

IBM InfoSphere QualityStage Standardization Rules
Designer
Version 11 Release 3

*Tutorial: Enhancing an address rule set
in the Standardization Rules Designer*



IBM InfoSphere QualityStage Standardization Rules
Designer
Version 11 Release 3

*Tutorial: Enhancing an address rule set
in the Standardization Rules Designer*



Note

Before using this information and the product that it supports, read the information in “Notices and trademarks” on page 39.

Contents

Tutorial: Enhancing an address rule set in the Standardization Rules Designer. . . 1

Setting up the tutorial environment.	5
Copying the tutorial files to a directory	5
Creating the tutorial project	6
Importing the tutorial rule set	6
Module 1: Creating a revision for a rule set and importing sample data	8
Lesson 1.1: Viewing rule sets and creating a revision in the Standardization Rules Designer	8
Lesson 1.2: Importing sample data.	10
Module 2: Classifying values	11
Lesson 2.1: Assigning values to an existing class by adding classification definitions	11
Lesson 2.2: Adding and populating a custom class.	15
Module 3: Adding a lookup table	18
Lesson 3.1: Adding a lookup table.	18
Lesson 3.2: Importing and adding lookup table definitions.	19
Module 4: Adding and modifying rules	22

Lesson 4.1: Modifying a rule to handle data correctly	22
Lesson 4.2: Identifying unhandled records and customizing output columns.	24
Lesson 4.3: Adding a rule for unhandled records	26
Lesson 4.4: Adding a rule to split values into different output columns	28
Lesson 4.5: Publishing a revision	30

Appendix A. Product accessibility . . . 33

Appendix B. Contacting IBM 35

Appendix C. Accessing the product documentation 37

Notices and trademarks 39

Index 45

Tutorial: Enhancing an address rule set in the Standardization Rules Designer

In this tutorial, you will use the IBM® InfoSphere® QualityStage® Standardization Rules Designer to enhance a rule set that standardizes address data. When you standardize data, you implement data quality standards that normalize data values and prepare data for uses such as matching and reporting.

In this tutorial, you will use data from the fictional Sample Outdoor Company, which sells and distributes products to third-party retailer stores and consumers. For the last several years, the company has steadily grown into a worldwide operation, selling their line of products to retailers in nearly every part of the world.

The fictional Sample Outdoor Company recently acquired several companies. As part of the acquisitions, the fictional Sample Outdoor Company received customer data from each company. The fictional Sample Outdoor Company wants to integrate the address data for these new customers into its current database, but the new data contains new types of information and is formatted inconsistently. The Sample Outdoor Company can use the IBM InfoSphere QualityStage Standardization Rules Designer to enhance the rule set that standardizes this type of data.

After the rule set is enhanced, the company will apply the rule set to a Standardize stage in a standardization job. When the standardization job is run, input data is standardized according to the logic that is specified in the enhanced rule set.

This tutorial guides you through some of the common tasks that you might complete when you enhance an address rule set in the Standardization Rules Designer. The following steps illustrate the sequence of actions in the tutorial:

1. In module 1, you open a revision for the rule set that requires enhancement in the Standardization Rules Designer. You also import sample data to use in the Standardization Rules Designer.
2. In module 2, you categorize the parts of the data. You add classification definitions that assign new values to existing classes and add a custom class for a new type of data. Figure 1 shows how each value in an example record for a fictional Sample Outdoor Company address can be assigned to a class.

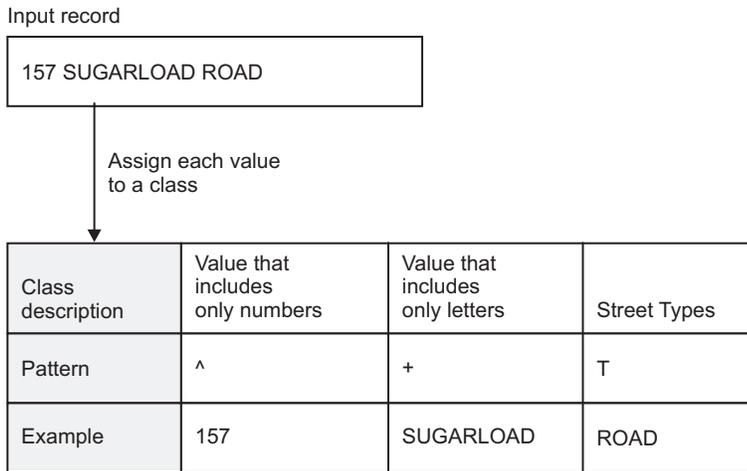


Figure 1. Each value in an example record is assigned to a class.

3. In module 3, you add a lookup table that converts abbreviated street names to complete street names. Figure 2 shows part of the lookup table that is added in the Standardization Rules Designer.

Value (abbreviated street name)	Returned value (complete street name)
MLK	MARTIN LUTHER KING
BWAY	BROADWAY
SM	STUART MILL

Figure 2. The lookup table converts abbreviated street names to complete street names.

4. In the first lesson of module 4, you modify a rule that was added in the Standardization Rules Designer. The rule is handling data incorrectly for some of the new addresses. Figure 3 shows the output with the current rule and the output after the rule is modified according to the data cleansing requirements of the fictional Sample Outdoor Company.

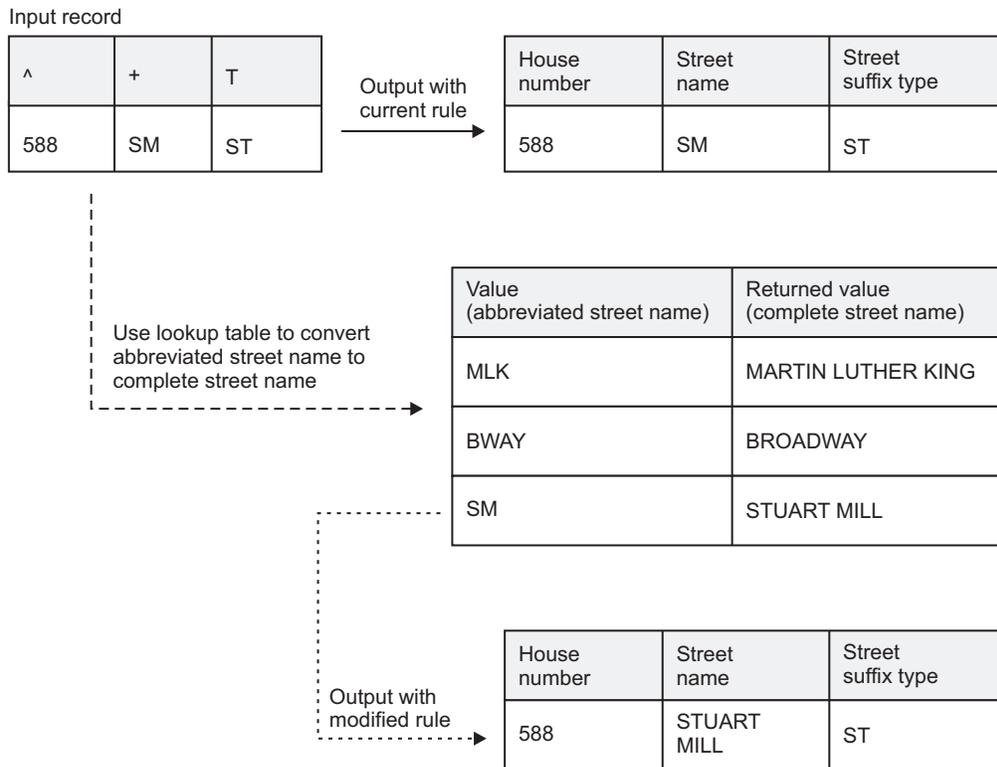


Figure 3. The current rule does not meet the data cleansing requirements of the fictional Sample Outdoor company. The modified rule uses a lookup table to convert abbreviated street names to complete street names.

- In the second and third lessons of module 4, you identify a pattern that is unhandled and add a rule to handle data that matches the pattern. Figure 4 shows how an example address record is handled by the new rule.

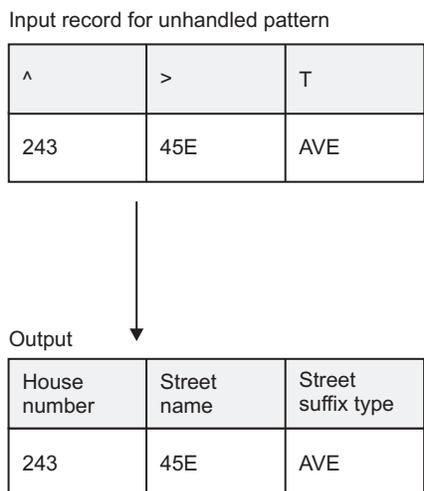


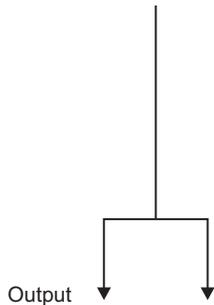
Figure 4. A new rule for an unhandled pattern adds address values to the appropriate output columns.

- In the fourth lesson of module 4, you add a rule that handles two distinct values that are concatenated in the input data. For example, if the input data contains the value 673HEGENBERGER, you can create a rule that splits the value into the values 673 and HEGENBERGER and places them in the appropriate output columns. Figure 5 shows how an example record for a fictional Sample

Outdoor Company product is handled by this rule.

Input record for concatenated values

>	T
673HEGENBERGER	ROAD



Output

House number	Sreet name	Street suffix type
673	HEGENBERGER	ROAD

Figure 5. The new rule splits a concatenated value into two distinct values and places them in the appropriate output columns. The rule also adds the street type value to the output column for street types.

Learning objectives

By completing the modules, you will learn about the concepts and tasks for enhancing rule sets:

- Import sample data to see how records from the sample data are affected by changes to parts of the rule set
- Use classifications to categorize parts of your data
- Add lookup tables to compare or convert data to specified values
- Add rules that apply actions to a group of related records

Time required

Before you begin the tutorial, you must set up your environment. The time that is required for setup depends on your current environment.

The modules each take 20 - 60 minutes to complete.

System requirements

The following components and applications must be installed on your system.

- IBM InfoSphere QualityStage Standardization Rules Designer
- IBM InfoSphere QualityStage with the following clients:
 - IBM InfoSphere DataStage® and QualityStage Designer
 - IBM InfoSphere DataStage and QualityStage Administrator

Prerequisites

Before you begin this tutorial, you must understand data quality concepts such as standardization, classification, and rules for data cleansing. Knowledge about InfoSphere DataStage and QualityStage concepts such as jobs, stages, and reports might be helpful, but is not required.

Notices: The Sample Outdoor Company, GO Sales, any variation of the Great Outdoors name, and Planning Sample, depict fictitious business operations with sample data used to develop sample applications for IBM and IBM customers. These fictitious records include sample data for sales transactions, product distribution, finance, and human resources. Any resemblance to actual names, addresses, contact numbers, or transaction values, is coincidental. Other sample files may contain fictional data manually or machine generated, factual data compiled from academic or public sources, or data used with permission of the copyright holder, for use as sample data to develop sample applications. Product names referenced may be the trademarks of their respective owners. Unauthorized duplication is prohibited.

Setting up the tutorial environment

Before you start this tutorial, you must prepare your system to run the tutorial.

Before you begin

- Ensure that the IBM InfoSphere QualityStage server component and the IBM InfoSphere DataStage and QualityStage Administrator client component is installed and configured appropriately.
- Ensure that user credentials are mapped to the engine tier operating system user credentials and that your user account has the following sets of security roles:
 - The Suite Administrator and DataStage and QualityStage Administrator roles are required to create a project. If you do not have these roles, a user with these roles must create a project for the tutorial and assign you the DataStage and QualityStage Developer or DataStage and QualityStage Production Manager role for the project.
 - The Suite User and DataStage and QualityStage Administrator or DataStage and QualityStage User roles are required to log in to the Standardization Rules Designer.

If you cannot log in to the Standardization Rules Designer, consult your system administrator.

About this task

The IBM InfoSphere QualityStage Standardization Rules Designer is a web-based application that interacts with the client, engine, and metadata repository tiers. Before you begin the tutorial, you must ensure that the sample data and tutorial rule set are available to the Standardization Rules Designer. You must also have a project that you can use for the tutorial.

Copying the tutorial files to a directory

You copy the tutorial files to a directory that you use during the tutorial. To complete the tutorial, you must have access to the tutorial files on the computer where you use the Standardization Rules Designer.

Procedure

1. Locate the `Standardization_Rules_Designer_address_tutorial.zip` file, which is on the installation media. In the directory that contains the installation media, the `Standardization_Rules_Designer_address_tutorial.zip` file is in the `parent_directory\TutorialData\QualityStage` directory. For example, the `Standardization_Rules_Designer_address_tutorial.zip` file might be in the `is-client\TutorialData\QualityStage` directory.
2. Extract the files from the `Standardization_Rules_Designer_address_tutorial.zip` file to a directory on the computer that you will use to connect to the Standardization Rules Designer.
3. If the directory that you extracted the files to is not on a computer where the client tier or engine tier is installed, copy the `Address_Rules_for_Tutorial.isx` file to a computer with the client tier or engine tier.

Creating the tutorial project

You create a new project for the tutorial to keep your tutorial exercises separate from the other work in InfoSphere QualityStage.

About this task

If you created a project for the tutorial that describes how to enhance a rule set for the product domain, you can use the same project for this tutorial.

Procedure

1. To start the IBM InfoSphere DataStage and QualityStage Administrator client, click **Start > All Programs > IBM InfoSphere Information Server > IBM InfoSphere DataStage and QualityStage Administrator**.
2. In the Attach to DataStage window, type your user name and password, and then click **Login**.
3. On the Projects page, click **Add**.
4. In the Add Project window, enter `StandardizationRulesDesignerTutorial` in the **Name** field, and then click **OK**.
5. If other users will complete the tutorial, assign the users an appropriate role for the project.
 - a. Select the `StandardizationRulesDesignerTutorial` project, and then click **Properties**.
 - b. Click the **Permissions** tab.
 - c. If the user that you want to add roles for is not in the list, add the user to the project.
 - d. From the list of project users, select the user that will complete the tutorial.
 - e. From the **User Roles** list, select **DataStage and QualityStage Developer** or **DataStage and QualityStage Production Manager**, and then click **OK**.
6. In the InfoSphere DataStage Administration window, click **Close**.

Importing the tutorial rule set

You use the InfoSphere Information Server Manager to import the tutorial rule set into the project that you created for the tutorial.

Before you begin

- “Copying the tutorial files to a directory” on page 5
- “Creating the tutorial project” on page 6

Procedure

1. Click **Start > Programs > IBM InfoSphere Information Server > IBM InfoSphere Information Server Manager**.
2. Connect to the metadata repository that contains the tutorial project:
 - a. In the Information Server Manager application window, right-click in the Repository view, and then click **Add Domain**.
 - b. Specify the application server and an InfoSphere Information Server administrator user name and password.
 - c. Click **OK**.
3. In the Repository navigation tree, right-click the StandardizationRulesDesignerTutorial project, and then click **Import**.
4. Import the Address_Rules_for_Tutorial.isx file:
 - a. Browse to the directory that contains the file. By default, the file is in the Standardization_Rules_Designer_address_tutorial directory where you copied the tutorial files.
 - b. Select the file, and then click **Open**.
 - c. Click **Import**.
 - d. Ensure that the StandardizationRulesDesignerTutorial project is selected, and then click **OK**.

Module 1: Creating a revision for a rule set and importing sample data

In this module, you will create a revision for a standardization rule set that requires enhancement in the Standardization Rules Designer. You will then import sample data to use in the Standardization Rules Designer.

Learning objectives

After completing the lessons in this module, you will understand the relevant concepts and tasks for rule sets and sample data:

- Create a revision for a rule set in the Standardization Rules Designer
- Import sample data

Time required

This module takes approximately 20 minutes to complete.

Prerequisites

Ensure that you completed the setup steps and have tutorial files and data loaded on your computer system.

Lesson 1.1: Viewing rule sets and creating a revision in the Standardization Rules Designer

In this lesson, you find the USADDRTutorial.SET rule set in the Standardization Rules Designer and create a revision for that rule set. When you create a revision, you can enhance the rule set through multiple iterations, working at different times and in collaboration with other users.

Overview

Because the USADDRTutorial.SET rule set does not meet the data quality requirements of the fictional Sample Outdoor Company, the rule set must be enhanced.

Learn more about rule sets:

Rule sets check and normalize input data. You can use rule sets to help understand the content and structure of source data and to standardize data and prepare the data for matching. Rule sets contain the following elements, which are discussed in greater detail as you work with them in the tutorial:

- Classifications
- Lookup tables
- Rules
- Output columns

Rule sets are used in jobs that include an Investigate stage or Standardize stage.

Before you can enhance this rule set, you must create a revision for the rule set in the Standardization Rules Designer.

Learn more about revisions:

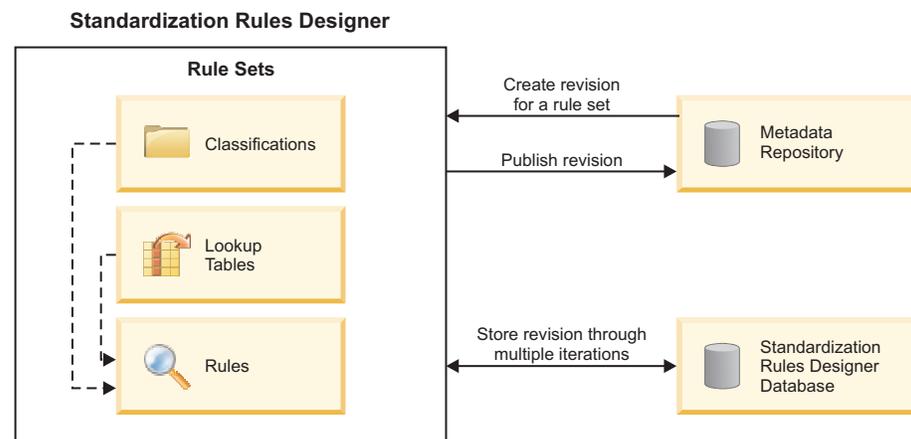
A revision is a collection of changes that were made to the rule set in the Standardization Rules Designer. One revision can be created for each rule set.

When you enhance a rule set in the Standardization Rules Designer, you work with a copy of the rule set that is stored in a database for the Standardization Rules Designer. You can keep a revision open for as long as you want, and the changes to the rule set can be viewed and modified by other users. When a revision is open, you can enhance the rule set in a collaborative way through multiple iterations that affect only the copy of the rule set. If you collaborate with other users, only one user can open the revision at a time.

When you want to update the actual rule set, you can publish the revision. When you publish the revision, the rule set in the metadata repository is updated with the changes that you made in the Standardization Rules Designer.

If you want to discard the current changes in the revision but keep the revision open, you can reset the revision. When you reset the revision, the version of the rule set in the Standardization Rules Designer database is replaced by the version in the metadata repository.

The following diagram shows how you can work with revisions throughout the lifecycle of a rule set in the Standardization Rules Designer.



Procedure

1. To open the Standardization Rules Designer, in your browser, go to https://host_name:secure_port_number/ibm/iis/qs/StandardizationRulesDesigner/, where *host_name* is the host name of the server where the Standardization Rules Designer is installed.
2. Log in to the Standardization Rules Designer. A list of servers that host IBM InfoSphere QualityStage projects is shown.
3. Expand the twistie for the server on which you created the StandardizationRulesDesignerTutorial project.
4. Expand **StandardizationRulesDesignerTutorial**.
5. Select the USADDRTutorial.SET rule set, and then click **Edit**.
6. In the window for the message about editing the rule set, click **OK**. A revision is created for the USADDRTutorial.SET rule set, and the Home page for the rule set is shown.

Lesson 1.2: Importing sample data

In this lesson, you import sample data to use in the Standardization Rules Designer.

Overview

The changes that you make to a rule set in the Standardization Rules Designer are informed by the sample data that you use. When you enhance a rule set in the Standardization Rules Designer, you can view how the changes affect the way that values and records in your sample data are handled. To work with your data, you must import it into the Standardization Rules Designer.

The fictional Sample Outdoor Company has prepared a sample data set to use in the Standardization Rules Designer. The sample data includes customer records from the companies that the fictional Sample Outdoor Company acquired.

Procedure

1. In the Standardization Rules Designer, click the **Home** tab.
2. In the navigation pane, click **Import Sample Data**.
3. Next to the **Select a Source** field, click **Import**.
4. Import the address_sample_data.csv file:
 - a. Browse to the directory that contains the file. By default, the file is in the Standardization_Rules_Designer_address_tutorial directory where you copied the tutorial files.
 - b. Select the file, and then click **Open**.

The Standardization Rules Designer shows the raw records that were imported and example records that were processed by the separation characters and strip characters.

What's next

In this module, you completed the following tasks:

- Created a revision for a rule set in the Standardization Rules Designer
- Imported sample data

You can log out of the Standardization Rules Designer and complete the next module later. The revision remains open.

In the next module, you will strengthen the contextual information that the new address data provides by categorizing parts of the data.

Module 2: Classifying values

In this module, you will strengthen the contextual information that the new data provides by categorizing parts of the data. You will add classification definitions that assign new values to existing classes and add a custom class to categorize the data.

The data for the new customer addresses that the fictional Sample Outdoor Company acquired includes the following types of values:

- New values to assign to an existing class
- A new type of value that the company wants to create a class for

Because some of the building types in the new customer address data are not assigned to the building type class, rules that handle patterns with the building type class do not handle the new data. In the first part of this module, you will identify how building types from the new addresses are currently classified. To ensure that the building types are handled by appropriate existing rules, you will add classification definitions that assign new building types to the building type class.

Additionally, some of the records for the addresses that the fictional Sample Outdoor Company recently acquired include information that is not part of the address, such as INVALID ADDRESS or BEWARE OF DOG. To categorize these values, you will add a custom class for exception data. After the class is added, you can write rules that properly handle exception data.

Learning objectives

After completing the lessons in this module, you will understand the relevant concepts and tasks for classifying values:

- Identify the class that a value belongs to
- Add classification definitions
- Add a custom class

Time required

This module takes approximately 30 minutes to complete.

Prerequisites

Ensure that you completed the setup steps and that you have the tutorial files and sample data loaded on your computer.

Lesson 2.1: Assigning values to an existing class by adding classification definitions

In this lesson, you identify how building types in new address data are currently classified. You then assign building type values to an existing custom class by adding classification definitions.

Overview

The fictional Sample Outdoor Company wants to ensure that records that contain the new building types are handled correctly. You can view the class that new

building type values belong to, and then use that information to see which rules, if any, handle the values. You can then assign building type values to a class by adding classification definitions.

Learn more about classifications:

In IBM InfoSphere QualityStage, records are represented as patterns. In the same way that a record is comprised of one or more values, patterns are comprised of one or more abstract characters, each of which represents a class. Classes strengthen the contextual information that patterns provide by identifying that the underlying value belongs to a particular category. For example, the fictional Sample Outdoor Company has records such as 157 SUGARLOAD ROAD that are represented by the pattern ^+T. The following table shows the contextual information that each class in the pattern ^+T provides.

Table 1. Each one-character class label indicates that the underlying value belongs to a particular category.

^	+	T
Value that includes only numbers	Value that includes only letters	Street Types

Information about values is provided by the following types of classes:

- Default classes provide basic information about the type of the value, such as whether the value is comprised of alphabetic characters, numeric characters, or some combination of both. Default classes are represented by a range of characters that is listed in the Standardization Rules Designer.
- Custom classes provide stronger contextual information about the type of the value, such as whether an alphabetic value is the name of a street type or the name of a building type. The one-character label for custom classes can be any letter in the Latin alphabet or 0, which indicates a null class.

Tasks in this lesson

1. “Identifying the class that a value belongs to”
2. “Assigning values to a class” on page 14

Identifying the class that a value belongs to

You can identify the default or custom class to which a building type value is currently assigned. If the building type values are not assigned to the L Building Types custom class, records that contain those values might not be handled by the appropriate rules.

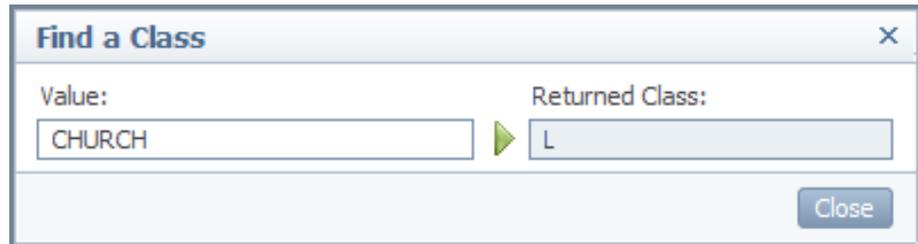
The customer address data that the fictional Sample Outdoor Company acquired includes the following building types:

- CHURCH
- INSTITUTE
- HOSTEL

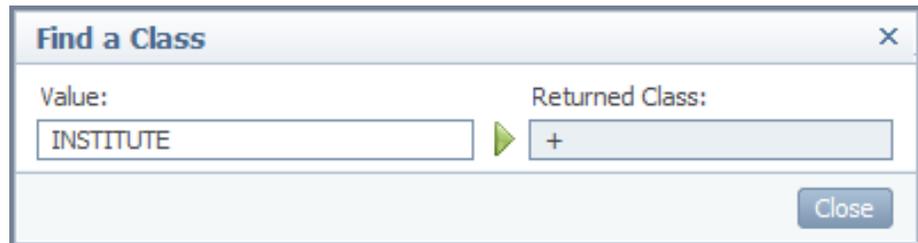
To ensure that records that contain these building types are handled properly, you must first identify what class the values belong to. You can also view which patterns contain the L Building Types class and view rules that handle records with those patterns.

Procedure

1. In the Standardization Rules Designer, click the **Classifications** tab.
2. Click **Find a Class**.
3. Identify the class that the building type values belong to:
 - a. In the **Value** field, enter the value CHURCH, and then click . The **Returned Class** field shows that this value is assigned to the L Building Types custom class.



- b. Repeat step 3a for the values INSTITUTE and HOSTEL. The **Returned Class** field shows that these values are assigned to the + default class.



- c. Click **Close**.
4. To access the patterns in the data and the rules for those patterns, open a rule group:
 - a. Click the **Rules** tab.
 - b. Select the Input_Overrides rule group, and then click **Open**.
5. Expand **Pattern Rule**. A list of patterns in the data is shown.

▼ Pattern Rule		Add Pattern...
Rule	Frequency	Unhandled
▶  ^+T	42.73%	0%
▶  ^D+T	5.34%	5.34%
▶  ^T+	4.59%	4.59%
▶  ^++T	3.65%	3.65%
▶  ^TT	3.37%	3.37%

6. Expand +L^+T, and then expand **Copy building details to building name output column**. A list of example records that are handled by this rule is shown. The list of records includes records that contain the value CHURCH. Because this value is assigned to the L Building Types class, the rule for this pattern handles records that contain the value.

7. Expand **++^+T**, and then expand **Unhandled Records**. A list of example records that match this pattern is shown. The list of records includes records that contain the values **INSTITUTE** and **HOSTEL**. Because these values are not assigned to the **L Building Types** class, records that contain the values are not handled by the appropriate rule.
8. To exit the **Input_Overrides** rule group, click **Select Rule Group** at the top of the **Rules** page.

To ensure that records that contain the values **INSTITUTE** and **HOSTEL** are handled by the appropriate rule, you can assign these values to the **L Building Types** class.

Assigning values to a class

You can assign the new building type values to the building types class by adding classification definitions manually.

Learn more about classification definitions:

Classification definitions include the following elements:

Value The input value that is defined.

Standard Value

A standardized spelling or representation of the value that an action or condition in a rule can use instead of the input value.

Class The class that the value is assigned to. The class is represented by a one-character class label.

Similarity Threshold

The degree of variation that is allowed in the spelling or representation of the value. The similarity threshold must be in the range 700 - 900. A similarity threshold of 900 indicates that only the exact value in the definition is affected by the definition.

If the similarity threshold is lower than 900, the definition might affect values that are similar to the value in the definition. For example, if you create a classification definition for the value **GRAMS** and set the similarity threshold to 800, the definition also affects the value **GRAM**.

Procedure

1. In the **Standardization Rules Designer**, click the **Classifications** tab.
2. Click **Define Values**.
3. Add a definition for a building type value:
 - a. Enter the value **INSTITUTE**.
 - b. From the **Class** list, select **L**.
 - c. Expand **Status of Definitions** and verify that **Active definition** is selected. An inactive definition does not affect the value.
 - d. Click **OK**.
4. To access the patterns in the data and the rules for those patterns, open a rule group:
 - a. Click the **Rules** tab.
 - b. Select the **Input_Overrides** rule group, and then click **Open**.
5. Expand **Pattern Rule**. A list of patterns in the data is shown.

6. Expand +L^+T, and then expand **Copy building details to building name output column**. A list of example records that are handled by this rule is shown. The list of records now includes records that contain the value INSTITUTE.
7. Expand ++^+T, and then expand **Unhandled Records**. A list of example records that have this pattern is shown. The list of records still includes records that contain the value HOSTEL.

Lesson 2.2: Adding and populating a custom class

In this lesson, you add a custom class to designate a new type of value. You then assign an initial set of values to this class by adding multiple classifications at the same time.

Overview

The address data that the fictional Sample Outdoor Company acquired includes values that are not part of the address. For example, some records contain information such as INVALID ADDRESS or BEWARE OF DOG. The company refers to this type of information as exception data. To distinguish exception data from address data, the company wants to add a separate custom class for values that indicate exception data. The company decides to use the X custom class for exception data.

Learn more about custom classifications:

A custom classification provides stronger contextual information about a value than a default classification. For example, if a definition does not assign the value TRAILCHEF to a custom class, the value is assigned to the + default class. This default class indicates that the value is a single alphabetic word. If a definition assigns the value to the Product Brand custom class, which is represented by the character B, the value provides more contextual information. When this information is provided, you can write rules that address a specific subset of the data and therefore handle that data more effectively.

Tasks in this lesson

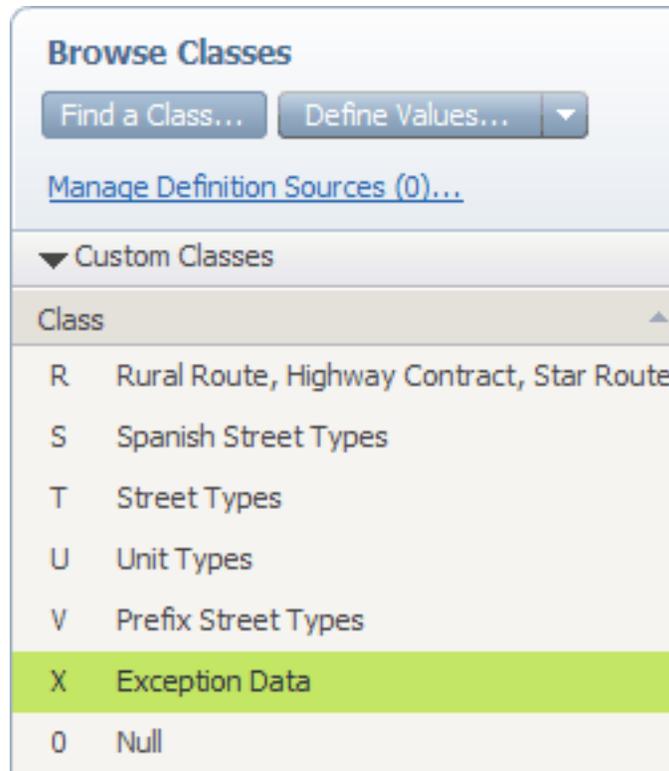
1. "Adding a custom class"
2. "Adding multiple classification definitions" on page 16

Adding a custom class

To create a class for exception data, you add the X custom class from the list of available classes.

Procedure

1. In the Standardization Rules Designer, click the **Classifications** tab.
2. In the **Custom Classes** section, click **Add Custom Class**.
3. From the **Available Classes** list, select **X**.
4. In the **Description** field, enter Exception Data.
5. Click **Add**. An entry for the new class is shown in the list.



6. Click **Close**.

Adding multiple classification definitions

You can assign more than one value to the new X class by adding multiple classification definitions at the same time.

If you add a new custom class in the Standardization Rules Designer, no values are assigned to that class initially. To populate the class, you add classification definitions that assign values to the class. For example, to populate the X custom class, you add classification definitions for values that represent exception data. After values are assigned to the class, you can add or modify rules to handle exception data values.

The fictional Sample Outdoor Company has identified values that indicate exception data in their company records. For example, the value USE is never part of an address in their data, but does appear in exception data such as USE LEFT MAILBOX and USE REAR ENTRANCE. In the new address data set, the following values indicate exception data:

- INVALID
- BEWARE
- SLEEPS
- USE

Procedure

1. In the Standardization Rules Designer, click the **Classifications** tab.
2. Click **Define Values**.
3. In the lower right of the Define Values window, click **Define Multiple**.
4. Select the **Apply class to all definitions** check box.

5. From the **Select Class** list, select **X**. In the Class column, **X** is selected automatically for each value that you define.
6. Enter **INVALID** in the first field in the Value column, then click outside the field. A second row is enabled.
7. Enter the following values by using the method from step 6.
 - **BEWARE**
 - **SLEEPS**
 - **USE**Enter each value in a separate row.
8. Expand **Status of Definitions** and verify that **Active definition** is selected. An inactive definition does not affect the value.
9. Click **OK**.

To view the values that are assigned to the X custom class, expand **Custom Classes** and select **X**.

What's next

In this module, you completed the following tasks:

- Identified the class that a value belongs to
- Added classification definitions
- Added a custom class

You can log out of the Standardization Rules Designer and complete the next module later. The revision remains open.

In the next module, you will add a lookup table that converts abbreviated street names to street names that are spelled out.

Module 3: Adding a lookup table

In this module, you will add a lookup table that converts abbreviated street names to street names that are spelled out.

The address data that the fictional Sample Outdoor Company acquired includes street names that are abbreviated. In some records, the abbreviations are initials. For example, the value MARTIN LUTHER KING is abbreviated as MLK, and DANIEL DAFOE is abbreviated as DD.

To ensure that street names are not ambiguous, the fictional Sample Outdoor Company created lookup table definitions for abbreviated street names. The company can use the lookup table definitions in rules that convert the abbreviated street names to street names that are spelled out.

In this module, you will add a new lookup table and import lookup table definitions into the Standardization Rules Designer. You will then add lookup table definitions for new abbreviations that are not included in the current lookup table.

Learning objectives

After completing the lessons in this module, you will understand the relevant concepts and tasks for lookup tables:

- Add a lookup table
- Import lookup table definitions
- Add lookup table definitions

Time required

This module takes approximately 20 minutes to complete.

Prerequisites

Ensure that you completed the setup steps and that you have the tutorial files and sample data loaded on your computer.

Lesson 3.1: Adding a lookup table

In this lesson, you add a lookup table to the USADDRTutorial.SET rule set that converts abbreviated street names to street names that are spelled out.

Overview

The fictional Sample Outdoor Company wants to use a lookup table to convert abbreviated street names to street names that are spelled out. You can add a lookup table to the rule set in the Standardization Rules Designer.

Learn more about using lookup tables:

Lookup tables contain definitions that rules can reference as part of actions or conditions. Actions or conditions might use a lookup table in the following ways:

- An action or condition can compare a particular value to a value in the lookup table. For example, a condition can stipulate that a rule handles a record only when a value in a particular position in the record is in the lookup table.

- An action can convert a particular value to a value in the lookup table. For example, an action might use a lookup table that contains geographic information to convert numeric place codes to place names.

Procedure

1. In the Standardization Rules Designer, click the **Lookup Tables** tab.
2. Click **Add Lookup Table**.
3. Enter information about the street name lookup table:
 - a. In the **Name** field, enter `Street_Name`.
 - b. In the **Description** field, enter Converts abbreviated street names to complete street names.
 - c. Click **OK**.

Lesson 3.2: Importing and adding lookup table definitions

In this lesson, you will import and add lookup table definitions to the `Street_Name` lookup table.

Overview

The fictional Sample Outdoor Company created a CSV file that contains lookup table definitions for the `Street_Name` lookup table. You can import these definitions into the lookup table that you added in the Standardization Rules Designer.

Recently, the company discovered a new abbreviation in the address data and identified a typographical error in a lookup table definition. You can update the lookup table by adding lookup table definitions in the Standardization Rules Designer.

Learn more about lookup table definitions:

Lookup table definitions include the following elements:

Value The input value that is defined.

Returned Value

A value that an action or condition in a rule can use instead of the input value.

Similarity Threshold

The degree of variation that is allowed in the spelling or representation of the value. The similarity threshold must be in the range 700 - 900. A similarity threshold of 900 indicates that only the exact value in the definition is affected by the definition.

If the similarity threshold is lower than 900, the definition might affect values that are similar to the value in the definition. For example, if you create a lookup table definition for the value `GRAMS` and set the similarity threshold to 800, the lookup table definition also affects the value `GRAM`.

Procedure

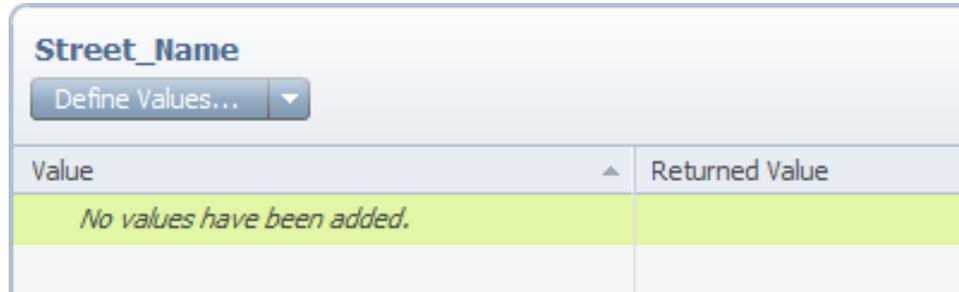
1. "Importing lookup table definitions" on page 20
2. "Adding lookup table definitions" on page 20

Importing lookup table definitions

You can import lookup table definitions from one or more files into a lookup table in the Standardization Rules Designer. You can also add lookup table definitions manually in the Standardization Rules Designer.

Procedure

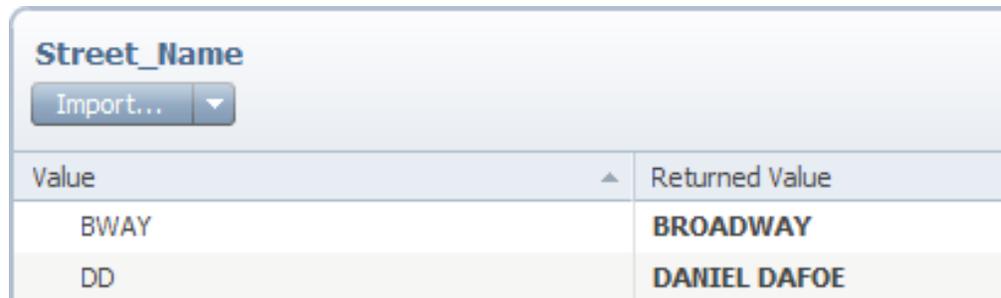
1. In the Standardization Rules Designer, click the **Lookup Tables** tab.
2. From the list of lookup tables, select the **Street_Name** table. A list of values in the lookup table is shown in the right pane. Because no values are defined for this lookup table, the list is empty.



The screenshot shows the 'Street_Name' lookup table interface. At the top, there is a 'Define Values...' button with a dropdown arrow. Below this is a table with two columns: 'Value' and 'Returned Value'. The table is currently empty, with a light green background and the text 'No values have been added.' centered in the 'Value' column.

3. From the **Define Values** list, select **Import**.
4. Import the `street_abbreviation_lookup.csv` file:
 - a. Browse to the directory that contains the file. By default, the file is in the `Standardization_Rules_Designer_address_tutorial` directory where you copied the tutorial files.
 - b. Select the file, and then click **Open**.
 - c. In the Import Definitions window, click **OK**.

The lookup table definitions are imported and the right pane shows the values in the lookup table that are defined.



The screenshot shows the 'Street_Name' lookup table interface after importing. The 'Define Values...' button has been replaced by an 'Import...' button with a dropdown arrow. The table now contains two rows of data:

Value	Returned Value
BWAY	BROADWAY
DD	DANIEL DAFOE

Adding lookup table definitions

You can add lookup table definitions manually in the Standardization Rules Designer.

Procedure

1. For the **Street_Name** table, select **Define Values** from the **Import** list.
2. Add a lookup table definition for the new value by completing the fields:
 - a. In the **Value** field, enter the abbreviation **IB**.
 - b. In the **Returned Value** field, enter the street name **IRLO BRONSON**.
 - c. Expand **Define Value As** and ensure that **Active definition** is selected. An inactive definition does not affect the value.

d. Click OK.

The right pane shows the value that you added a definition for.

Value	Returned Value
DD	DANIEL DAFOE
IB	IRLO BRONSON

3. Create a new lookup table definition for a value by repeating step 2 on page 20, but entering SP for the value and SARA PELLETIER for the returned value.
4. Verify that the new lookup table definition is the active definition:
 - a. From the list of lookup tables, select the Street_Name table. A list of values in the lookup table is shown in the right pane.
 - b. From the list of values, select SP. A list of definitions for the value is shown on the Define Value page. The definition that you added is selected, and is therefore the active definition.

Value:	Returned Value:
<input type="radio"/> SP	SARA PELLETIER
<input checked="" type="radio"/> SP	SARA PELLETIER

What's next

In this module, you completed the following tasks:

- Added a lookup table
- Imported lookup table definitions
- Added lookup table definitions

In the next module, you will create and modify rules that handle the addresses in customer records for the fictional Sample Outdoor Company.

Module 4: Adding and modifying rules

In this module, you will add and modify rules to ensure that records for new address data are standardized correctly.

The fictional Sample Outdoor Company wants to add or modify rules so that the records from their new address data are handled properly.

First, the company looks at the rule that handles the most common pattern in the data. The records that match this pattern are handled relatively well, but some of the street names are abbreviated. The company wants to use the lookup table that was added in module 3 to convert abbreviated street names to street names that are spelled out.

Next, the company determines that street names that contain leading digits followed by letters are not handled at all. For example, street names such as 45TH and 76E are not handled by any rules. The company must add a rule to handle this data.

When the company looks at the remaining unhandled patterns, the company finds a pattern in which the house number and street name are concatenated into one value. For example, in the following record, the values 637 and HEGENBERGER are concatenated into 637HEGENBERGER: 673HEGENBERGER ROAD. The company must add a rule for this pattern that splits the data into the appropriate output columns.

After rules are added and modified, the company decides to update the actual rule set by publishing the revision.

Learning objectives

After completing the lessons in this module you will understand the concepts and tasks for adding and modifying rules and publishing revisions:

- Modify a rule
- Identify unhandled patterns and view example records for those patterns
- Customizing the output columns that are available in the Standardization Rules Designer
- Add a basic rule by mapping values in a record to the appropriate output columns
- Add a rule that splits a concatenated value into two distinct values
- Publish a revision

Time required

This module takes approximately 60 minutes to complete.

Prerequisites

Ensure that you completed the setup steps and have the tutorial files and sample data loaded on your computer system.

Lesson 4.1: Modifying a rule to handle data correctly

In this lesson, you ensure that all records that match a pattern are handled correctly by modifying the rule that handles those records.

Overview

To prevent delivery errors, the fictional Sample Outdoor Company wants to ensure that all street names in address data are spelled out. For the most common pattern in the data, some of the street names are abbreviated. You can modify the rule that handles this data to use the lookup table that you added in module 3. The rule uses the lookup table to convert abbreviated street names to street names that are spelled out.

Learn more about rules and rule groups:

Rules are processes that standardize groups of related records. Rules can apply to records that match the same pattern or to exact strings of text. When you add or modify a rule, you map values in the input records to output columns, specify actions that manipulate the data, and identify conditions to ensure that rules apply only to the correct records.

A rule group is a collection of rules that are applied to records at the same point in the standardization process. To ensure that rules are applied in a particular order, you can organize the rules into rule groups in the Standardization Rules Designer. Then, you can invoke the rule groups from the pattern-action specification (previously called the pattern-action file).

Procedure

1. Click the **Rules** tab, select the Input_Overrides rule group, and then click **Open**.
2. Expand **Pattern Rule**. A list of patterns in the data is shown.

▼ Pattern Rule		Add Pattern...	
Rule	Frequency	Unhandled	
▶  ^+T	42.73%	0%	
▶  ^D+T	5.34%	5.34%	
▶  ^T+	4.59%	4.59%	
▶  ^++T	3.65%	3.65%	
▶  ^TT	3.37%	3.37%	

3. Expand ^+T, and then expand **Copy address data to output columns**.
4. Right-click the value that is in the StreetName output column, and then click **Edit Action**. In the Actions window, you can manipulate the data that is sent to the output column.
5. In the **Look up the Object** section, click **Yes**.
6. Use the Street_Name lookup table to convert abbreviated street names to complete street names:
 - a. From the **Source** list, select **Lookup Table**.
 - b. From the **Lookup Table** list, select Street_Name.
 - c. From the **If Found** list, select **Convert to Returned Value**. If a street name value is in the Street_Name lookup table, the abbreviated street name will be converted to the street name that is spelled out.

- d. From the **If Not Found** list, select **Use Value**.
 - e. Click **OK**.
7. Click **Apply**. On the Define Rule page, example records that contain an abbreviated street name show the complete street name in the StreetName output column. For example, if the example record contains the abbreviated street name SM, the value STUART MILL is in the StreetName output column.

The screenshot shows an 'Example Record' window with two rows of input and output fields. The first row has input fields for '588', 'SM', and 'ST', and output fields for '588' and 'STUART MILL'. The second row has input fields for 'ST' and output fields for 'ST'.

Lesson 4.2: Identifying unhandled records and customizing output columns

In this lesson, you identify records that are not handled by rules in the rule set so that you can add a rule for those records. You then hide output columns that you will not need when you add a rule for the unhandled records.

Tasks in this lesson

1. "Identifying unhandled records"
2. "Customizing output columns" on page 25

Identifying unhandled records

A subject matter expert from the fictional Sample Outdoor Company ran a Standardization Quality Assessment (SQA) report for a job that standardized the acquired customer data. The job applied the USADDRTutorial.SET rule set to the data. When the subject matter expert reviewed the SQA report, the subject matter expert identified unhandled records and the patterns that the unhandled records match.

The subject matter expert determined that some of the unhandled records include street names that contain leading digits followed by letters. For example, records that include street names such as 45TH and 76E were not handled by any rules. The subject matter expert also noticed that records that contain these street names often match the pattern $\wedge > T$. For example, the record 243 45E AV matches the pattern $\wedge > T$.

Table 2. The values 243 45E AV match the pattern $\wedge > T$

\wedge	$>$	T
Value that includes only numbers	Leading digits followed by letters	Street Type

You can identify unhandled records that contain the type of street name that the subject matter expert wants a rule to be added for. Later, you can add a rule that handles these records.

Learn more about unhandled records:

Unhandled records are records that no rule in the rule set applies to. Unhandled records can be records from the sample data or records that are generated by the system.

When you use the Standardization Rules Designer, you manage your standardization processes most effectively when you add rules that affect a large number of records in your data. After you identify unhandled records, you can add rules that apply to those records.

Procedure

1. Click the **Rules** tab, select the Input_Overrides rule group, and then click **Open**.
2. Expand **Pattern Rule**. A list of patterns in the data is shown.

▼ Pattern Rule		Add Pattern...	
Rule	Frequency	Unhandled	
▶ ^+T	42.73%	0%	
▶ ^D+T	5.34%	5.34%	
▶ ^T+	4.59%	4.59%	
▶ ^++T	3.65%	3.65%	
▶ ^TT	3.37%	3.37%	

3. Filter the list of patterns to show only the pattern ^>T:
 - a. In the upper-right corner of the Browse Rules pane, click .
 - b. Verify that the **Enable filter** check box is selected.
 - c. From the **Condition** list, select **Equals**.
 - d. In the **Pattern** field, enter ^>T, and then click **OK**.

The list is filtered, and only the pattern ^>T is shown.

▼ Pattern Rule		Add Pattern...	
Rule	Frequency	Unhandled	
▶ ^>T	1.68%	1.68%	

4. Expand ^>T. No rules exist for this pattern. All of the records are unhandled.

Because no rules have been created for this pattern, and a subject matter expert has determined that this pattern is important, you can affect important records in your data by adding a rule for this pattern. In the next lesson, you will add a rule for this pattern.

Customizing output columns

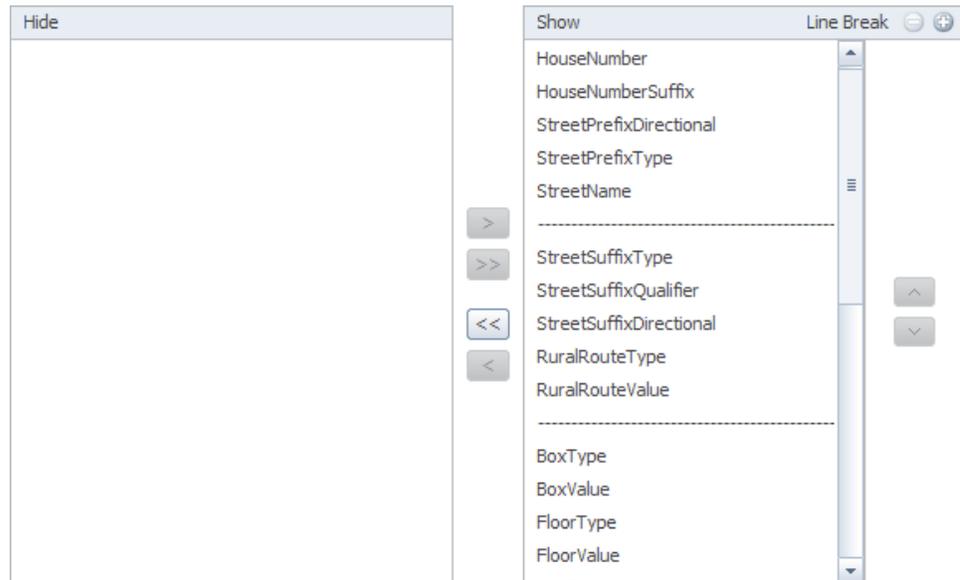
You can hide output columns that are not required for a rule.

The rule that the fictional Samples Outdoor Company wants to add for the ^>T pattern does not require many of the output columns that are shown on the Define

Rule page. To simplify the pane where the rule is added, the company wants to hide the output columns that are not needed from the Define Rule page.

Procedure

1. From the ^>T pattern, click **Unhandled Records**. An example record and list of output columns is shown.
2. On the Unhandled Records pane, click **Customize Output Columns**. A list of the columns that are hidden and the columns that are displayed is shown.



3. Customize the output columns so that only a subset of the output columns are shown:

- a. Click  to move all of the output columns to the list of columns that are hidden.
- b. Select the HouseNumber output column and click  to move the output column to the list of columns that are shown.
- c. Repeat step 3b for the StreetName, StreetSuffixType, and AdditionalAddress output columns.
- d. Click **OK**.

The Define Rule page does not show the output columns that you just hid.

Lesson 4.3: Adding a rule for unhandled records

In this lesson, you add a rule to handle an important pattern in the data.

Overview

The fictional Sample Outdoor Company wants to add a rule for a pattern that a subject matter expert identified as an important pattern.

Learn more about actions:

An action is a part of a rule that specifies how the rule processes a record. An action can specify what information is moved from the input data to the output

columns and how that data is manipulated. In the Standardization Rules Designer, you can use an action to manipulate data in the following ways:

- Specify the part of the input data that is acted on. The part can be an exact value, standard value, one or more characters in the value, or a literal.
- Compare or convert the selected part of the input data to values in a lookup table or list.
- Add the output data to an output column and specify a leading separator between that data and other data in the column.

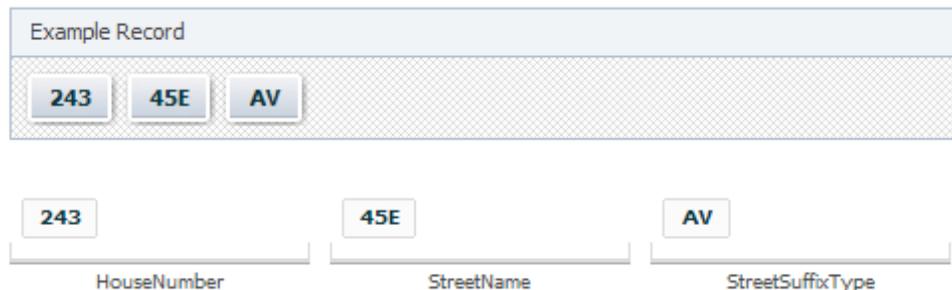
Procedure

1. Click the **Rules** tab, select the Input_Overrides rule group, and then click **Open**.
2. Expand **Pattern Rule**. A list of patterns in the data is shown.
3. Expand **^>T**, and then click **Unhandled Records**. An example record and list of output columns is shown.
4. If the example record on the Define Rule page is not 243 45E AV, select 243 45E AV from the list of example records.
5. To map the address values from the example record to the appropriate output columns, drag the values to the column. The following table shows the output column for each address value.

Table 3. Output columns for each address value in the example record

Value	Output column
243	HouseNumber
45E	StreetName
AV	StreetSuffixType

The values from the example records are mapped to the output columns.



6. Apply the rule to the pattern:
 - a. Click **Apply as New Rule**.
 - b. In the **Rule Name** field, enter Map address values to output columns.
 - c. Click **OK**.

The rule is added and is shown under the ^>T pattern.

▼ 📁 ^>T	1.68%
▶ 📌 Map address values to output columns	1.68%

7. If the list of patterns is still filtered, remove the filter.
 - a. In the upper-right corner of the Browse Rules pane, click .

- b. Clear the **Enable filter** check box, and then click **OK**.

Lesson 4.4: Adding a rule to split values into different output columns

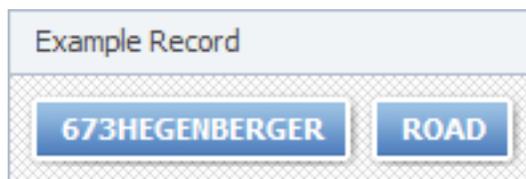
In this lesson, you add a rule that splits a value and maps each part of the split value to a different output column.

Overview

Records in input data might contain information that is concatenated. For example, the input data might contain the value 637HEGENBERGER. This value concatenates 637, which represents a house number, and HEGENBERGER, which represents a street name. The fictional Sample Outdoor Company determined that many records that contain concatenated values match the pattern >T. The company wants to distinguish the house number from the street name in these records. To split the concatenated values, you can add a rule that maps each part of the input value to a different output column.

Procedure

1. Click the **Rules** tab, select the Input_Overrides rule group, and then click **Open**.
2. Expand **Pattern Rule**. A list of patterns in the data is shown.
3. Expand >T, and then click **Unhandled Records**. An example record is shown. One of the values in the record, which is represented in the pattern as >, includes leading digits followed by letters.

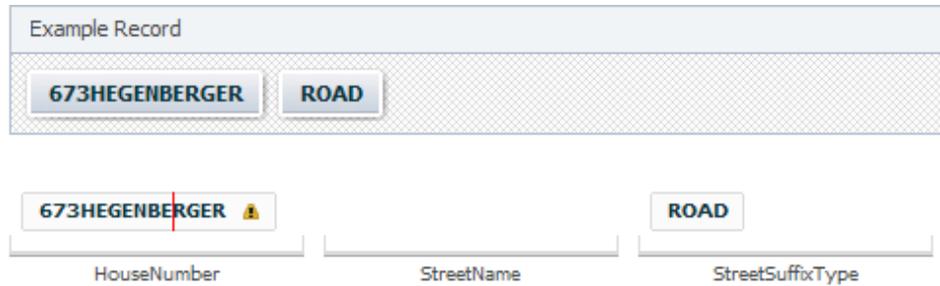


4. If the example record on the Define Rule page is not 673HEGENBERGER ROAD, select 673HEGENBERGER ROAD from the list of example records. If 673HEGENBERGER ROAD is not in the list of example records, increase the number of records in the list.
 - a. In the upper-right corner of the Browse Rules pane, click .
 - b. From the **Records to Display** list, select **100**.
 - c. Click **OK**. The list of example records includes 673HEGENBERGER ROAD. You can select the record from the list.
5. Map the values from the example record to the appropriate output columns by dragging the values to the column. The following table shows the output column for each value.

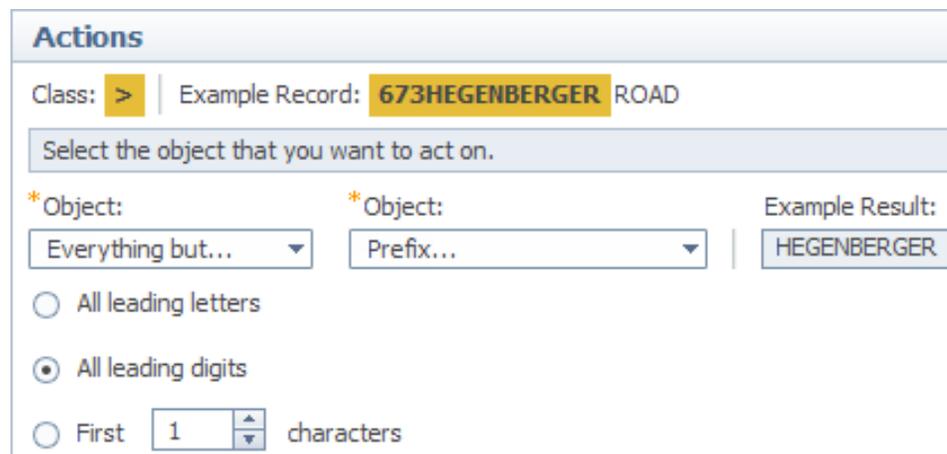
Table 4. Output columns for each address value in the example record

Value	Output column
673HEGENBERGER	HouseNumber
ROAD	StreetSuffixType

The values from the example records are mapped to the output columns. However, the value in the HouseNumber output column contains a value that belongs in the StreetName output column.

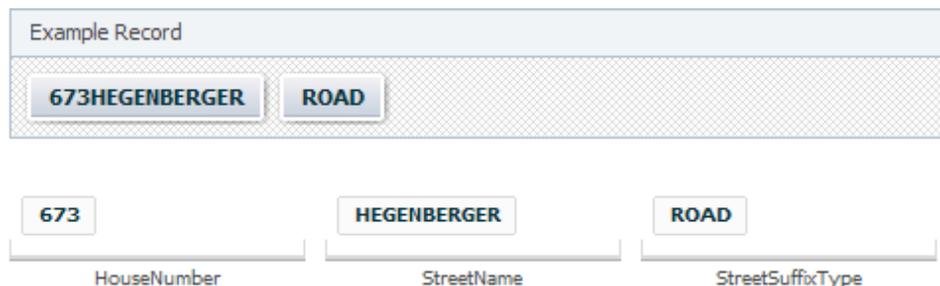


6. In the HouseNumber output column, right-click 673HEGENBERGER, and then click **Edit Action**. In the Action windows, you can manipulate the data that is sent to the output column.
7. Edit the action to add only the leading digits to the HouseNumber output column:
 - a. From the **Object** list, select **Prefix**.
 - b. Click **All leading digits**. The **Example Result** field shows the value 673.
8. Click **Specify action for remaining characters**, and then click **OK**. The **Actions** window opens. Options are selected that specify an action for the characters that you did not map to the HouseNumber output column.



9. Map the remaining characters to the StreetName output column:
 - a. From the **Output Column** list, select **StreetName**.
 - b. Click **OK**.

The values are shown in the correct output columns.



10. Apply the rule to the pattern:
 - a. Click **Apply as New Rule**.
 - b. In the **Name** field, enter Split house number and street name.

c. Click OK.

The rule is added and is shown under the >T pattern.

▼ 📁 >T	1.49%
▶ 📁 Split house number and street name	1.49%

Lesson 4.5: Publishing a revision

In this lesson, you publish a revision to the metadata repository so that the enhanced rule set can be used in a standardization job.

Overview

After you make changes to a rule set in the Standardization Rules Designer, you can apply those changes to the rule set in the metadata repository by publishing the revision. After you publish the revision, the rule set must be provisioned in the Designer client before the rule set can be applied in a standardization job.

The fictional Sample Outdoor Company has updated the classifications, lookup tables, and rules in the version of the USADDRTutorial.SET rule set that is stored on the Standardization Rules Designer database. To ensure that the version of the rule set that is used in standardization jobs reflects these changes, you publish the revision to the metadata repository.

Procedure

1. In the Standardization Rules Designer, click the **Home** tab.
2. In the navigation pane, click **View and Publish Revision**.
3. Optional: In the Properties pane, enter information about the rule set:
 - a. In the **Notes** field, enter a description of the changes that you made to the rule set since the revision was published.
 - b. Click **Apply**.
4. Click **Publish**. In the window for the message about publishing changes, click **Yes**.

The rule set in the metadata repository is updated with the changes that you made in the Standardization Rules Designer.

What's next

In this module, you completed the following tasks:

- Modified a rule
- Identified unhandled patterns and viewed example records for those patterns
- Managed the output columns that are shown in the Standardization Rules Designer
- Added a basic rule by mapping values in a record to the appropriate output columns
- Added a rule that splits a concatenated value into two distinct values
- Published a revision

You have completed the tutorial for enhancing an address rule set in the IBM InfoSphere QualityStage Standardization Rules Designer.

You can now complete one or more of the following tasks:

- Complete the tutorial for enhancing a product rule set in the Standardization Rules Designer.
- Import your own sample data and use the Standardization Rules Designer to enhance a rule set that applies to the data
- Learn more about IBM InfoSphere QualityStage and the Standardization Rules Designer by reading the *IBM InfoSphere QualityStage User's Guide*

Appendix A. Product accessibility

You can get information about the accessibility status of IBM products.

The IBM InfoSphere Information Server product modules and user interfaces are not fully accessible.

For information about the accessibility status of IBM products, see the IBM product accessibility information at http://www.ibm.com/able/product_accessibility/index.html.

Accessible documentation

Accessible documentation for InfoSphere Information Server products is provided in an information center. The information center presents the documentation in XHTML 1.0 format, which is viewable in most web browsers. Because the information center uses XHTML, you can set display preferences in your browser. This also allows you to use screen readers and other assistive technologies to access the documentation.

The documentation that is in the information center is also provided in PDF files, which are not fully accessible.

IBM and accessibility

See the IBM Human Ability and Accessibility Center for more information about the commitment that IBM has to accessibility.

Appendix B. Contacting IBM

You can contact IBM for customer support, software services, product information, and general information. You also can provide feedback to IBM about products and documentation.

The following table lists resources for customer support, software services, training, and product and solutions information.

Table 5. IBM resources

Resource	Description and location
IBM Support Portal	You can customize support information by choosing the products and the topics that interest you at www.ibm.com/support/entry/portal/Software/Information_Management/InfoSphere_Information_Server
Software services	You can find information about software, IT, and business consulting services, on the solutions site at www.ibm.com/businesssolutions/
My IBM	You can manage links to IBM Web sites and information that meet your specific technical support needs by creating an account on the My IBM site at www.ibm.com/account/
Training and certification	You can learn about technical training and education services designed for individuals, companies, and public organizations to acquire, maintain, and optimize their IT skills at http://www.ibm.com/training
IBM representatives	You can contact an IBM representative to learn about solutions at www.ibm.com/connect/ibm/us/en/

Appendix C. Accessing the product documentation

Documentation is provided in a variety of formats: in the online IBM Knowledge Center, in an optional locally installed information center, and as PDF books. You can access the online or locally installed help directly from the product client interfaces.

IBM Knowledge Center is the best place to find the most up-to-date information for InfoSphere Information Server. IBM Knowledge Center contains help for most of the product interfaces, as well as complete documentation for all the product modules in the suite. You can open IBM Knowledge Center from the installed product or from a web browser.

Accessing IBM Knowledge Center

There are various ways to access the online documentation:

- Click the **Help** link in the upper right of the client interface.
- Press the F1 key. The F1 key typically opens the topic that describes the current context of the client interface.

Note: The F1 key does not work in web clients.

- Type the address in a web browser, for example, when you are not logged in to the product.

Enter the following address to access all versions of InfoSphere Information Server documentation:

```
http://www.ibm.com/support/knowledgecenter/SSZJPZ/
```

If you want to access a particular topic, specify the version number with the product identifier, the documentation plug-in name, and the topic path in the URL. For example, the URL for the 11.3 version of this topic is as follows. (The ⇒ symbol indicates a line continuation):

```
http://www.ibm.com/support/knowledgecenter/SSZJPZ_11.3.0/⇒  
com.ibm.swg.im.iis.common.doc/common/accessingiidoc.html
```

Tip:

The knowledge center has a short URL as well:

```
http://ibm.biz/knowctr
```

To specify a short URL to a specific product page, version, or topic, use a hash character (#) between the short URL and the product identifier. For example, the short URL to all the InfoSphere Information Server documentation is the following URL:

```
http://ibm.biz/knowctr#SSZJPZ/
```

And, the short URL to the topic above to create a slightly shorter URL is the following URL (The ⇒ symbol indicates a line continuation):

```
http://ibm.biz/knowctr#SSZJPZ_11.3.0/com.ibm.swg.im.iis.common.doc/⇒  
common/accessingiidoc.html
```

Changing help links to refer to locally installed documentation

IBM Knowledge Center contains the most up-to-date version of the documentation. However, you can install a local version of the documentation as an information center and configure your help links to point to it. A local information center is useful if your enterprise does not provide access to the internet.

Use the installation instructions that come with the information center installation package to install it on the computer of your choice. After you install and start the information center, you can use the **iisAdmin** command on the services tier computer to change the documentation location that the product F1 and help links refer to. (The \Rightarrow symbol indicates a line continuation):

Windows

```
IS_install_path\ASBServer\bin\iisAdmin.bat -set -key  $\Rightarrow$   
com.ibm.iis.infocenter.url -value http://<host>:<port>/help/topic/
```

AIX® Linux

```
IS_install_path/ASBServer/bin/iisAdmin.sh -set -key  $\Rightarrow$   
com.ibm.iis.infocenter.url -value http://<host>:<port>/help/topic/
```

Where <host> is the name of the computer where the information center is installed and <port> is the port number for the information center. The default port number is 8888. For example, on a computer named server1.example.com that uses the default port, the URL value would be `http://server1.example.com:8888/help/topic/`.

Obtaining PDF and hardcopy documentation

- The PDF file books are available online and can be accessed from this support document: <https://www.ibm.com/support/docview.wss?uid=swg27008803&wv=1>.
- You can also order IBM publications in hardcopy format online or through your local IBM representative. To order publications online, go to the IBM Publications Center at <http://www.ibm.com/e-business/linkweb/publications/servlet/pbi.wss>.

Notices and trademarks

This information was developed for products and services offered in the U.S.A. This material may be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

Notices

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785 U.S.A.

For license inquiries regarding double-byte character set (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Corporation
J46A/G4
555 Bailey Avenue
San Jose, CA 95141-1003 U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information is for planning purposes only. The information herein is subject to change before the products described become available.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Each copy or any portion of these sample programs or any derivative work, must include a copyright notice as follows:

© (your company name) (year). Portions of this code are derived from IBM Corp. Sample Programs. © Copyright IBM Corp. _enter the year or years_. All rights reserved.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

Privacy policy considerations

IBM Software products, including software as a service solutions, ("Software Offerings") may use cookies or other technologies to collect product usage information, to help improve the end user experience, to tailor interactions with the end user or for other purposes. In many cases no personally identifiable information is collected by the Software Offerings. Some of our Software Offerings can help enable you to collect personally identifiable information. If this Software Offering uses cookies to collect personally identifiable information, specific information about this offering's use of cookies is set forth below.

Depending upon the configurations deployed, this Software Offering may use session or persistent cookies. If a product or component is not listed, that product or component does not use cookies.

Table 6. Use of cookies by InfoSphere Information Server products and components

Product module	Component or feature	Type of cookie that is used	Collect this data	Purpose of data	Disabling the cookies
Any (part of InfoSphere Information Server installation)	InfoSphere Information Server web console	<ul style="list-style-type: none"> • Session • Persistent 	User name	<ul style="list-style-type: none"> • Session management • Authentication 	Cannot be disabled
Any (part of InfoSphere Information Server installation)	InfoSphere Metadata Asset Manager	<ul style="list-style-type: none"> • Session • Persistent 	No personally identifiable information	<ul style="list-style-type: none"> • Session management • Authentication • Enhanced user usability • Single sign-on configuration 	Cannot be disabled

Table 6. Use of cookies by InfoSphere Information Server products and components (continued)

Product module	Component or feature	Type of cookie that is used	Collect this data	Purpose of data	Disabling the cookies
InfoSphere DataStage	Big Data File stage	<ul style="list-style-type: none"> • Session • Persistent 	<ul style="list-style-type: none"> • User name • Digital signature • Session ID 	<ul style="list-style-type: none"> • Session management • Authentication • Single sign-on configuration 	Cannot be disabled
InfoSphere DataStage	XML stage	Session	Internal identifiers	<ul style="list-style-type: none"> • Session management • Authentication 	Cannot be disabled
InfoSphere DataStage	IBM InfoSphere DataStage and QualityStage Operations Console	Session	No personally identifiable information	<ul style="list-style-type: none"> • Session management • Authentication 	Cannot be disabled
InfoSphere Data Click	InfoSphere Information Server web console	<ul style="list-style-type: none"> • Session • Persistent 	User name	<ul style="list-style-type: none"> • Session management • Authentication 	Cannot be disabled
InfoSphere Data Quality Console		Session	No personally identifiable information	<ul style="list-style-type: none"> • Session management • Authentication • Single sign-on configuration 	Cannot be disabled
InfoSphere QualityStage Standardization Rules Designer	InfoSphere Information Server web console	<ul style="list-style-type: none"> • Session • Persistent 	User name	<ul style="list-style-type: none"> • Session management • Authentication 	Cannot be disabled
InfoSphere Information Governance Catalog		<ul style="list-style-type: none"> • Session • Persistent 	<ul style="list-style-type: none"> • User name • Internal identifiers • State of the tree 	<ul style="list-style-type: none"> • Session management • Authentication • Single sign-on configuration 	Cannot be disabled
InfoSphere Information Analyzer	Data Rules stage in the InfoSphere DataStage and QualityStage Designer client	Session	Session ID	Session management	Cannot be disabled

If the configurations deployed for this Software Offering provide you as customer the ability to collect personally identifiable information from end users via cookies and other technologies, you should seek your own legal advice about any laws applicable to such data collection, including any requirements for notice and consent.

For more information about the use of various technologies, including cookies, for these purposes, see IBM's Privacy Policy at <http://www.ibm.com/privacy> and IBM's Online Privacy Statement at <http://www.ibm.com/privacy/details> the section entitled "Cookies, Web Beacons and Other Technologies" and the "IBM Software Products and Software-as-a-Service Privacy Statement" at <http://www.ibm.com/software/info/product-privacy>.

Trademarks

IBM, the IBM logo, and [ibm.com](http://www.ibm.com)[®] are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at www.ibm.com/legal/copytrade.shtml.

The following terms are trademarks or registered trademarks of other companies:

Adobe is a registered trademark of Adobe Systems Incorporated in the United States, and/or other countries.

Intel and Itanium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows and Windows NT are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java[™] and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

The United States Postal Service owns the following trademarks: CASS, CASS Certified, DPV, LACS^{Link}, ZIP, ZIP + 4, ZIP Code, Post Office, Postal Service, USPS and United States Postal Service. IBM Corporation is a non-exclusive DPV and LACS^{Link} licensee of the United States Postal Service.

Other company, product or service names may be trademarks or service marks of others.

Index

C

customer support
 contacting 35

L

legal notices 39

P

product accessibility
 accessibility 33
product documentation
 accessing 37

R

rule set extensions
 address data tutorial 1

S

software services
 contacting 35
Standardization Rules Designer
 address data tutorial 1
support
 customer 35

T

trademarks
 list of 39



Printed in USA

SC19-4334-00

