet generates NYSIIS codes from the surname values and uses the NYSIIS code as group keys.
rule_Company_Name_MatchScore	Generates a match score based on company names.
rule_DEU_Company_Name_and_Address_MatchScore	Generates a match score based on company names and addresses.
rule_DEU_Familyname_and_Address_MatchScore	Generates a match score based on surnames and addresses.

Name	Description
rule_DEU_Firstname_3CharsSurname_DOB_and_Postcode_MatchScore	Generates a match score based on the first names, the first three characters of surnames, the date of birth, and the postal codes.
rule_DEU_Firstname_and_PID_MatchScore	Generates a match score based on first names and any data in the personal data column such as telephone number, email, or personal ID.
rule_DEU_Firstname_Surname_2ElementsDOB_and_Postcode_MatchScore	Generates a match score based on personal names, date of birth, and postal codes. **Note:** The input format of the date of birth is assumed to be DD/MM/YYYY.
rule_DEU_Firstname_Surname_DOB_and_Postcode_MatchScore	Generates a match score based on the surnames, date of birth, and postal codes.
rule_DEU_Individual_Name_and_Phone_MatchScore	Generates a match score based on the person names and the telephone numbers.
rule_Familyname_and_Address_MatchScore	Generates a match score based on the family names and addresses.
rule_Individual_Name_and_Date_MatchScore	Generates a match score based on person names and dates.
rule_Individual_Name_and_Email_MatchScore	Generates a match score based on person names and email addresses.
rule_Individual_Name_and_SSN_MatchScore	Generates a match score based on the firstnames and any data in the personal data column such as telephone number, email, or the SSN number.
rule_Individual_Name_MatchScore	Generates a match score based on person names.

Germany Demonstration Mappings

The demonstration mappings in the Germany accelerator use multiple rules to demonstrate data quality processes.

Find the demonstration mappings in the following repository location:

    [Informatica_DQ_Content]\Rules_Demo\DEU_Accelerator

The accelerator contains the following demonstration mappings:

m_DEU_customer_data_demo

Parses, standardizes, and validates Germany data.

m_DEU_customer_matching_demo

Parses and standardizes identity data from Germany and performs identity match analysis on the data.

The mapping analyzes the following data combinations and generates match clusters for each combination:

- Person name and address data
- Person name and telephone number

CHAPTER 10

India Accelerator

This chapter includes the following topics:

India Accelerator Overview

Use the rules in the India accelerator to verify and enhance data from organizations in India.

The India accelerator includes rules that perform the following data quality processes:

- Address data cleansing
- Contact data cleansing
- Corporate data cleansing
- General data cleansing
- Financial data cleansing
- Matching and deduplication
- Product data cleansing

The accelerator depends on data cleansing rules that the Core accelerator installs.

India Address Data Cleansing Rules

Use the address data cleansing rules to parse, standardize, and validate address data.

Find the address data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
```

The following table describes the address data cleansing rules in the India accelerator:

Name	Description
mplt_IND_Address_Info_From_Postcode	Derives the names of subdivisions such as state, district, and taluk from the input postcode.
mplt_IND_Latitude_Longitude	Returns the latitude and longitude coordinates of each input address. The rule also returns state names.
rule_IND_Address_Parse_Hybrid	Parses unstructured India addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_IND_Address_Parse_Multiline	Parses unstructured India addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_IND_Address_Validation_Discrete	Validates the deliverability of India addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_IND_Address_Validation_Hybrid	Validates the deliverability of India addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_IND_Address_Validation_Multiline	Validates the deliverability of India addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_IND_Postcode_Parse	Parses the postcode from the input data.
rule_IND_Postcode_Standardize	Standardizes India postal codes to the standard format.
rule_IND_Postcode_Validation	Validates India postal codes. The rule matches standardized postal codes with valid India postal codes. If the rule does not find a matching postal code, it verifies whether the postal code follows the standard India postal code pattern.
rule_IND_State_Validation	Validates India state names.

India Contact Data Cleansing Rules

Use the contact data cleansing rules to parse, standardize, and validate data about business contacts and individuals.

Find the contact data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
```

The following table describes the contact data cleansing rules in the India accelerator:

Name	Description
rule_IND_Aadhar_Number_Validation	Validates Aadhar numbers or unique identification numbers in India data. The rule validates an Aadhar number based on the Verhoeff algorithm. The number contains twelve digits, and the twelfth digit is a checksum.
rule_IND_Gender_Assignment	Assigns gender according to first name in India data. The rule returns "M" for male names, "F" for female names, and "U" if the gender is unknown. For example, the rule assigns the name "Neha Kapur" a gender of "F" for female.
rule_IND_Given_Name_Standard	Returns standard given name equivalents for common nicknames in India data. For example, the rule returns Vishwanath for Vishwa.
rule_IND_Phone_Number_Parse	Parses an India telephone number from the input data. The rule parses the first telephone number in the data, reading from left to right. The rule returns a telephone number and also returns a string that contains the input data with the telephone number removed.
rule_IND_Phone_Number_Standardization	Standardizes India telephone numbers. The rule returns the telephone number in the following formats: - Standard. (nn) nnn- nnnnnn - Dashes. nn-nnn-nnnnnnn - No spaces. nnnnnnnnnn
rule_IND_Phone_Number_Validation	Validates the area code and length of India telephone numbers. The rule returns the region of the telephone number, as well as codes that indicate if the area code and length of a telephone number are valid.

India Corporate Data Cleansing Rules

Use the corporate data cleansing rules in the India accelerator to standardize and validate corporate data.

Find the corporate data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing

The following table describes the corporate data cleansing rules in the India accelerator:

Name	Description
mplt_IND_Company_Data	Parses data in India input records into categories of corporate information, such as address, corporate listing status, and principal business activity.
rule_IND_CIN_Parse	Parses India corporate identification numbers (CIN) from input data.
rule_IND_Company_CIN_Standardize_and_Validate	Standardizes and validates India corporate identification numbers (CIN) in input data. The rule returns the 21-character number in the format $xnnnnnnxxnnnnxxxnnnnnn$.
rule_IND_Company_Name_Standardization	Standardizes India company names to reference table values.

India Financial Data Cleansing Rules

Use the financial data cleansing rules to parse, standardize, and validate financial data.

Find the financial data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Financial_Data_Cleansing
```

The following table describes the financial data cleansing rules in the India accelerator:

Name	Description
rule_IND_Bank_Data_Specifics	Derives the bank name, branch name, and branch code that an Indian Financial System Code (IFSC) represents. An IFSC is an eleven-character code in alpha-numeric format that the Reserve Bank of India uses to identify all bank branches within the NEFT (National Electronic Funds Transfer) network.
rule_IND_Bank_IFSC_Code_Validation	Validates an IFSC code. An IFSC is an eleven-character code in alpha-numeric format that the Reserve Bank of India uses to identify all bank branches within the NEFT (National Electronic Funds Transfer) network.
rule_IND_Bank_Name_From_IFSC_Code	Derives the bank name that an Indian Financial System Code (IFSC) represents.

India General Data Cleansing Rules

Use the general data cleansing rules to identify the type of information contained in input fields.

Find the general data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\General_Data_Cleansing
```

The following table describes the general data cleansing rules in the India accelerator:

Name	Description
rule_UpperCase_Multiport	Returns all alphabetic characters in upper case across multiple ports.
rule_Verhoeff_Algorithm	Validates the check digit of a number according to the Verhoeff Algorithm. The check digit is the final digit in the number. Returns Valid if the check digit is correct. Returns Invalid for an incorrect check digit and any non-numeric input.

Dependencies on Core General Data Cleansing Rules

The India accelerator depends on the following general data cleansing rules from the Core accelerator:

- rule_Assign_DQ_Match_Code_Description
- rule_Remove_Extra_Spaces
- rule_Remove_Leading_Zero
- rule_Remove_Limited_Punctuation
- rule_Remove_Non_Numbers

- rule_Remove_Punctuation
- rule_Remove_Punctuation_and_Space
- rule_Replace_Limited_Punct_with_Space
- rule_UpperCase

For more information about these rules, see "Core General Data Cleansing Rules" on page 22.

India Matching and Deduplication Rules

Use matching and deduplication rules to generate match scores and identify duplicate records.

Matching and deduplication rules install to the following repository location:

```
[Informatica_DQ_Content]\Rules\Matching_Deduplication
```

The following table describes the matching and deduplication rules:

Name	Description
mplt_IND_IMO_Company_Name_and_Address_Match	Uses identity match strategies to identify duplicate rows for India data based on company names and addresses. The mapplet generates group keys from the postal code data.
mplt_IND_IMO_Familyname_and_Address_Match	Uses identity match strategies to identify duplicate rows in India data based on family names and addresses. The mapplet generates group keys from the postal code data.
mplt_IND_IMO_Individual_Name_and_Address_Match	Uses identity match strategies to identify duplicate rows in India data based on person names and addresses. The mapplet generates group keys from the postal code data.
mplt_IND_IMO_Personal_Name_and_Data_Match	Uses identity match strategies to identify duplicate rows in India data based on person names and personal data. The fields in the personal data column must contain a single type of data, such as telephone number, driver license number, tax number, or email address. The mapplet generates group keys from the personal data.
mplt_IND_Individual_Name_and_Address_Match	Uses field match strategies to identify duplicate rows in India data based on person names and address data. The mapplet generates a NYSIIS value from the first four digits of the postcode, the first letter of the surname, and the first five characters of the first name. The mapplet uses the NYSIIS value to generate group keys.
mplt_IND_Individual_Name_and_Date_Match	Uses field match strategies to identify duplicate rows in India data based on person names and dates. The mapplet generates group keys from the date data.
mplt_IND_Individual_Name_and_Email_Match	Uses field match strategies to identify duplicate rows in India data based on person names and email addresses. The mapplet generates group keys from email address data.
mplt_IND_Individual_Name_and_Phone_Match	Uses field match strategies to identify duplicate rows in India data based on person names and telephone numbers. The mapplet generates group keys from the telephone number data.

Name	Description
mplt_IND_Individual_Name_Match h	Uses field match strategies to identify duplicate rows in India data based on person names. The mapplet generates NYSIIS codes from the surname values and uses the NYSIIS codes as group keys.
rule_IND_Individual_Name_and_ Address_MatchScore	Compares person name and address fields across records in India data and generates a match score.
rule_IND_Individual_Name_and_ Date_MatchScore	Compares person name and date fields across records in India data and generates a match score.
rule_IND_Individual_Name_and_ Email_MatchScore	Compares person name and email fields across records in India data and generates a match score.
rule_IND_Individual_Name_and_ Phone_MatchScore	Compares person name and telephone number fields across records in India data and generates a match score.
rule_IND_Individual_Name_Matc hScore	Compares person name fields across records in India data and generates a match score.

India Product Data Cleansing Rules

Use the product data cleansing rules to parse, standardize, and validate product data.

Find the product data cleansing rules in the following repository location:

`[Informatica_DQ_Content]\Rules\Product_Data_Cleansing`

The following table describes the product data cleansing rules in the India accelerator:

Name	Description
mplt_IND_Product_Info	Standardizes product codes in India data and returns the product description and sector for each product code.

CHAPTER 11

Italy Accelerator

This chapter includes the following topics:

Italy Accelerator Overview

Use the rules in the Italy accelerator to verify and enhance data from organizations in Italy.

The Italy accelerator includes rules that perform the following data quality processes:

- Address data cleansing
- Contact data cleansing
- Corporate data cleansing
- Financial data cleansing
- General data cleansing
- Matching and deduplication

Italy Address Data Cleansing Rules

Use the address data cleansing rules to parse, standardize, and validate address data.

Find the address data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
```

The following table describes the address data cleansing rules in the Italy accelerator:

Name	Description
rule_ITA_Address_Parse_Hybrid	Parses unstructured Italy addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_ITA_Address_Parse_Multiline	Parses unstructured Italy addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_ITA_Address_Validation_Discrete	Validates the deliverability of Italy addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_ITA_Address_Validation_Discrete_w_Geocoding	Validates the deliverability of Italy addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_ITA_Address_Validation_Hybrid	Validates the deliverability of Italy addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_ITA_Address_Validation_Hybrid_w_Geocoding	Validates the deliverability of Italy addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_ITA_Address_Validation_Multiline	Validates the deliverability of Italy addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_ITA_Address_Validation_Multiline_w_Geocoding	Validates the deliverability of Italy addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_ITA_Comune_Validation	Validates a comune name in Italy. A comune is a basic administrative division in Italy, roughly equivalent to a township or municipality.
rule_ITA_External_Postcode_Validation	Validates that an input string is a valid post code for a mail item that originates outside of Italy. An external post code includes the prefix I or IT.
rule_ITA_Local_Postcode_Validation	Validates that an input string is a valid post code for a mail item that originates within Italy. An internal post code does not use a prefix.
rule_ITA_Postcode_Parse	Parses an internal or external Italian post code from an input string.
rule_ITA_Postcode_Standardization	Standardizes the postcode in an Italy address to the standard format.

Italy Contact Data Cleansing Rules

Use the contact data cleansing rules to parse, standardize, and validate data about business contacts and individuals.

Find the contact data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

The following table describes the contact data cleansing rules in the Italy accelerator:

Name	Description
rule_ITA_Gender_Assignment	Assigns gender according to first names. The rule returns "M" for male names, "F" for female names, and "U" if the gender is unknown. For example, the rule assigns the name "Alessandro Scarlatti" a gender of "M" for male.
rule_ITA_Given_Name_Standardization	Returns a standard name for a common nickname.
rule_ITA_Lastname_Validation	Validates that a name is an Italian surname.
rule_ITA_Phone_Number_Parse	Parses an Italian telephone number from a string. The rule parses the first telephone number in the data, reading from right to left. The rule returns a telephone number and also returns a string that contains the input text with the telephone number removed.
rule_ITA_Phone_Number_Standardization	Standardizes Italian telephone numbers to international and local dialing formats. The rule recognizes telephone numbers that use leading zeros, international dialing codes, or extensions that begin with the hash symbol.
rule_ITA_Phone_Number_Validation	Validates the area code and length of Italian telephone numbers. The rule returns the region of the telephone number, as well as codes that indicate if the area code and length of a telephone number are valid.
rule_ITA_Prename_Assignment	Generates an honorific according to the gender.
rule_ITA_Salutation_Assignment	Generates formal and casual greetings from prenames and name tokens. You can change the prefix and punctuation by editing the variables in the dq_Generate_Salutation Expression transformation.

Italy Corporate Data Cleansing Rules

Use the corporate data cleansing rules in the Italy accelerator to standardize corporate data.

Find the corporate data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing

The following table describes the corporate data cleansing rules in the Italy accelerator:

Name	Description
rule_ITA_Company_Name_Standardization	Standardizes Italian company names to reference table values.

Italy Financial Data Cleansing Rules

Use the financial data cleansing rules to parse and validate financial data.

Find the financial data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Financial_Data_Cleansing

The following table describes the financial data cleansing rules in the India accelerator:

Name	Description
rule_ITA_Codice_Fiscale_Parse	Parses an Italian fiscal code, or Codice fiscale, from a string.
rule_ITA_Codice_Fiscale_Validation	Validates an Italian fiscal code, or Codice fiscale.
rule_ITA_VAT_Number_Validation	Validates an Italian VAT (Value Added Tax) number.

Italy General Data Cleansing Rules

Use the general data cleansing rules to perform general validation and standardization tasks on data and to describe the information in address status fields.

Find the general data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\General_Data_Cleansing

The following table describes the general data cleansing rules in the Italy accelerator:

Name	Description
rule_Assign_DQ_GeocodingStatus_Description	Assigns a description to the Geocoding Status output from the Address Validator transformation. The description corresponds to the output from Data Quality transformations in releases prior to Data Quality 9.0.
rule_Assign_DQ_Mailability_Score_Description	Assigns a description to the Mailability Score output from the Address Validator transformation. The description corresponds to the output from Data Quality transformations in releases prior to Data Quality 9.0.
rule_Assign_DQ_Match_Code_Description	Assigns a description to the Match Code output from the Address Validator transformation. The description corresponds to the output from Data Quality transformations in releases prior to Data Quality 9.0.
rule_Date_Validation	Validates date strings that appear in a single format in a data column. To configure the date format that the rule uses for validation, open the dq_ValidateDate Expression transformation in the rule and update the In_Date_Format expression variable. The default format is "MM/DD/YYYY." The rule returns "Valid" or "Invalid."
rule_ITA_Registration_Number_Plate_Validation	Validates an Italian vehicle registration number.

Name	Description
rule_Remove_Extra_Spaces	Replaces all consecutive spaces with a single space and trims leading and trailing spaces.
rule_Remove_Hyphen	Removes hyphens.
rule_Remove_Non_Numbers	Removes all characters that are not numeric.
rule_Remove_Period_Parentheses	Removes the following characters: - Left and right parentheses - Periods
rule_Remove_Punctuation	Removes punctuation symbols.
rule_Remove_Punctuation_and_Space	Removes all punctuation and all space characters.
rule_Remove_Space	Removes all character spaces.
rule_Replace_Limited_Punct_with_Space	Replaces the following punctuation characters with a single space: dash, back slash, period, exclamation mark, and underscore. The rule also replaces two, three, and four consecutive spaces with a single space.
rule_Replace_Punctuation_with_Space	Replaces all punctuation with spaces.
rule_UpperCase	Returns all alphabetic characters in upper case.

Italy Matching and Deduplication Rules

Use matching and deduplication rules to generate match scores and identify duplicate records.

Matching and deduplication rules install to the following repository location:

    [Informatica_DQ_Content]\Rules\Matching_Deduplication

The following table describes the matching and deduplication rules:

Name	Description
mplt_Company_Name_Match	Uses field match strategies to identify duplicate rows based on company names. The mapplet generates Soundex codes from the company name values and uses the Soundex codes as group keys.
mplt_ITA_Company_Name_and_Address_Match	Identifies duplicate rows in data based on company name and address data from Italy. The mapplet uses a combination of characters from the company name values and the postal code values to generate group keys.
mplt_ITA_Familyname_and_Address_Match	Identifies duplicate rows based on surname and address data from Italy. The mapplet uses a combination of characters from the surname values and the postal code values to generate group keys.

Name	Description
mplt_ITA_Firstname_3CharsSurname_DOB_and_Postcode_Match	Uses field match strategies to identify duplicate rows in Italian data based on person names, the first three characters of the family names, dates of birth, and postal codes. The mapplet generates group keys from the postal code data.
mplt_ITA_Firstname_and_PID_Match	Uses field match strategies to identify duplicate rows in Italian data based on person names and personal identification numbers. The mapplet generates group keys from the personal ID data.
mplt_ITA_Firstname_Surname_2ElementsDOB_and_Postcode_Match	Uses field match strategies to identify duplicate rows in Italian data based on person names, two elements of the date of birth, and postal codes. The mapplet generates group keys from the postal code data.
mplt_ITA_Firstname_Surname_DOB_and_Postcode_Match	Uses field match strategies to identify duplicate rows in Italian data based on person names, dates of birth, and postal codes. The mapplet generates group keys from the postal code data.
mplt_ITA_IMO_Familyname_and_Address_Match	Uses identity match strategies to identify duplicate rows in Italian data based on family names and addresses. The mapplet generates group keys from the postal code data.
mplt_ITA_Individual_Name_and_Phone_Match	Uses field match strategies to identify duplicate rows based on Italian person names and telephone numbers. The mapplet generates group keys from the telephone number data.
mplt_ITA_Individual_Name_Match	Uses field match strategies to identify duplicate rows in Italian data based on person names. The mapplet generates NYSIIS codes from the surname values and uses the NYSIIS codes as group keys.
rule_Company_Name_MatchScore	Generates a match score based on company names.
rule_Individual_Name_MatchScore	Generates a match score based on person names.
rule_ITA_Company_Name_and_Address_MatchScore	Generates a match score based on company names and Italian address data.
rule_ITA_Familyname_and_Address_MatchScore	Generates a match score based on surnames and Italian address data.
rule_ITA_Firstname_3CharsSurname_DOB_and_Postcode_MatchScore	Generates a match score based on first names, the first three characters of the surnames, dates of birth, and postal codes.
rule_ITA_Firstname_and_PID_MatchScore	Generates a match score based on first names and personal identification numbers.
rule_ITA_Firstname_Surname_2ElementsDOB_and_Postcode_MatchScore	Generates a match score based on the following information: - Person names - Any two date of birth elements, such as month and year - Postal codes The rule assumes that the date format is DD/MM/YYYY.

Name	Description
rule_ITA_Firstname_Surname_DOB_and_Postcode_MatchScore	Generates a match score based on the surnames, dates of birth, and postal codes.
rule_ITA_Individual_Name_and_Phone_MatchScore	Generates a match score based on person names and telephone numbers.

Italy Demonstration Mappings

The demonstration mappings in the Italy accelerator use multiple rules to demonstrate data quality processes.

Find the demonstration mappings in the following repository location:

    [Informatica_DQ_Content]\Rules_Demo\ITA_Accelerator

The accelerator contains the following demonstration mappings:

m_ITA_customer_data_demo

Parses, standardizes, and validates Italian data.

m_ITA_customer_matching_demo

Parses and standardizes identity data from Italy and performs identity match analysis on the data.

The mapping analyzes the following data combinations and generates match clusters for each combination:

- Person name and address data
- Person name and telephone number

CHAPTER 12

Portugal Accelerator

This chapter includes the following topics:

Portugal Accelerator Overview

Use the rules in the Portugal accelerator to verify and enhance data from organizations in Portugal.

The Portugal accelerator includes rules that perform the following data quality processes:

- Address data cleansing
- Contact data cleansing
- Corporate data cleansing
- General data cleansing
- Matching and deduplication

The accelerator depends on data cleansing rules that the Core accelerator installs.

Portugal Address Data Cleansing Rules

Use the address data cleansing rules to parse and validate address data.

Find the address data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
```

The following table describes the address data cleansing rules in the Portugal accelerator:

Name	Description
rule_PRT_Address_Parse_Hybrid	Parses unstructured Portuguese addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_PRT_Address_Parse_Multiline	Parses unstructured Portuguese addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_PRT_Address_Validation_Discrete_w_Geocoding	Validates the deliverability of Portuguese addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_PRT_Address_Validation_Discrete	Validates the deliverability of Portuguese addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_PRT_Address_Validation_Hybrid_w_Geocoding	Validates the deliverability of Portuguese addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_PRT_Address_Validation_Hybrid	Validates the deliverability of Portuguese addresses and. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_PRT_Address_Validation_Multiline_w_Geocoding	Validates the deliverability of Portuguese addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_PRT_Address_Validation_Multiline	Validates the deliverability of Portuguese addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.

Portugal Composite Rules

Use the composite rules in the Portugal accelerator to add a set of rules to a mapping as a single object. A composite rule is a rule that makes use of the logic of other accelerator rules.

Find the composite rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Composite_Rules
```

Composite Rule for Portuguese Contact Data

The following table describes the composite rule for Portuguese contact data in the Portugal accelerator:

Name	Description
rule_PRT_Contact_Data	Parses, standardizes, and validates Portuguese contact data, such as addresses, telephone numbers, and Número de Identificação Fiscal (NIF) numbers.

The rule `rule_PRT_Contact_Data` reads mapplets from multiple folders in the repository. The rule also includes a nonreusable transformation.

The following table lists the names and the repository locations of the rules and the transformation in `rule_PRT_Contact_Data`:

Rule	Location
Case_Converter	Nonreusable transformation
rule_Assign_DQ_Mailability_Score_Description	[Informatica_DQ_Content]\Rules\General_Data_Cleansing
rule_Email_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_PRT_Address_Validation_Hybrid	[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
rule_PRT_Company_Name_Standardization	[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing
rule_PRT_NIF_Standardization	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_PRT_NIF_Validate	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_PRT_Personal_Name_Parse_Validate	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_PRT_Phone_Number_Standardization	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_PRT_Phone_Number_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_PRT_Prename_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_PRT_Salutation_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

Portugal Contact Data Cleansing Rules

Use the contact data cleansing rules to parse, standardize, and validate data about business contacts and individuals.

Find the contact data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
```

The following table describes the contact data cleansing rules in the Portugal accelerator:

Name	Description
rule_PRT_Gender_Assignment	Assigns gender according to first names. The rule returns "M" for male names, "F" for female names, and "U" if the gender is unknown. For example, the rule assigns the name "Artur Cruz" a gender of "M" for male.
rule_PRT_Given_Name_Standard	Generate given names from Portuguese nicknames.
rule_PRT_NIF_Parse	Parses Número de Identificação Fiscal (NIF) numbers from strings. The rule returns the ID numbers and also returns a string that contains the input text with the ID numbers removed.
rule_PRT_NIF_Standardization	Standardizes Número de Identificação Fiscal (NIF) numbers into a nine-digit string. The rule removes alphabetic characters, symbols, and spaces.
rule_PRT_NIF_Validate	Validates Número de Identificação Fiscal (NIF) numbers based on the check digit in each number. The rule requires that the input is a nine-digit numeric string with no spaces.
rule_PRT_Personal_Name_Parse_Validate	Parses person name values into separate ports. The rule creates ports for values such as title, first name, middle name, and surname. The rule also indicates if the name might be a company name and validates the spelling of the name. The rule output includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping.
rule_PRT_Phone_Number_Parse	Parses a Portuguese telephone number from a string. The rule parses the first telephone number in the data, reading from right to left. The rule returns a telephone number and also returns a string that contains the input text with the telephone number removed.
rule_PRT_Phone_Number_Standardization	Standardizes Portuguese telephone numbers to international and local dialing formats.
rule_PRT_Phone_Number_Validation	Validates the area code and length of Portuguese telephone numbers. The rule returns the region of the telephone number, as well as codes that indicate if the area code and length of a telephone number are valid.
rule_PRT_Prename_Assignment	Generates an honorific according to the gender. You can change the female_prename expression variable from "Sra" to "Sta".
rule_PRT_Salutation_Assignment	Generates formal and casual greetings from prenames and name tokens. For example, when input data contains "Sr. Artur Cruz," the rule generates the formal greeting "Prezado Sr. Cruz," and the casual greeting "Prezado Artur,". You can change the prefix and punctuation by editing the variables in the dq_Generate_Salutation Expression transformation.

Dependencies on Core Contact Data Cleansing Rules

The Portugal accelerator depends on the following contact data cleansing rules from the Core accelerator:

- rule_Email_Validation

For more information about these rules, see "Core Contact Data Cleansing Rules" on page 21.

Portugal Corporate Data Cleansing Rules

Use the corporate data cleansing rules to parse, standardize, and validate corporate data.

Find the corporate data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing

The following table describes the corporate data cleansing rules in the Portugal accelerator:

Name	Description
rule_PRT_Company_Name_Standardization	Standardizes Portuguese company names to reference table values.
rule_PRT_NIPC_Parse	Parses a Número de Identificação Pessoa Colectiva (NIPC). The rule returns the NIPC and also returns a string that contains the input text with the NIPC removed.
rule_PRT_NIPC_Standardize	Standardizes a Número de Identificação Pessoa Colectiva (NIPC) into a nine-digit string. The rule removes alphabetic characters, symbols, and spaces.
rule_PRT_NIPC_Validate	Validates a Número de Identificação Pessoa Colectiva (NIPC) based on the check digit in each number. The rule requires that the input is a nine-digit string.

Portugal General Data Cleansing Rules

Use the general data cleansing rules to identify the type of information contained in input fields.

Find the general data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\General_Data_Cleansing

The following table describes the general data cleansing rules in the Portugal accelerator:

Name	Description
rule_PRT_NER_Field_Identification	Identifies the type of information contained in an input field. The rule can identify names, Personal IDs, company names, dates, and Portuguese address data. The rule returns a label that describes the type of input data. The rule uses reference data to identify the types of information. The rule uses probabilistic matching techniques to identify the types of information.

Dependencies on Core General Data Cleansing Rules

The Portugal accelerator depends on the following general data cleansing rules from the Core accelerator:

- rule_Assign_DQ_GeocodingStatus_Description
- rule_Assign_DQ_ElementResultStatus_Description
- rule_Assign_DQ_Mailability_Score_Description
- rule_Assign_DQ_Match_Code_Descriptions
- rule_Parse_First_Word
- rule_Remove_Extra_Spaces

- rule_Remove_Non_Numbers
- rule_Remove_Punctuation
- rule_Remove_Punctuation_and_Space
- rule_Replace_Limited_Punct_with_Space
- rule_UpperCase

For more information about these rules, see "Core General Data Cleansing Rules" on page 22.

Portugal Matching and Deduplication Rules

Use the matching and deduplication rules to generate match scores and identify duplicate records.

Find the matching and deduplication rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Matching_Deduplication

The following table describes the matching and deduplication rules in the Portugal accelerator:

Name	Description
mplt_Company_Name_Match	Uses field match strategies to identify duplicate rows based on company name. The mapplet generates Soundex codes from the company name values and uses the Soundex codes as group keys.
mplt_PRT_Company_Name_and_Address_Match	Uses field match strategies to identify duplicate rows in Portuguese data based on company name and address data. The mapplet uses a combination of characters from the company name values and the postal code values to generate group keys.
mplt_PRT_Familyname_and_Address_Match	Uses field match strategies to identify duplicate rows in Portuguese data based on surname and address data. The mapplet uses a combination of characters from the surname values and the postal code values to generate group keys.
mplt_PRT_Firstname_and_NIF_BI_Match	Uses field match strategies to identify duplicate rows in Portuguese data based on first name and personal identification numbers such as Número de Indentificação Fiscal (NIF) and Bilhete de Identidade (BI). The mapplet generates group keys from the personal identification number data.
mplt_PRT_IMO_Company_Name_and_Address_Match	Uses identity match strategies to identify duplicate rows in Portuguese data based on company names and addresses. The mapplet generates group keys from the postal code data.
mplt_PRT_IMO_Familyname_and_Address_Match	Uses identity match strategies to identify duplicate rows in Portuguese data based on family names and addresses. The mapplet generates group keys from the postal code data.
mplt_PRT_IMO_Individual_Name_and_Address_Match	Uses identity match strategies to identify duplicate rows in Portuguese data based on person names and addresses. The mapplet generates group keys from the postal code data.

Name	Description
mplt_PRT_IMO_Personal_Name_and_Data_Match	Uses identity match strategies to identify duplicate rows in Portuguese data based on person names and personal data. The fields in the personal data column must contain a single type of data, such as telephone number, email, or Número de Indentificação Fiscal (NIF). The mapplet generates group keys generated from the personal data.
mplt_PRT_Individual_Name_and_Address_Match	Uses field match strategies to identify duplicate rows in Portuguese data based on person names and address data. The mapplet uses a combination of characters from the surname values and the postal code values to generate group keys.
mplt_PRT_Individual_Name_and_Date_Match	Uses field match strategies to identify duplicate rows in Portuguese data based on person names and date data. The mapplet generates group keys from the date data.
mplt_PRT_Individual_Name_and_Email_Match	Uses field match strategies to identify duplicate rows in Portuguese data based on person names and email addresses. The mapplet generates group keys from the email address data.
mplt_PRT_Individual_Name_and_Phone_Match	Uses field match strategies to identify duplicate rows in Portuguese data based on person names and telephone numbers. The mapplet generates group keys from the telephone number data.
mplt_PRT_Individual_Name_Match	Uses field match strategies to identify duplicate rows in Portuguese data based on person names. The mapplet generates NYSIIS codes from the surname values and uses the NYSIIS codes as group keys.
rule_Company_Name_MatchScore	Generates a match score based on company names.
rule_PRT_Company_Name_and_Address_MatchScore	Generates a match score based on company names and Portuguese address data.
rule_PRT_Familyname_and_Address_MatchScore	Generates a match score based on surnames and Portuguese address data.
rule_PRT_Firstname_and_NIF_BI_MatchScore	Generates a match score based on first name data, Número de Indentificação Fiscal (NIF), and Bilhete de Identidade (BI) numbers.
rule_PRT_Individual_Name_and_Address_MatchScore	Generates a match score based on person names and Portuguese address data.
rule_PRT_Individual_Name_and_Date_MatchScore	Generates a match score based on person names and dates.
rule_PRT_Individual_Name_and_Email_MatchScore	Generates a match score based on person names and email addresses.
rule_PRT_Individual_Name_and_Phone_MatchScore	Generates a match score based on person names and telephone numbers.
rule_PRT_Individual_Name_MatchScore	Generates a match score based on person names.

Portugal Demonstration Mappings

The demonstration mappings in the Portugal accelerator use multiple rules to demonstrate data quality processes.

Find the demonstration mappings in the following repository location:

    [Informatica_DQ_Content]\Rules_Demo\PRT_Accelerator

The accelerator contains the following demonstration mappings:

m_PRT_customer_data_demo

Parses, standardizes, and validates Portuguese data.

m_PRT_customer_matching_demo

Parses and standardizes identity data from Portugal and performs identity match analysis on the data.

The mapping analyzes the following data combinations and generates match clusters for each combination:

- Person name and address data
- Person name and telephone number

CHAPTER 13

Spain Accelerator

This chapter includes the following topics:

Spain Accelerator Overview

Use the rules in the Spain accelerator to verify and enhance data from organizations in Spain.

The Spain accelerator includes rules that perform the following data quality processes:

- Address data cleansing
- Contact data cleansing
- Corporate data cleansing
- General data cleansing
- Matching and deduplication

The accelerator depends on data cleansing rules that the Core accelerator installs.

Spain Address Data Cleansing Rules

Use the address data cleansing rules to parse, standardize, and validate address data.

Find the address data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
```

The following table describes the address data cleansing rules in the Spain accelerator:

Name	Description
rule_ESP_Address_Parse_Hybrid	Parses unstructured Spanish addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_ESP_Address_Parse_Multiline	Parses unstructured Spanish addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_ESP_Address_Validation_Discrete_w_Geocoding	Validates the deliverability of Spanish addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_ESP_Address_Validation_Discrete	Validates the deliverability of Spanish addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_ESP_Address_Validation_Hybrid_w_Geocoding	Validates the deliverability of Spanish addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_ESP_Address_Validation_Hybrid	Validates the deliverability of Spanish addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_ESP_Address_Validation_Multiline_w_Geocoding	Validates the deliverability of Spanish addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_ESP_Address_Validation_Multiline	Validates the deliverability of Spanish addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.

Spain Contact Data Cleansing Rules

Use the contact data cleansing rules to parse, standardize, and validate data about business contacts and individuals.

Find the contact data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

The following table describes the contact data cleansing rules in the Spain accelerator:

Name	Description
rule_ESP_CIF_Parse	Parses the Spanish Certificado de Identificación Fiscal (CIF).
rule_ESP_CIF_Standardization	Standardizes the Spanish Certificado de Identificación Fiscal (CIF).
rule_ESP_CIF_Validation	Validates the Spanish Certificado de Identificación Fiscal (CIF).
rule_ESP_DNI_Parse	Parses the Spanish Documento Nacional de Identidad (DNI).
rule_ESP_DNI_Standardization	Standardizes the Spanish Documento Nacional de Identidad (DNI).
rule_ESP_DNI_Validation	Validates the Spanish Documento Nacional de Identidad (DNI).
rule_ESP_Gender_Assignment	Assigns gender according to first names. The rule returns "M" for male names, "F" for female names, and "U" if the gender is unknown. For example, the rule assigns the name "Juan Garcia" a gender of "M" for male.
rule_ESP_Given_Name_Standard	Generates given names from Spanish nicknames.
rule_ESP_NIE_Parse	Parses the Spanish Número de Identidad de Extranjero (NIE).
rule_ESP_NIE_Standardization	Standardizes the Spanish Número de Identidad de Extranjero (NIE).
rule_ESP_NIE_Validation	Validates the Spanish Número de Identidad de Extranjero (NIE).
rule_ESP_NIF_Parse	Parses the Spanish Número de Identificación Fiscal (NIF) from a string.
rule_ESP_NIF_Standardization	Standardizes the Spanish Número de Identificación Fiscal (NIF).
rule_ESP_NIF_Validation	Validates the Spanish Número de Identificación Fiscal (NIF).
rule_ESP_Personal_Name_Parse	Parses Spanish person names.
rule_ESP_Phone_Number_Parse	Parses a Spanish telephone number from a string. The rule parses the first telephone number in the data, reading from right to left. The rule returns a telephone number and also returns a string that contains the input text with the telephone number removed.
rule_ESP_Phone_Number_Standardization	Standardizes Spanish telephone numbers to international and local dialing formats. The rule recognizes telephone numbers that use leading zeros, international dialing codes, or extensions that begin with the hash symbol.

Name	Description
rule_ESP_Phone_Number_Validation	Validates the area code and length of Spanish telephone numbers. The rule returns the region of the telephone number, as well as codes that indicate if the area code and length of a telephone number are valid.
rule_ESP_Phone_Parse_Standardize_Validate	Parses Spanish telephone numbers and standardizes format. Validates the area code and length of Spanish telephone numbers.
rule_ESP_Prename_Assignment	Generates an honorific according to the gender.
rule_ESP_Salutation_Assignment	Generates formal and casual greetings from prenames and name tokens. For example, when input data contains " Sr. Juan Garcia," the rule generates the formal greeting " Estimado Sr. Garcia," and the casual greeting "Querido Juan,". You can change the prefix and punctuation by editing the variables in the dq_Generate_Salutation Expression transformation.

Dependencies on Core Contact Data Cleansing Rules

The Spain accelerator depends on the following contact data cleansing rules from the Core accelerator:

- rule_Email_Validation

For more information about these rules, see "Core Contact Data Cleansing Rules" on page 21.

Spain Corporate Data Cleansing Rules

Use the corporate data cleansing rules to standardize corporate data.

The Spain accelerator depends on the following corporate data cleansing rule from the Core accelerator:

- rule_Company_Name_Standardization

Spain General Data Cleansing Rules

Use the general data cleansing rules to identify the type of information contained in input fields.

Find the general data cleansing rules in the following repository location:

    [Informatica_DQ_Content]Rules\General_Data_Cleansing

The following table describes the general data cleansing rules in the Spain accelerator:

Name	Description
rule_ESP_NER_Field_Identification	Identifies the type of information contained in an input field. The rule can identify names, Personal IDs, company names, dates, and Spanish address data. The rule returns a label that describes the type of input data. The rule uses probabilistic matching techniques to identify the types of information.

The Spain accelerator depends on the following general data cleansing rules from the Core accelerator:

- rule_Assign_DQ_ElementResultStatus_Description
- rule_Assign_DQ_GeocodingStatus_Description
- rule_Assign_DQ_Match_Code_Descriptions
- rule_Remove_Extra_Spaces
- rule_Remove_Leading_Zero
- rule_Remove_Limited_Punctuation
- rule_Remove_Non_Numbers
- rule_Remove_Punctuation_and_Space
- rule_Remove_Punctuation
- rule_Replace_limited_Punct_with_Space
- rule_Translate_Diacritic_Characters
- rule_UpperCase

Spain Matching and Deduplication Rules

Use the matching and deduplication rules to generate match scores and identify duplicate records.

Find the matching and deduplication rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Matching_Deduplication
```

The following table describes the matching and deduplication rules in the Spain accelerator:

Name	Description
mplt_Company_Name_Match	Uses field match strategies to identify duplicate rows based on company name.
mplt_ESP_Company_Name_and_Address_Match	Uses field match strategies to identify duplicate rows in Spanish data based on company name and address data. The mapplet uses a combination of characters from the company name values and the postal code values to generate group keys.
mplt_ESP_Familyname_and_Address_Match	Uses field match strategies to identify duplicate rows in Spanish data based on surname and address data. The mapplet uses a combination of characters from the surname values and the postal code values to generate group keys.
mplt_ESP_Firstname_and_NIF_BI_Match	Uses field match strategies to identify duplicate rows in Spanish data based on first name and personal identification numbers, such as the Número de Identificación Fiscal (NIF). The mapplet generates group keys from the personal identification number data.
mplt_ESP_IMO_Company_Name_Match	Uses identity match strategies to identify duplicate rows in Spanish data based on company names. The mapplet generates Soundex codes from the company name values and uses the Soundex codes as group keys.

Name	Description
mplt_ESP_IMO_Company_Name_and_Address_Match	Uses identity match strategies to identify duplicate rows in Spanish data based on company names and addresses. The mapplet generates group keys from the postal code data.
mplt_ESP_IMO_Familyname_and_Address_Match	Uses identity match strategies to identify duplicate rows in Spanish data based on family names and addresses. The mapplet generates group keys generate from the postal code data.
mplt_ESP_IMO_Individual_Name_and_Address_Match	Uses identity match strategies to identify duplicate rows in Spanish data based on person names and addresses. The mapplet generates group keys from the postal code data.
mplt_ESP_IMO_Personal_Name_and_Data_Match	Uses identity match strategies to identify duplicate rows in Spanish data based on person names and personal data. The fields in the personal data column must contain a single type of data, such as telephone number or email. The mapplet generates group keys from the personal data.
mplt_ESP_Individual_Name_Match	Uses field match strategies to identify duplicate rows in Spanish data based on person names. The mapplet generates NYSIIS codes from the surname values and uses the NYSIIS codes as group keys.
mplt_ESP_Individual_Name_and_Address_Match	Uses field match strategies to identify duplicate rows in Spanish data based on person names and address data. The mapplet uses a combination of characters from the surname values and the postal code values to generate group keys.
mplt_ESP_Individual_Name_and_Date_Match	Uses field match strategies to identify duplicate rows in Spanish data based on person names and dates. The mapplet generates group keys from the date data.
mplt_ESP_Individual_Name_and_Email_Match	Uses field match strategies to identify duplicate rows in Spanish data based on email addresses and person names. The mapplet generates group keys from the email address data.
mplt_ESP_Individual_Name_and_Phone_Match	Uses field match strategies to identify duplicate rows in Spanish data based on person names and telephone numbers. The mapplet generates group keys from the telephone number data.
rule_Company_Name_MatchScore	Generates a match score based on company names.
rule_ESP_Company_Name_and_Address_MatchScore	Generates a match score based on company names and Spanish address data.
rule_ESP_Familyname_and_Address_MatchScore	Generates a match score based on surnames and Spanish address data.
rule_ESP_Firstname_and_NIF_BI_Matchscore	Generates a match score based on first name and Número de Identificación Fiscal (NIF) numbers.
rule_ESP_Individual_Name_MatchScore	Generates a match score based on person names.

Name	Description
rule_ESP_Individual_Name_and_Address_MatchScore	Generates a match score based on person names and Spanish address data.
rule_ESP_Individual_Name_and_Date_MatchScore	Generates a match score based on person names and dates.
rule_ESP_Individual_Name_and_Email_MatchScore	Generates a match score based on person names and email addresses.
rule_ESP_Individual_Name_and_Phone_MatchScore	Generates a match score based on person names and telephone numbers.

Spain Demonstration mappings

The demonstration mappings in the Spain accelerator use multiple rules to demonstrate data quality processes.

Find the demonstration mappings in the following repository location:

    [Informatica_DQ_Content]\Rules_Demo\ESP_Accelerator

The accelerator contains the following demonstration mappings:

m_ESP_customer_data_demo

Parses, standardizes, and validates Spanish data.

m_ESP_customer_matching_demo

Parses and standardizes identity data from Spain and performs identity match analysis on the data.

The mapping analyzes the following data combinations and generates match clusters for each combination:

- Person name and address data
- Person name and telephone number

CHAPTER 14

United Kingdom Accelerator

This chapter includes the following topics:

United Kingdom Accelerator Overview

Use the rules in the United Kingdom accelerator to verify and enhance data from organizations in the United Kingdom.

The United Kingdom accelerator includes rules that perform the following data quality processes:

- Address data cleansing
- Contact data cleansing
- Financial data cleansing
- Matching and deduplication

The accelerator depends on data cleansing rules that the Core accelerator installs.

United Kingdom Address Data Cleansing Rules

Use the address data cleansing rules to parse, standardize, and validate address data.

Find the address data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
```

The following table describes the address data cleansing rules in the United Kingdom accelerator:

Name	Description
rule_GBR_Address_Parse_Hybrid	Parses unstructured United Kingdom addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_GBR_Address_Parse_Multiline	Parses unstructured United Kingdom addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_GBR_Address_Validation_Discrete_w_Geocoding	Validates the deliverability of United Kingdom addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_GBR_Address_Validation_Discrete	Validates the deliverability of United Kingdom addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_GBR_Address_Validation_Hybrid_w_Geocoding	Validates the deliverability of United Kingdom addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_GBR_Address_Validation_Hybrid	Validates the deliverability of United Kingdom addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_GBR_Address_Validation_Multiline_w_Geocoding	Validates the deliverability of United Kingdom addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_GBR_Address_Validation_Multiline	Validates the deliverability of United Kingdom addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_GBR_Postcode_Parse	Parses United Kingdom postal codes.

Name	Description
rule_GBR_Postcode_Standardize	Standardizes United Kingdom postal codes. The rule requires that the input follows predefined formats. The rule standardizes inputs that match the following patterns: - A9 9AA - A99 9AA - AA9 9AA - AA99 9AA - A9A 9AA - AA9A 9AA - GIR 0AA The letter A represents an alphabetic character and the number 9 represents a digit.
rule_GBR_Postcode_Validate	Validates United Kingdom postal codes. The rule matches standardized postal codes with valid United Kingdom postal codes. If the rule does not find a matching postal code, it verifies whether the postal code follows the standard United Kingdom postal code pattern.

United Kingdom Composite Rules

Use the composite rules in the United Kingdom accelerator to add a set of rules to a mapping as a single object. A composite rule is a rule that makes use of the logic of other accelerator rules.

Find the composite rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Composite_Rules

United Kingdom Composite Rule for Contact Data

The following table describes the composite rule for contact data in the United Kingdom accelerator:

Name	Description
rule_GBR_Contact_Data	Parses, standardizes, and validates United Kingdom contact data, such as addresses, telephone numbers, and National Insurance Numbers (NINO).

The rule `rule_GBR_Contact_Data` reads mapplets from multiple folders in the repository. The rule also includes a nonreusable transformation.

The following table lists the names and the repository locations of the rules and the transformation in `rule_GBR_Contact_Data`:

Name	Location
Case_Converter	Nonreusable transformation
rule_Assign_DQ_Mailability_Score_Description	[Informatica_DQ_Content]\Rules\General_Data_Cleansing
rule_Email_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

Name	Location
rule_GBR_Address_Validation_Hybrid	[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
rule_GBR_Company_Name_Standardization	[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing
rule_GBR_Gender_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_GBR_Multi_Person_Name_Parse	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_GBR_NINO_Standardization	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_GBR_NINO_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_GBR_Phone_Number_Standardization	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_GBR_Phone_Number_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_Prename_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_Salutation_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

United Kingdom Contact Data Cleansing Rules

Use the contact data cleansing rules to parse, standardize, and validate data about business contacts and individuals.

Find the contact data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

The following table describes the contact data cleansing rules in the United Kingdom accelerator:

Name	Description
rule_GBR_Driver_Number_Parse	Parses strings that match the format of United Kingdom driver's license numbers.
rule_GBR_Driver_Number_Validation	Validates United Kingdom driver's license numbers based on the requirements of the United Kingdom Government Data Standards Catalog.
rule_GBR_Gender_Assignment	Assigns gender according to first names. The rule returns "M" for male names, "F" for female names, and "U" if the gender is unknown. For example, the rule assigns the name "John Smith" a gender of "M" for male.
rule_GBR_Given_Name_Standard	Generate given names from United Kingdom nicknames.

Name	Description
rule_GBR_Multi_Person_Name_Parse	Parses person name values into separate ports. The rule creates ports for values such as title, first name, middle name, and surname.
	The rule output includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping.
	When the name data identifies more than one person, the rule creates an output port for each full name. For example, the rule can read the name "John and Jane Smith" and create output ports for "John Smith" and "Jane Smith."
rule_GBR_NHS_Number_Parse	Parses National Health Service (NHS) numbers from a string. The rule returns the NHS number and also returns a string that contains the input text with the NHS number removed.
rule_GBR_NHS_Number_Standardize	Standardizes National Health Service (NHS) numbers into the standard format (999 999 9999). The rule requires that the input is a 10-digit string.
rule_GBR_NHS_Number_Validate	Validates National Health Service (NHS) numbers based on the check digit in each number. The rule requires that the input is a 10-digit string.
rule_GBR_NINO_Conformity_Check	Validates the standard pattern for a United Kingdom National Insurance Number (NINO). The rule does not verify that a NINO is accurate or active.
rule_GBR_NINO_Parse	Parses United Kingdom National Insurance Numbers (NINO) from strings. The rule returns the NINO and also returns a string that contains the input text with the NINO removed.
rule_GBR_NINO_Standardization	Standardizes United Kingdom National Insurance Numbers (NINO) into the two most typical formats. The rule returns the following formats, where C represents alphabetic characters and N represents numerals: - CC NN NN NN C - CCNNNNNNC The rule formats all alphabetic characters as uppercase. The rule requires that the input conforms to the pattern of a NINO.
rule_GBR_NINO_Validation	Validates a United Kingdom National Insurance Number (NINO). The rule does not verify that a NINO is active.
rule_GBR_Passport_Number_MR_Parse	Parses United Kingdom passport numbers in extended format. The extended format is the machine readable format for passport numbers.
rule_GBR_Passport_Number_Parse	Parses United Kingdom passport numbers that use the format specified by the Government Data Standards Catalogue. The rule parses all nine-digit strings.
rule_GBR_Passport_Number_Validation	Validates United Kingdom passport numbers that use the format specified by the Government Data Standards Catalogue.
rule_GBR_Personal_Name_Parsing_FML	Parses the values in a person name into separate ports. The rule creates the ports in the following sequence: - First name, middle name, last name The rule output also includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping.

Name	Description
rule_GBR_Personal_Name_Parsing_LFM	Parses the values in a person name into separate ports. The rule creates the ports in the following sequence: - Last name, first name, middle name The rule output also includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping.
rule_GBR_Phone_Number_Parse	Parses a United Kingdom telephone number from a string. The rule parses the first telephone number in the data, reading from right to left. The rule recognizes telephone numbers that use leading zeros, the "+44" international dialing code, and extensions that begin with the hash symbol. The rule processes the following punctuation symbols: the plus sign, parentheses, and the hash symbol. Before you run the rule, remove all other punctuation, including double spaces. The rule returns a telephone number and also returns a string that contains the input text with the telephone number removed.
rule_GBR_Phone_Number_Standardization	Standardizes United Kingdom telephone numbers to international and local dialing formats. The rule recognizes telephone numbers that use leading zeros, the "+44" international dialing code, and extensions that begin with the hash symbol.
rule_GBR_Phone_Number_Validation	Validates the area code and length of United Kingdom telephone numbers. The rule returns the region of the telephone number as well as codes that indicate if the area code and length of a telephone number are valid.
rule_Prename_Assignment	Generates an honorific according to the gender. You can change the female_prename expression variable from Ms. to Mrs.
rule_Salutation_Assignment	Generates formal and casual greetings from prenames and name tokens. For example, when input data contains "Mr. John Smith," the rule generates the formal greeting "Dear Mr. Smith," and the casual greeting "Dear John,". You can change the prefix and punctuation by editing the variables in the dq_Generate_Salutation Expression transformation.

United Kingdom Corporate Data Cleansing Rules

Use the corporate data cleansing rules to standardize corporate data.

Find the corporate data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing
```

The following table describes the corporate data cleansing rules from the United Kingdom accelerator:

Name	Description
rule_GBR_Company_Name_Standardization	Standardizes a company name and provides the acronym for the name if it is possible to do so.

United Kingdom Financial Data Cleansing Rules

Use the financial data cleansing rules to parse, standardize, and validate financial data.

Find the financial data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Financial_Data_Cleansing

The following table describes the financial data cleansing rules in the United Kingdom accelerator:

Name	Description
rule_GBR_Bank_Account_Parse	Parses eight-digit numeric strings as United Kingdom bank account numbers.
rule_GBR_Bank_Account_Validation	Validates United Kingdom bank account numbers. The rule returns codes that indicate whether the input is numeric and whether it is the correct number of digits.
rule_GBR_Bank_Sort_Code_Parse	Parses six-digit numeric strings as United Kingdom bank sort codes. The rule parses strings of numbers in the following formats: - Consecutive numbers (999999) - Numbers delimited with a dash (99-99-99)
rule_GBR_Bank_Sort_Code_Validation	Validates the format and length of United Kingdom bank sort codes that are standardized to the dash-delimited format (99-99-99). The rule returns a Status port that describes the validity of the sort code and a Validation Note port that explains the status. If the sort code prefix matches a known assignment for a United Kingdom bank, the Validation Note port includes the bank name.
rule_GBR_Bank_Sort_Code_Standardize	Standardizes a United Kingdom bank sort code to the format "NN-NN-NN."

United Kingdom General Data Cleansing Rules

Use the general data cleansing rules to identify the type of information contained in input fields.

Find the general data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\General_Data_Cleansing

The following table describes the general data cleansing rules in the United Kingdom accelerator:

Name	Description
rule_GBR_NER_Field_Identification	Identifies the type of information contained in an input field. The rule can identify names, Personal IDs, company names, dates, and United Kingdom address data. The rule returns a label that describes the type of input data. The rule uses reference data to identify the types of information. The rule uses probabilistic matching techniques to identify the types of information.

The United Kingdom accelerator depends on the following general data cleansing rules from the Core accelerator:

- rule_Assign_DQ_GeocodingStatus_Description
- rule_Assign_DQ_Mailability_Score_Description

- rule_Assign_DQ_Match_Code_Descriptions
- rule_Remove_Extra_Spaces
- rule_Remove_Leading_Zero
- rule_Remove_Period_Parentheses
- rule_Remove_Punctuation
- rule_Remove_Punctuation_and_Space
- rule_Remove_Space
- rule_Replace_Limited_Punct_with_Space
- rule_UpperCase

For more information about these rules, see "Core General Data Cleansing Rules" on page 22.

United Kingdom Matching and Deduplication Rules

Use the matching and deduplication rules to measure the levels of similarity between the records in data sets.

Find the matching and deduplication rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Matching_Deduplication

The following table describes the matching and deduplication rules in the United Kingdom accelerator:

Name	Description
mplt_GBR_Company_Name_Postcode_Match	Uses field match strategies to identify duplicate rows in United Kingdom data based on company name and postal code. The mapplet generates group keys from the postal code.
mplt_GBR_Familyname_and_NINO_Match	Uses field match strategies to identify duplicate rows in United Kingdom data based on surname and National Insurance Number (NINO). The mapplet generates group keys from the NINO data.
mplt_GBR_Familyname_and_Postcode_Match	Uses field match strategies to identify duplicate rows in United Kingdom data based on surname and United Kingdom postal code. The mapplet generates group keys from the postal code data.
mplt_GBR_Firstname_3CharsSurname_DOB_and_Postcode_Match	Uses field match strategies to identify duplicate rows in United Kingdom data based on the following data: - First name - The first three characters in the surname - Date of birth - postal code The mapplet generates group keys from the postal code data.
mplt_GBR_Firstname_Surname_2ElementsDOB_and_Postcode_Match	Uses field match strategies to identify duplicate rows in United Kingdom data based on the following data: - Person names - Any two date of birth elements, such as month and year - United Kingdom postal code The mapplet generates group keys from the postal code data.

Name	Description
mplt_GBR_Firstname_Surname_DOB_and_Postcode_Match	Uses field match strategies to identify rows based on the following data: - Person names - Date of birth - postal code The mapplet generates group keys from the postal code data.
mplt_GBR_IMO_Company_Name_and_Address_Match	Uses identity match strategies to identify duplicate rows in United Kingdom data based on company names and addresses. The mapplet generates group keys from the postal code data.
mplt_GBR_IMO_Familyname_and_Address_Match	Uses identity match strategies to identify duplicate rows in United Kingdom data based on family names and addresses. The mapplet generates group keys from the postal code data.
mplt_GBR_IMO_Individual_Name_and_Address_Match	Uses identity match strategies to identify duplicate rows in United Kingdom data based on person names and addresses. The mapplet generates group keys from the postal code data.
mplt_GBR_IMO_Personal_Name_and_Data_Match	Uses identity match strategies to identify duplicate rows in United Kingdom data based on person names and personal data. The fields in the personal data column must contain a single type of data, such as telephone number, email, or National Insurance Number. The mapplet generates group keys from the personal data.
mplt_GBR_Individual_Name_and_Date_Match	Uses field match strategies to identify duplicate rows in United Kingdom data based on person names and date data. The mapplet generates group keys from the date data.
mplt_GBR_Individual_Name_and_Email_Match	Uses field match strategies to identify duplicate rows in United Kingdom data based on person names and the email address data. The mapplet generates group keys from the email address data.
mplt_GBR_Individual_Name_and_NINO_Match	Uses field match strategies to identify duplicate rows in United Kingdom data based on person names and National Insurance Numbers (NINO). The mapplet generats group keys from the NINO data.
mplt_GBR_Individual_Name_and_Phone_Match	Uses field match strategies to identify duplicate rows in United Kingdom data based on person names and telephone numbers. The mapplet generates group keys from the telephone number data.
mplt_GBR_Individual_Name_and_Postcode_Match	Uses field match strategies to identify duplicate rows in United Kingdom data based on person names and the postal code data. The mapplet generates group keys from the postal code data.
mplt_GBR_Individual_Name_Match	Uses field match strategies to identify duplicate rows in United Kingdom data based on person names. The mapplet generates NYSIIS codes from the surname values and uses the NYSIIS codes as group keys.
rule_GBR_Familyname_and_NINO_MatchScore	Generates a match score based on surnames and United Kingdom National Identification Numbers (NINO).
rule_GBR_Familyname_and_Postcode_MatchScore	Generates a match score based on surnames and United Kingdom postal codes.

Name	Description
rule_GBR_Firstname_3CharsSurname_DOB_and_Postcode_MatchScore	Generates a match score based on the following information: - First name - The first three characters in the surname - Date of birth - Postal code
rule_GBR_Firstname_Surname_2ElementsDOB_and_Postcode_MatchScore	Generates a match score based on the following information: - Person names - Any two date of birth elements, such as month and year - United Kingdom postal code
rule_GBR_Firstname_Surname_DOB_and_Postcode_MatchScore	Generates a match score based on person names, date of birth, and postal code.
rule_GBR_Individual_Name_and_NINO_MatchScore	Generates a match score based on person names and United Kingdom National Insurance Numbers (NINO).
rule_GBR_Individual_Name_and_Phone_MatchScore	Generates a match score based on person names and telephone numbers.
rule_GBR_Individual_Name_and_Postcode_MatchScore	Generates a match score based on person names and United Kingdom postal codes.
rule_Individual_Name_and_Date_MatchScore	Generates a match score based on person names and dates.
rule_Individual_Name_MatchScore	Generates a match score based on person names.
rule_GBR_Company_Name_Postcode_MatchScore	Generates a match score based on company name and United Kingdom postal codes.
rule_Individual_Name_and_Email_MatchScore	Generates a match score based on person names and email addresses.

United Kingdom Demonstration Mappings

The demonstration mappings in the United Kingdom Accelerator use multiple rules to demonstrate data quality processes.

Find the demonstration mappings in the following repository location:

    [Informatica_DQ_Content]\Rules_Demo\GBR_Accelerator

The United Kingdom accelerator contains the following demonstration mappings:

m_GBR_customer_data_demo

Parses, standardizes, and validates United Kingdom customer data.

m_GBR_customer_matching_demo

Parses and standardizes identity data from the United Kingdom and performs identity match analysis on the data.

The mapping analyzes the following data combinations and generates match clusters for each combination:

- Person name and address data

- Person name and telephone number

CHAPTER 15

U.S./Canada Accelerator

This chapter includes the following topics:

U.S./Canada Accelerator Overview

Use the rules in the U.S./Canada accelerator to verify and enhance data from organizations in the United States and Canada.

The U.S./Canada accelerator includes rules that perform the following data quality processes:

- Address data cleansing
- Contact data cleansing
- General data cleansing
- Matching and deduplication

The accelerator depends on data cleansing rules that the Core accelerator installs.

U.S./Canada Address Data Cleansing Rules

Use address data cleansing rules to parse, standardize, and validate address data.

The address data cleansing rules in the following repository location:

```
[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
```

The following table describes the address data cleansing rules in the U.S./Canada accelerator:

Name	Description
rule_CAN_Address_Certification_Hybrid	Validates the deliverability of Canadian addresses to the Software Evaluation and Recognition Program (SERP) standards that Canada Post maintains for Canadian addresses. The rule corrects errors in the input addresses where possible and includes a status port that identifies the addresses that meet the SERP standard. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation. **Note:** When you certify a set of address records to the SERP standard, you must submit a certification report to Canada Post. The Address Validator transformation includes property fields that you can populate with information for the report. Save or print the report, and include it with the address details that you submit to Canada Post.
rule_CAN_Address_Parse_Hybrid	Parses unstructured Canadian addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_CAN_Address_Parse_Multiline	Parses unstructured Canadian addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_CAN_Address_Validation_Discrete_w_Geocoding	Validates the deliverability of Canadian addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_CAN_Address_Validation_Discrete	Validates the deliverability of Canadian addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_CAN_Address_Validation_Hybrid_w_Geocoding	Validates the deliverability of Canadian addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_CAN_Address_Validation_Hybrid	Validates the deliverability of Canadian addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_CAN_Address_Validation_Multiline_w_Geocoding	Validates the deliverability of Canadian addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_CAN_Address_Validation_Multiline	Validates the deliverability of Canadian addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_CAN_Postcode_Validation	Validates Canadian postal codes. The rule returns "Valid" or "Invalid."
rule_CAN_Province_Validation	Validates Canadian province names. The rule returns "Valid" or "Invalid."

Name	Description
rule_USA_Address_Certification_Hybrid	Validates the deliverability of United States addresses to the Coding Accuracy Support System (CASS) standards that the United States Postal Service maintains for the addresses. The rule corrects errors in the input addresses where possible and includes a status port that identifies the addresses that meet the CASS standard. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation. **Note:** When you certify a set of address records to the CASS standard, you must submit a certification report to the USPS. The Address Validator transformation includes property fields that you can populate with information for the report. Save or print the report, and include it with the address details that you submit to the USPS.
rule_USA_Address_Parse_Hybrid	Parses unstructured United States addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_USA_Address_Parse_Multiline	Parses unstructured United States addresses into address elements. The rule does not validate the addresses. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_USA_Address_Validation_Discrete_w_Geocoding	Validates the deliverability of United States addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_USA_Address_Validation_Discrete	Validates the deliverability of United States addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Discrete input ports on the Address Validator transformation.
rule_USA_Address_Validation_Hybrid_w_Geocoding	Validates the deliverability of address records from United States addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_USA_Address_Validation_Hybrid	Validates the deliverability of address records from United States addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Hybrid input ports on the Address Validator transformation.
rule_USA_Address_Validation_Multiline_w_Geocoding	Validates the deliverability of United States. addresses and adds latitude and longitude coordinates to each output address. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_USA_Address_Validation_Multiline	Validates the deliverability of U.S. addresses. The rule corrects errors in the input addresses where possible. Use the rule when you can connect the input address fields to the Multiline input ports on the Address Validator transformation.
rule_USA_County_Validation	Validates United States county names. The rule compares input data against county names in all states. The rule returns "Valid" or "Invalid."

Name	Description
rule_USA_State_Validation	Validates United States state names. The rule returns "Valid" or "Invalid."
rule_USA_ZIPCode_Validation	Validates five-digit United States Zone Improvement Plan (ZIP) Codes. The rule returns "Valid" or "Invalid."

U.S./Canada Composite Rules

Use the composite rules in the U.S./Canada accelerator to add a set of rules to a mapping as a single object. A composite rule is a rule that makes use of the logic of other accelerator rules.

Find the composite rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Composite_Rules

The following table describes the composite rules in the U.S./Canada accelerator:

Name	Description
rule_CAN_Contact_Data	Parses, standardizes, and validates Canada contact data, such as addresses, telephone numbers, and Social Insurance Numbers (SIN).
rule_USA_Contact_Data	Parses, standardizes, and validates United States contact data, such as addresses, telephone numbers, and Social Security Numbers (SSN).

Composite Rule for Canada Contact Data

The rule `rule_CAN_Contact_Data` reads mapplets from multiple folders in the repository. The rule also includes a nonreusable transformation.

The following table lists the names and the repository locations of the rules and the transformation in `rule_CAN_Contact_Data`:

Rule	Location
Case_Converter	Nonreusable transformation
rule_Assign_DQ_Mailability_Score_Description	Informatica_DQ_Content]\Rules\General_Data_Cleansing
rule_CAN_Address_Validation_Hybrid	[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
rule_CAN_Gender_Assignment	Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_CAN_Multi_Person_Name_Parse	Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_CAN_Phone_Number_Standardization	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_CAN_Phone_Number_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_CAN_SIN_Standardization	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

Rule	Location
rule_CAN_SIN_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_Company_Name_Standardization	[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing
rule_Email_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_Prename_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_Salutation_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

Composite Rule for United States Contact Data

The rule `rule_USA_Contact_Data` reads mapplets from multiple folders in the repository. The rule also includes a nonreusable transformation.

The following table lists the names and the repository locations of the rules and the transformation in `rule_USA_Contact_Data`:

Rule	Location
Case_Converter	Nonreusable tranformation
rule_Assign_DQ_Mailability_Score_Description	[Informatica_DQ_Content]\Rules\General_Data_Cleansing
rule_Company_Name_Standardization	[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing
rule_Email_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_Prename_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_Salutation_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_USA_Address_Validation_Hybrid	[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
rule_USA_Gender_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_USA_Multi_Person_Name_Parse	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_USA_Phone_Number_Standardization	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_USA_Phone_Number_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_USA_SSN_Standardization	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_USA_SSN_Validation_post_June2011	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

U.S./Canada Contact Data Cleansing Rules

Use contact data cleansing rules to parse, standardize, and validate data about business contacts and individuals.

Find the contact data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

The following table describes the contact data cleansing rules in the U.S./Canada accelerator:

Name	Description
rule_CAN_Gender_Assignment	Assigns gender according to first names. The rule returns "M" for male names, "F" for female names, and "U" if the gender is unknown. For example, the rule assigns the name "John Smith" a gender of "M" for male.
rule_CAN_Given_Name_Standard	Generate given names from Canadian nicknames. For example, the rule standardizes the nickname "Bob" to the given name "Robert."
rule_CAN_Multi_Person_Name_Parse	Parses person name values into separate ports. The rule creates ports for values such as title, first name, middle name, and surname. The rule output includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping. When the name data identifies more than one person, the rule creates an output port for each full name. For example, the rule can read the name "John and Jane Smith" and create output ports for "John Smith" and "Jane Smith."
rule_CAN_Personal_Name_Parse_and_Standardize_FML	Parses the values in a person name into separate ports. The rule also standardizes the name values. The rule creates the ports in the following sequence: - First name, middle name, last name The rule output also includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping.
rule_CAN_Personal_Name_Parse_and_Standardize_LFM	Parses the values in a person name into separate ports. The rule also standardizes the name values. The rule creates the ports in the following sequence: - Last name, first name, middle name The rule output also includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping.
rule_CAN_Personal_Name_Parsing_FML	Parses the values in a person name into separate ports. The rule creates the ports in the following sequence: - Last name, first name, middle name The rule output also includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping. **Note:** The rule does not standardize the name values. To standardize and parse Canadian name values in the sequence that the rule defines, select rule_CAN_Personal_Name_Parse_and_Standardize_FML.

Name	Description
rule_CAN_Personal_Name_Parsing_LFM	Parses the values in a person name into separate ports. The rule creates the ports in the following sequence: - Last name, first name, middle name The rule output also includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping. **Note:** The rule does not standardize the name values. To standardize and parse Canadian name values in the sequence that the rule defines, select ule_CAN_Personal_Name_Parse_and_Standardize_LFM.
rule_CAN_Phone_Number_Parse	Parses a Canadian telephone number from a string. The rule parses the first telephone number in the data, reading from right to left. The rule returns a telephone number and also returns a string that contains the input text with the telephone number removed.
rule_CAN_Phone_Number_Standardization	Standardizes Canadian telephone numbers. The rule returns the telephone number in the following formats: - Standard - (nnn) nnn-nnnn - Dashes - nnn-nnn-nnnn - No Spaces - nnnnnnnnnn
rule_CAN_Phone_Number_Validation	Validates the area code and length of Canadian telephone numbers. The rule returns codes that indicate telephone number type and validity. Types describe categories such as "toll-free."
rule_CAN_Phone_Parse_Standardize_Validate	Parse a telephone number from a string of text and verifies that the area code is valid for Canada. If the area code is valid, the rule returns the telephone number in three standard formats. The rule also returns a status value to indicate whether the data conforms to the standard format for a Canadian telephone number.
rule_CAN_Phone_w_Extension_Parse	Parses a number from a string of text if the number conforms to the standard format for a Canadian telephone number. The rule includes any telephone extension data when it parses the telephone number.
rule_CAN_SIN_Parse	Parses a Canadian Social Insurance Number (SIN) from a string. The rule returns the SIN and also returns a string that contains the input text with the SIN removed.
rule_CAN_SIN_Standardization	Standardizes Canadian Social Insurance Numbers (SIN). The rule can output the following formats: - No Punctuation - nnnnnnnnn - Space - nnn nnn nnn - Dash - nnn-nnn-nnn To change the format, edit the SIN_Format expression variable in the dq_Format_SIN Expression transformation. Default is "No_Punctuation."
rule_CAN_SIN_Validation	Validates Canadian Social Insurance Numbers (SIN). The rule uses the Luhn algorithm to verify whether or not a SIN is valid. The rule returns "Valid" or "Invalid."
rule_Prename_Assignment	Generates an honorific according to the gender. You can change the female_prename expression variable from Ms. to Mrs.

Name	Description
rule_Salutation_Assignment	Generates formal and casual greetings from prenames and name tokens. For example, when input data contains "Mr. John Smith," the rule generates the formal greeting "Dear Mr. Smith," and the casual greeting "Dear John,". You can change the prefix and punctuation by editing the variables in the dq_Generate_Salutation Expression transformation.
rule_USA_Gender_Assignment	Assigns gender according to first name. The rule returns "M" for male names, "F" for female names, and "U" if the gender is unknown. For example, the rule assigns the name "John Smith" a gender of "M" for male.
rule_USA_Given_Name_Standard	Generate given names from U.S. nicknames. For example, the rule standardizes the nickname "Bob" to the given name "Robert."
rule_USA_Multi_Person_Name_Parse	Parses person name values into separate ports. The rule creates ports for values such as title, first name, middle name, and surname. The rule output includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping. When the name data identifies more than one person, the rule creates an output port for each full name. For example, the rule can read the name "John and Jane Smith" and create output ports for "John Smith" and "Jane Smith."
rule_USA_Personal_Name_Parse_and_Standardize_FML	Parses the values in a person name into separate ports. The rule also standardizes the name values. The rule creates the ports in the following sequence: - First name, middle name, last name The rule output also includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping.
rule_USA_Personal_Name_Parse_and_Standardize_LFM	Parses the values in a person name into separate ports. The rule also standardizes the name values. The rule creates the ports in the following sequence: - Last name, first name, middle name The rule output also includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping.
rule_USA_Personal_Name_Parse_Validation	Validates the gender assignment for a name. The rule calculates the probabilities that a data value is a male name or a female name. If the gender is unknown, the rule uses the probability calculations to assign a gender to the name.
rule_USA_Personal_Name_Parsing_FML	Parses the values in a person name into separate ports. The rule creates the ports in the following sequence: - First name, middle name, last name The rule output also includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping. **Note:** The rule does not standardize the name values. To standardize and parse United States name values in the sequence that the rule defines, select rule_USA_Personal_Name_Parse_and_Standardize_FML .

Name	Description
rule_USA_Personal_Name_Parsing_LFM	Parses the values in a person name into separate ports.
	The rule creates the ports in the following sequence:
	- Last name, first name, middle name
	The rule output also includes a port that contains the full name of the person in the record. You can use the full name port as an input to a Match transformation in an identity match analysis mapping.
	Note: The rule does not standardize the name values. To standardize and parse United States name values in the sequence that the rule defines, select rule_USA_Personal_Name_Parse_and_Standardize_LFM.
rule_USA_Phone_Number_Parse	Parses a United States telephone number from a string. The rule parses the first telephone number in the data, reading from right to left. The rule returns a telephone number and also returns a string that contains the input text with the telephone number removed.
rule_USA_Phone_Number_Standardization	Standardizes United States telephone numbers. The rule returns the telephone number in the following formats:
	- Standard - (nnn) nnn-nnnn
	- Dashes - nnn-nnn-nnnn
	- No Spaces - nnnnnnnnnn
rule_USA_Phone_Number_Validation	Validates the area code and length of United States telephone numbers. The rule returns codes that indicate if the area code and length of a telephone number are valid.
rule_USA_Phone_Parse_Standardize_Validate	Parse a telephone number from a string of text and verifies that the area code is valid for the United States.
	If the area code is valid, the rule returns the telephone number in three standard formats. The rule also returns a status value to indicate whether the data conforms to the standard format for a United States telephone number.
rule_USA_Phone_w_Extension_Parse	Parses a number from a string of text if the number conforms to the standard format for a United States telephone number. The rule includes any telephone extension data when it parses the telephone number.
rule_USA_SSN_Parse	Parses United States Social Security Numbers (SSN).
rule_USA_SSN_Parse_Standardize_and_Validate	Parses, standardizes, and validates United States Social Security Numbers from a larger string of text. The rule can parse numbers that include or omit dashes.
	By default, the rule writes Social Security Numbers without any punctuation. To change the standardization format, open the dq_SSN_Format transformation in the rule and update the expression on the SSN_Format port.
rule_USA_SSN_Standardization	Standardizes United States Social Security Numbers (SSN). The rule can output the following formats:
	- No Punctuation - nnnnnnnnn
	- Space - nnn nnn nnn
	- Dash - nnn-nnn-nnn
	To change the format, edit the SSN_format expression variable in the dq_SSN_Format Expression transformation. Default is "No_Punctuation."

Name	Description
rule_USA_SSN_Validation	Validates United States Social Security Numbers (SSN). The rule validates each SSN for length, numeric values, and known minimum and maximum values in the Area, Group, and Serial Number sections.
	The Area section comprises the first three digits of the SSN, and the Group section comprises the fourth and fifth digits. The Serial Number section comprises the final four digits.
	If the SSN was issued prior to June 2011, the rule also verifies that the Area value and Group value are a valid combination. The rule does not verify that the SSN is an issued number. The rule returns "Valid" or "Invalid."
rule_USA_SSN_Validation_post_June2011	Validates United States Social Security Numbers (SSN). The rule validates each SSN for length, numeric values, and known minimum and maximum values in the Area, Group, and Serial Number sections.
	The Area section comprises the first three digits of the SSN, and the Group section comprises the fourth and fifth digits. The Serial Number section comprises the final four digits.
	The rule does not verify that the Area value and Group value are a valid combination. The rule does not verify that the SSN is an issued number. The rule returns "Valid" or "Invalid."

Dependencies on Core Contact Data Cleansing Rules

The U.S./Canada accelerator depends on the following contact data cleansing rules from the Core accelerator:

- rule_Email_Validation

For more information about these rules, see "Core Contact Data Cleansing Rules" on page 21.

U.S./Canada Corporate Data Cleansing Rules

Use the corporate data cleansing rules in the U.S./Canada accelerator to parse, standardize, and validate corporate data.

The following table describes the corporate data cleansing rules in the U.S./Canada accelerator:

Name	Description
rule_NAICS_Code_Validation	Validates North American Industry Classification System (NAICS) codes.
rule_USA_SIC_Code_Validation	Validates Standard Industrial Classification (SIC) codes.

The U.S./Canada accelerator depends on the following corporate data cleansing rule from the Core accelerator:

- rule_Company_Name_Standardization

U.S./Canada General Data Cleansing Rules

Use the general data cleansing rules to identify the type of information contained in input fields.

Find the general data cleansing rules in the following repository location:

    [Informatica_DQ_Content]\Rules\General_Data_Cleansing

The following table describes the general data cleansing rules in the U.S./Canada accelerator:

Name	Description
rule_CAN_Field_Identification	Identifies the type of information that an input field contains. The rule can identify names, personal IDs, company names, dates, and Canadian address data. The rule returns a label that describes the type of input data. The rule uses reference data to identify the types of information.
rule_CAN_NER_Field_Identification	Identifies the type of information that an input field contains. The rule can identify names, personal IDs, company names, dates, and Canadian address data. The rule returns a label that describes the type of input data. The rule uses reference data to identify the types of information. The rule uses probabilistic matching techniques to identify the types of information.
rule_USA_Field_Identification	Identifies the type of information that an input field contains. The rule can identify names, personal IDs, company names, dates, and United States address data. The rule returns a label that describes the type of input data. The rule uses reference data to identify the types of information.
rule_Field_North_American_Data	Identifies the following types of fields: name, occupation title, company, address, city, state or province, postcode, country, personal ID, email, telephone, credit card, and date. The rule generates a score that indicates the degree of confidence in the field identification. Higher scores indicate greater levels of confidence. If the rule cannot assign a field type, the rule writes the data on the Out_Undetermined port.
rule_USA_NER_Field_Identification	Identifies the type of information that an input field contains. The rule can identify names, personal IDs, company names, dates, and United States address data. The rule returns a label that describes the type of input data. The rule uses reference data to identify the types of information. The rule uses probabilistic matching techniques to identify the types of information.

Dependencies on Core General Data Cleansing Rules

The U.S./Canada accelerator depends on the following general data cleansing rules from the Core accelerator:

- rule_Assign_DQ_GeocodinStatus_Description
- rule_Assign_DQ_Mailability_Score_Description
- rule_Assign_DQ_Match_Code_Descriptions
- rule_Date_Validation
- rule_Remove_Extra_Spaces
- rule_Remove_Punctuation
- rule_Replace_Limited_Punct_with_Space
- rule_UpperCase

For more information about these rules, see "Core General Data Cleansing Rules" on page 22.

U.S./Canada Matching and Deduplication Rules

Use the matching and deduplication rules to measure the levels of similarity between the records in data sets.

Find the matching and deduplication rules in the following repository location:

    [Informatica_DQ_Content]\Rules\Matching_Deduplication

The following table describes the matching and deduplication rules in the U.S./Canada accelerator:

Name	Description
mplt_CAN_IMO_Company_Name_and_Address_Match	Uses identity match strategies to identify duplicate rows in Canadian data based on company names and addresses. The mapplet generates group keys from the postal code data.
mplt_CAN_IMO_Familyname_and_Address_Match	Uses identity match strategies to identify duplicate rows in Canadian data based on family names and addresses. The mapplet generates group keys from the postal code data.
mplt_CAN_IMO_Individual_Name_and_Address_Match	Uses identity match strategies to identify duplicate rows in Canadian data based on person names and addresses. The mapplet generates group keys from the postal code data.
mplt_CAN_IMO_Personal_Name_and_Data_Match	Uses identity match strategies to identify duplicate rows in Canadian data based on person names and personal data. The fields in the personal data column should contain a single type of data, such as telephone number, email, or Social Insurance Number. The mapplet generates group keys from the personal data.
mplt_Company_Name_and_Address_Match	Uses field match strategies to identify duplicate rows based on company name and address data. The mapplet uses a combination of characters from the company name values and the postal code values to generate group keys.
mplt_Company_Name_Match	Uses field match strategies to identify duplicate rows based on company name. The mapplet generates Soundex codes from the company name values and uses the Soundex codes as group keys.
mplt_Familyname_and_Address_Match	Uses field match strategies to identify duplicate rows based on surname and address data. The mapplet uses a combination of characters from the surname values and the postal code values to generate group keys.
mplt_Firstname_and_SSN_Match	Uses field match strategies to identify duplicate rows based on first names and United States Social Security numbers. The mapplet generates group keys from the Social Security number data.
mplt_Individual_Name_and_Address_Match	Uses field match strategies to identify duplicate rows based on person names and United States address data. The mapplet uses a combination of characters from the surname values and the postal code values to generate group keys.
mplt_Individual_Name_and_Date_Match	Uses field match strategies to identify duplicate rows based on person names and date data. The mapplet generates group keys from the date data.
mplt_Individual_Name_and_Email_Match	Uses field match strategies to identify duplicate rows based on person names and email addresses. The mapplet generates group keys from the email address data.
mplt_Individual_Name_and_Phone_Match	Uses field match strategies to identify duplicate rows based on person names and telephone numbers. The mapplet generates group keys from the telephone number data.

Name	Description
mplt_Individual_Name_and_SSN_Match	Uses field match strategies to identify duplicate rows based on person names and United States Social Security numbers. The mapplet generates keys generated from the Social Security number data.
mplt_Individual_Name_Match	Uses field match strategies to identify duplicate rows based on person names. The mapplet generates NYSIIS codes from the surname values and uses the NYSIIS codes as group keys.
mplt_USA_Address_Match	Uses field match strategies to identify duplicate rows in United States data based on United States address data. The mapplet generates group keys from the postal code data.
mplt_USA_IMO_Company_Name_and_Address_Match	Uses identity match strategies to identify duplicate rows in United States data based on company names and addresses. The mapplet generates group keys from the postal code data.
mplt_USA_IMO_Familyname_and_Address_Match	Uses identity match strategies to identify duplicate rows in United States data based on family names and addresses. The mapplet generates group keys from the postal code data.
mplt_USA_IMO_Individual_Name_and_Address_Match	Uses identity match strategies to identify duplicate rows in United States data based on person names and addresses. The mapplet generates group keys from the postal code data.
mplt_USA_IMO_Personal_Name_and_Data_Match	Uses identity match strategies to identify duplicate rows in United States data based on person names and personal data. The fields in the personal data column must contain a single type of data, such as telephone number, email, or Social Security number. The mapplet generates group keys from the personal data.
rule_Company_Name_and_Address_MatchScore	Generates a match score based on company names and United States address data.
rule_Company_Name_MatchScore	Generates a match score based on company names.
rule_Familyname_and_Address_MatchScore	Generates a match score based on surnames and United States address data.
rule_Firstname_and_SSN_MatchScore	Generates a match score based on first names and United States address data.
rule_Individual_Name_and_Address_MatchScore	Generates a match score based on person names and United States address data.
rule_Individual_Name_and_Date_MatchScore	Generates a match score based on person names and dates.
rule_Individual_Name_and_Email_MatchScore	Generates a match score based on person names and email addresses.
rule_Individual_Name_and_Phone_MatchScore	Generates a match score based on person names and telephone numbers.
rule_Individual_Name_and_SSN_MatchScore	Generates a match score based on person names, Social Security numbers, and identification data.

Name	Description
rule_Individual_Name_MatchScore	Generates a match score based on person names.
rule_USA_Address_MatchScore	Generates a match score based on United States address data.

U.S./Canada Demonstration Mappings

The demonstration mappings in the U.S./Canada accelerator use multiple rules to demonstrate data quality processes.

Find the demonstration mappings in the following repository location:

    [Informatica_DQ_Content]\Rules_Demo\US_Canada_Accelerator

The U.S./Canada accelerator includes the following demonstration mappings:

m_customer_data_US_demo

Parses, standardizes, and validates United States and Canadian data.

m_customer_matching_US_demo

Parses and standardizes identity data from the United States and perofrms identity match analysis on the data.

The mapping analyzes the following data combinations and generates match clusters for each combination:

- Person name and address data
- Person name and telephone number