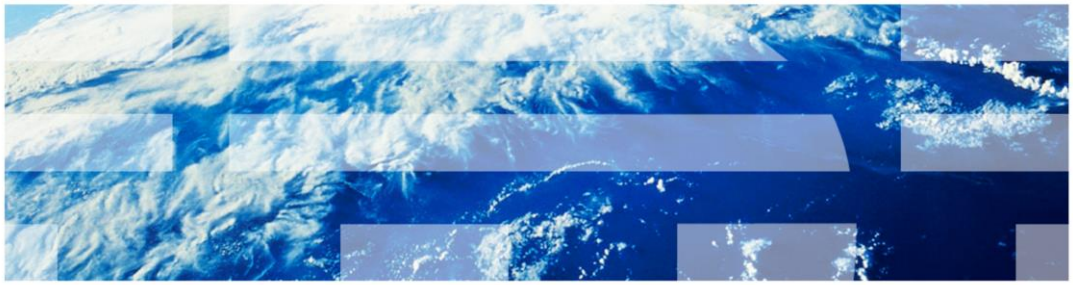
IBM

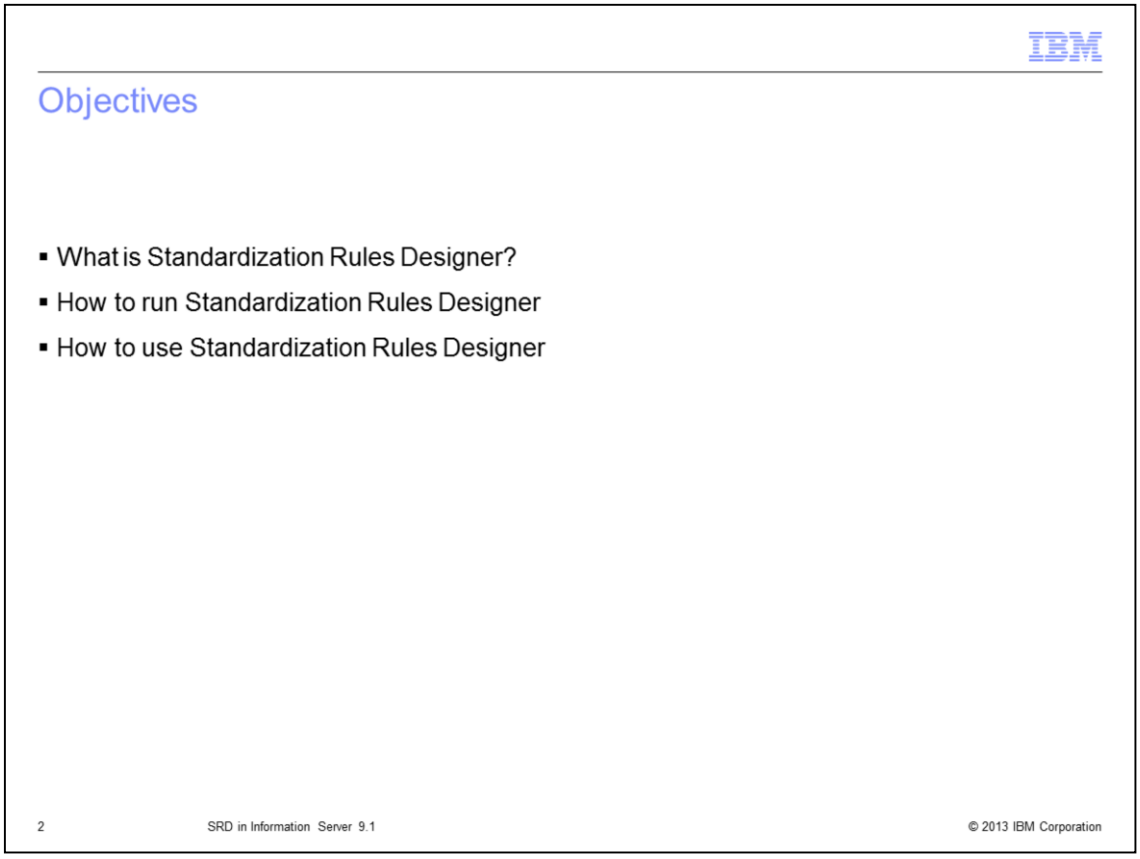
InfoSphere Information Server

Using the QualityStage Standardization Rules Designer (SRD) in Information Server 9.1

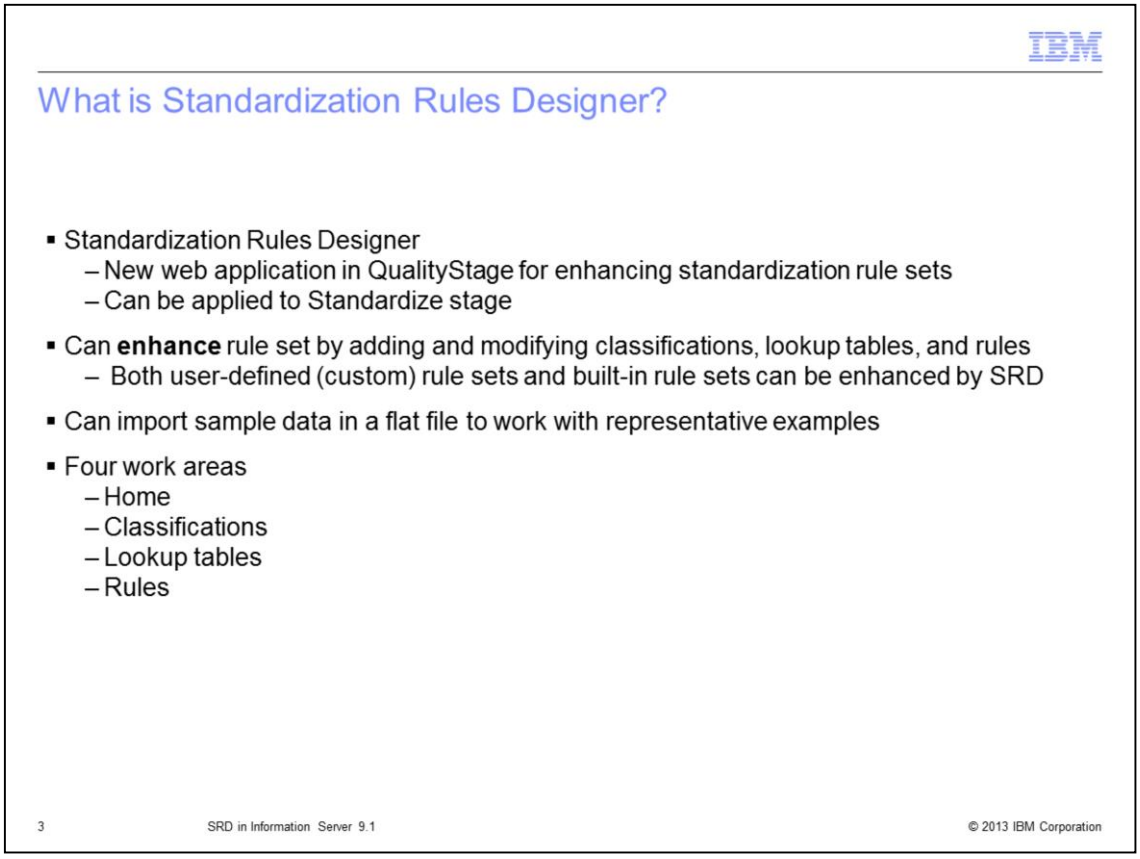


© 2013 IBM Corporation

This presentation will discuss what the QualityStage Standardization Rules Designer is and how to use the SRD with QualityStage version 9.1. The QualityStage Standardization Rules Designer is referred to as SRD throughout this presentation.



The objectives of this presentation are to understand what SRD is, to understand how to launch the SRD and how to use the SRD.



The SRD is a new web application in QualityStage. It is designed to aid in the enhancement of standardization rule sets. These enhanced rule sets can then be applied to the Standardization stage.

You can enhance a rule set by adding and modifying classifications, lookup tables, and rules. The SRD can enhance user-defined or custom rule sets and the built-in standardization rule sets, such as USNAME and USADDR.

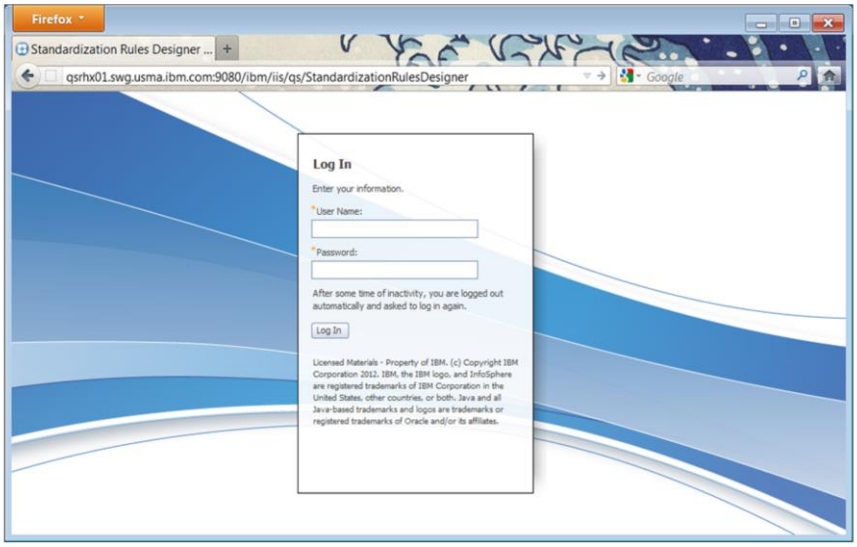
The SRD also has the ability to import sample data in a flat file. This provides a representative sample to work with during the enhancement process.

There are four main screens in SRD. These main screens are called work areas, which are the areas separated by navigation tabs in the user interface. These work areas are Home, Classifications, Lookup Tables, and Rules.

There are three different ways to launch the SRD. The SRD can be launched directly through a web browser using a uniform resource locator or URL. It can also be launched indirectly from the QualityStage Rules Management dialog within the DataStage®/QualityStage Designer Client. Additionally, SRD can be opened from the IBM InfoSphere Information Server Launchpad.

Launching SRD from web browser

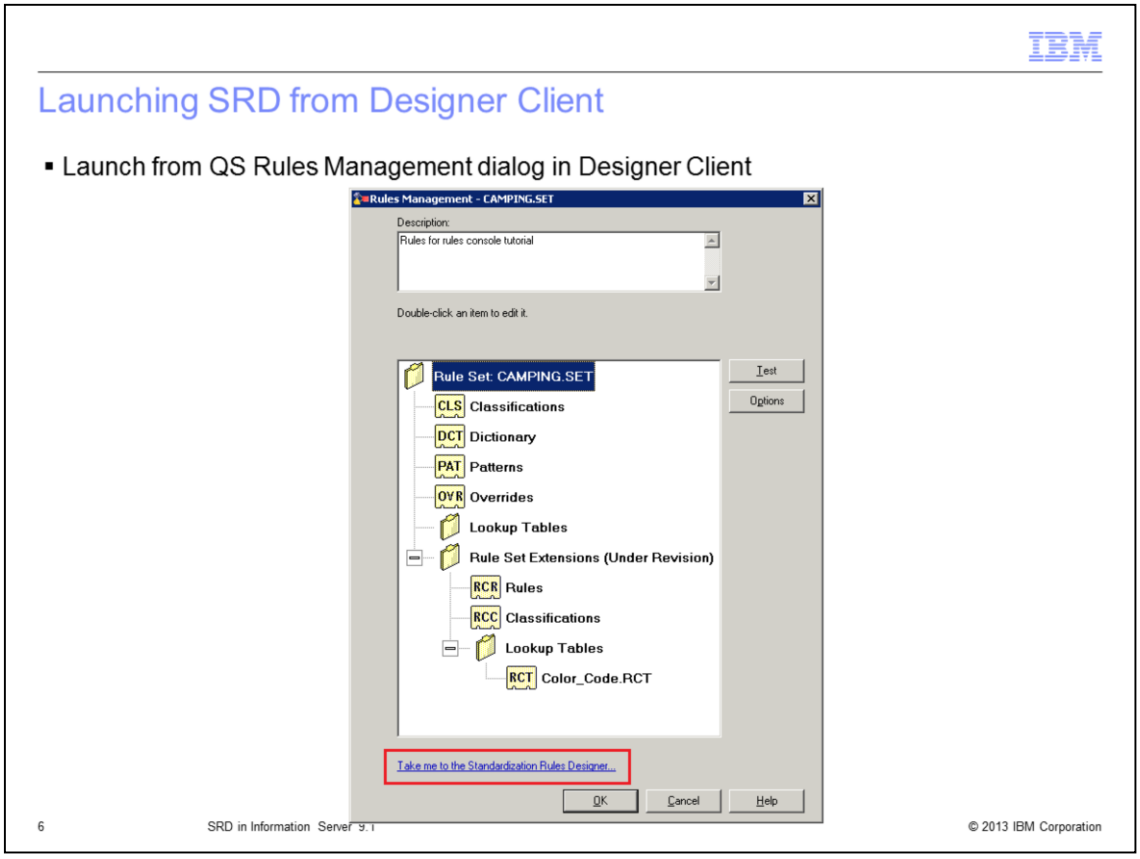
- Type URL in web browser or double click shortcut icon on client machine's desktop
- Login screen opens



5 SRD in Information Server 9.1 © 2013 IBM Corporation

If you are launching SRD from a web browser, launch your web browser, which can be Internet Explorer, Firefox or Chrome. Type in the URL for the SRD. The URL format for SRD is: http://<server_name>:<port_number>/ibm/iis/qs/StandardizationRulesDesigner. The default port is 9080. You must have the DataStage/QualityStage Administrator role or the DataStage/QualityStage User role to log in.

The inactive session is terminated after 30 minutes, after which you will automatically be brought back to the login screen, as displayed on this slide.



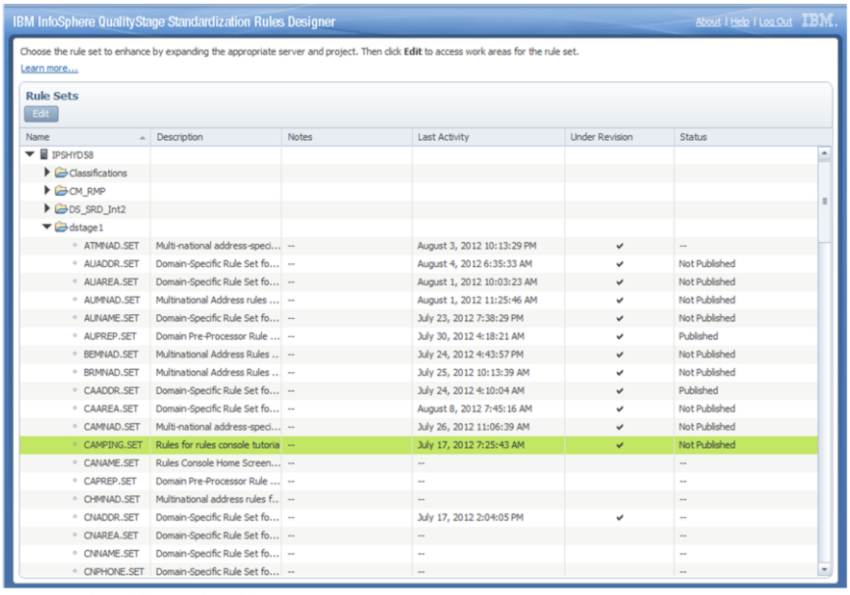
Launch SRD from the Designer Client by clicking “Take me to the Standardization Rules Designer” link at the bottom of the Rule Management dialog. This slide displays the link enclosed in a red box. Doing this will take you directly to the Home screen.

If you click “RCR”, you are taken directly to the Rules Work Area. If you click “RCC”, you are taken directly to the Classifications Work Area.

If you click “RCT”, you are taken directly to the Lookup Tables Work Area.

Rule Set Selection screen

- Displays rule sets under servers and DataStage Projects
- Allows user to start working with rule set by creating a revision



© 2013 IBM Corporation

The Rule Set Selection screen is the first screen displayed after you have logged in by way of the login screen, or have launched SRD from the Designer Client.

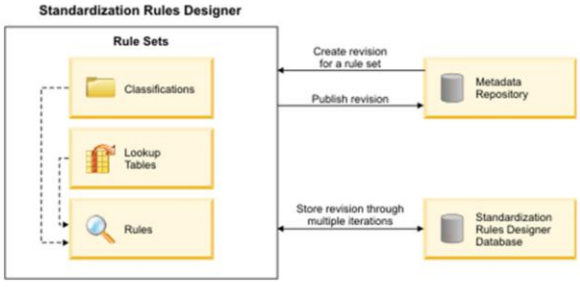
The rule sets and their related information are displayed in a table. The rule sets are under DataStage project nodes and server nodes. Project nodes and Server nodes are expandable items. For example, CAMPING.SET, which is the selected rule set in this screen capture, is a rule set under the dstage1 project which is under IPSHYD58 server.

You can choose the rule set to enhance by expanding the appropriate server and project. Then, click Edit to access work areas for the rule set. If the rule set is not yet under the revision, clicking Edit will create the revision. If the rule set is already under the revision, clicking Edit will re-open the existing revision.

"Revision" is discussed in detail on the next slide.

Rule Set revision

- A way to save and revert changes to rule set
- User creates a revision
- Changes are saved in SRD database
- Can roll back changes by resetting revision or publish changes by publishing revision
- Only published changes can be used in Standarize stage



8 SRD in Information Server 9.1 © 2013 IBM Corporation

Rule sets are stored in the metadata repository. When you create a revision in the SRD, a copy of the rule set is created in a database for the SRD.

The SRD database is created automatically during the Information Server installation if the metadata tier is selected during the installer's interview screen. If the metadata tier is not selected during the installation, then there is a different interview screen to specify the database registration, which is used to generate a script to register the database outside of the installer.

To begin, create a revision for a rule set by editing that rule set in the SRD. Next, the rule set is enhanced through multiple iterations. You can keep a revision for as long as you want. The changes that you make in the SRD affects only the copy of the rule set in the SRD database. You can change the rule set yourself or collaborate with other users. If you collaborate with other users, only one user can open the revision at a time.

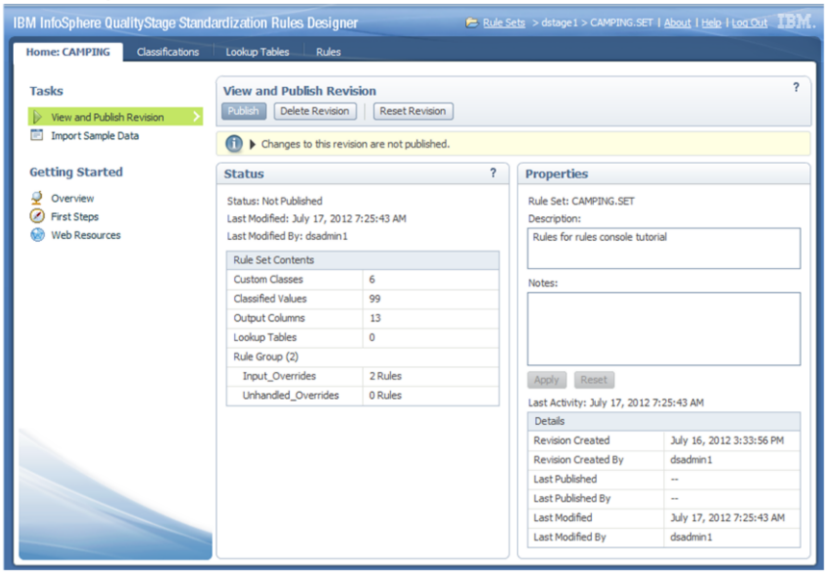
To discard changes that you made in the SRD, reset the revision. When you reset the revision, the revision remains open and is updated with the version of the rule set in the metadata repository.

When you reach a particular milestone, you publish the revision. Publishing the revision updates the rule set in the metadata repository by applying the changes that you made in the SRD. When you publish a revision, you can delete the revision or keep the revision and continue to enhance the rule set.

When your enhancements are complete, delete the revision. Deleting the revision discards the copy of the rule set that is stored in the SRD database. You can begin the life-cycle again by creating a new revision for the rule set.

Home work area (1 of 3)

- View and Publish Revision panel displays current status of rule sets
- Revision can be published/reset/deleted



9 SRD in Information Server 9.1 © 2013 IBM Corporation

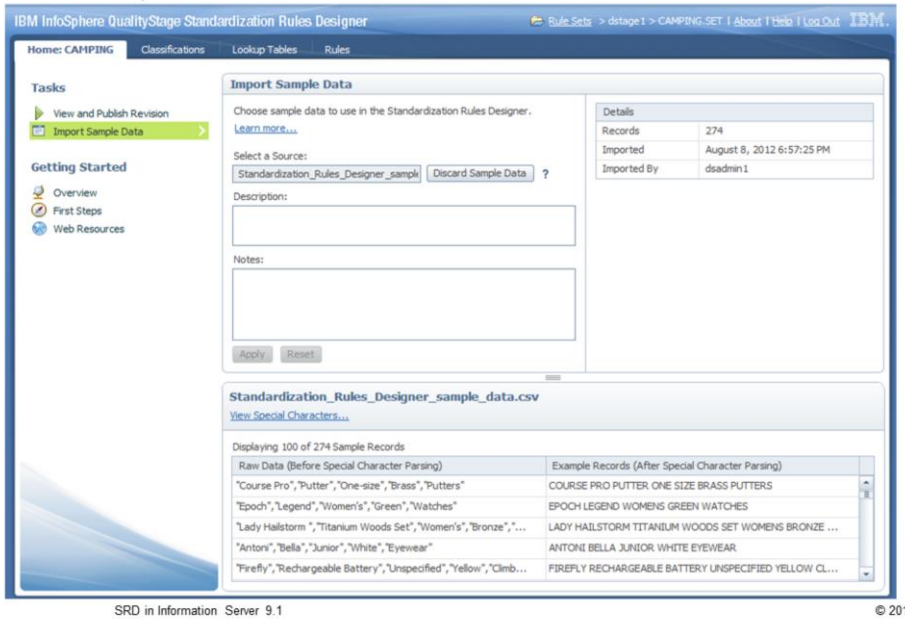
A revision is a collection of changes that were made to the rule set in the SRD. Review the status and properties for the current revision, then determine what you want to do with the revision.

Click Publish to update the rule set in the metadata repository by applying the changes that you made in the SRD. After you publish a revision, you can provision the updated rule set and use that rule set in a job.

When you publish a revision, you choose to keep or delete the revision in the SRD. Click Delete Revision to discard sample data and any changes in the revision that has not been published. Click Reset Revision to discard any changes in the revision that have not been published. When you reset the revision, the revision stays open and the sample data is kept.

Home work area (2 of 3)

- Import Sample Data panel Sample
- Data can be imported and discarded



10 SRD in Information Server 9.1 © 2013 IBM Corporation

You can import sample data from a flat file to work with representative examples with the Import Sample Data Panel. Files that contain sample data must meet specific requirements. Files must be plain text files. Files must have only one record per line and files must use UTF-8 character encoding.

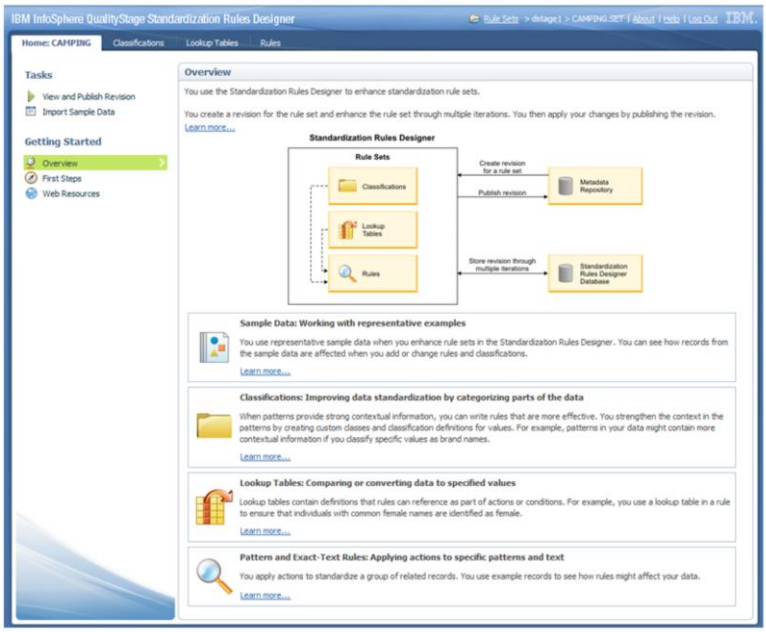
Imported sample data is processed in different ways. Each line in the file is processed as a one column record. Only the first 4097 characters and 750 values in a record are processed, and only the first 600 characters in a value are processed.

To start the file import, click the “Import...” button next to the “Select a source” text field. Select a file from the dialog, then click the Open button. The first 100 records are populated in the table at the bottom of the screen, and the “Import...” button is changed to “Discard Sample Data”. Only one sample data file can be imported per rule set. To change to another sample data file, click “Discard Sample Data”, click the Yes button, then import another file.

If the sample data is not imported, SRD can generate sample data for you for each pattern that you select in the Rules Browser inside the Rules work area.

Home work area (3 of 3)

- Overview, First Steps, and Web Resource panels
  - Provide links to user guide topics, tutorial, QS user developerWorks® forum, and QS product support website

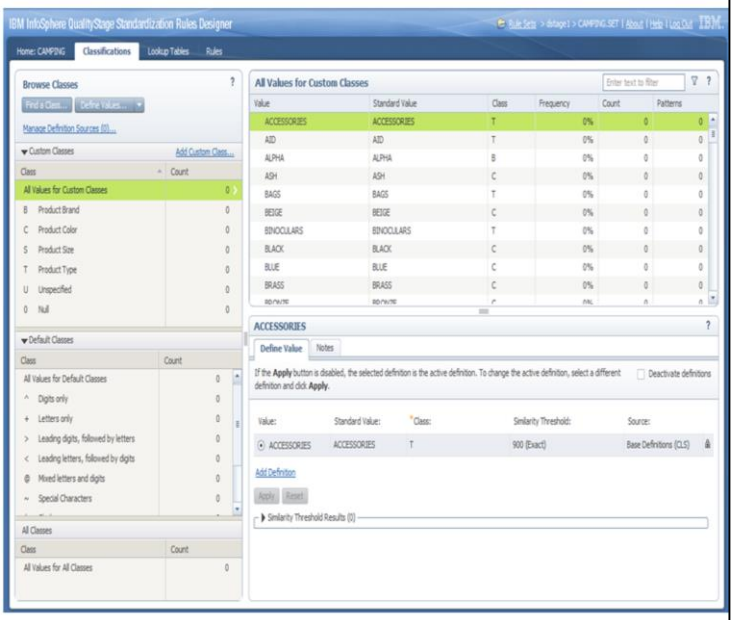


© 2013 IBM Corporation

The screen capture displayed on this slide shows the Overview panel which contains a diagram of SRD workflow, the links to the user guide topics about the rule set, the sample data, classifications, lookup tables, and rules. Clicking any of the links will open another web browser window or a new web browser tab (depending on your web browser's settings) and re-direct you to the appropriate user guide topic. Details are discussed on a later slide.

Classifications work area (1 of 2)

- View values assigned to each custom class and each default class
- Can add custom classes and specific classification definitions
- Can add classification definitions manually by using Define Values dialog
- Can import classification definitions from a source by using Import Definitions dialog
- If more than one definition specified for a value
  - Choose which definition is active



12 SRD in Information Server 9.1 © 2013 IBM Corporation

Patterns provide contextual information. You can strengthen this contextual information by using custom classifications to categorize values that share common attributes. For example, if a record in the sample data contains the word ALPHA (the third value in the values table), then this word is represented inside a pattern as a letter B, because the word ALPHA is one of the class values in custom class B, The Product Brand.

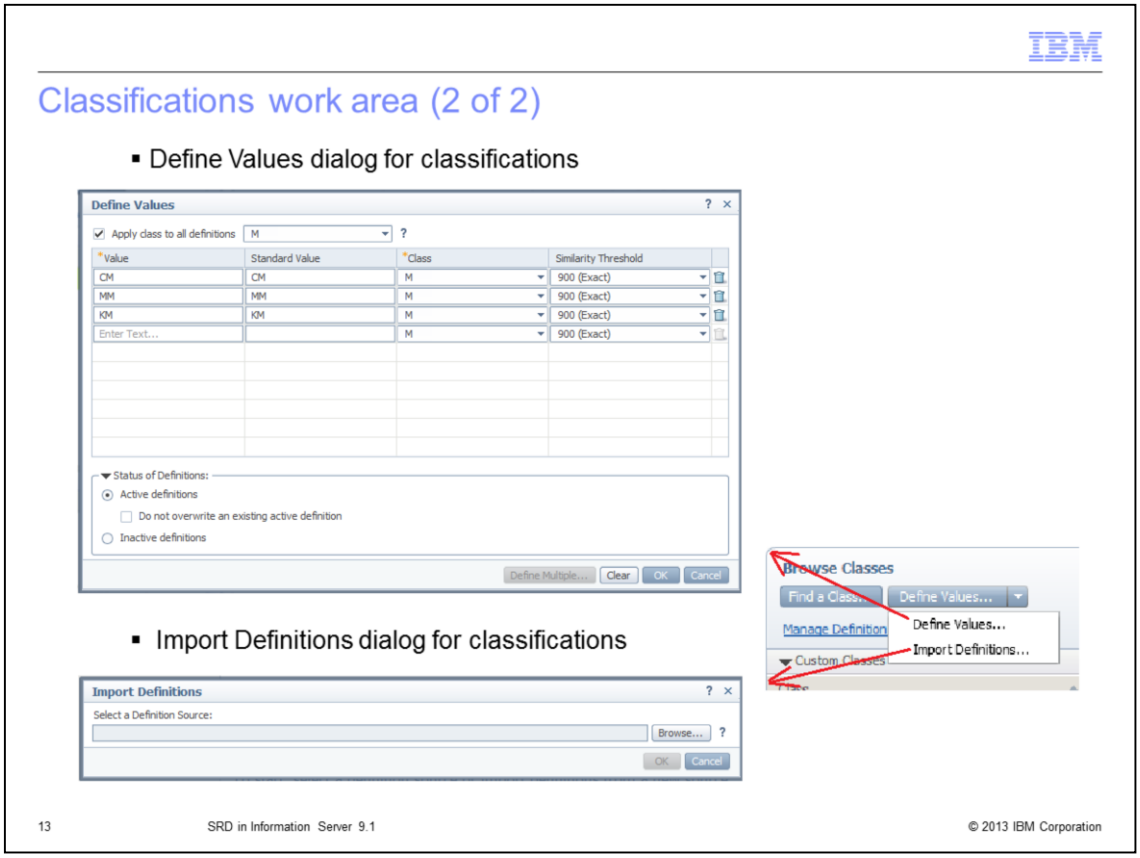
You create classification definitions, import definitions from a source, and add custom classes. The count indicates the number of values in the sample data that are assigned to the class. Values that appear more than once are counted each time that they appear.

The standard value is a standardized spelling or representation of the value that can be used as part of an action or condition in a rule.

The frequency is the percentage of the records in the sample data that contain the value. The count is the number of times that the value appears in the sample data. The patterns are the number of distinct patterns in the sample data that contain this value.

In the Define Value panel, you add classification definitions, choose the active definition, and view the other values that are affected by the currently selected definition.

The similarity threshold is the degree of variation that is allowed in the spelling or representation of the value. Choose lower similarity thresholds to allow more variations affected by this definition.

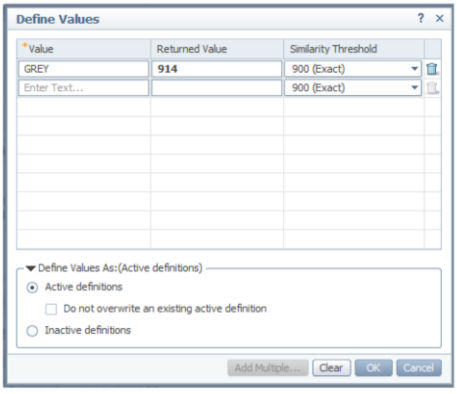


This slide displays screen captures of the Define Values dialog and the Import Definitions dialog. Both dialogs can be invoked from the pop-up menu in the Browse Classes area as displayed in the screen capture on the right side of the slide.

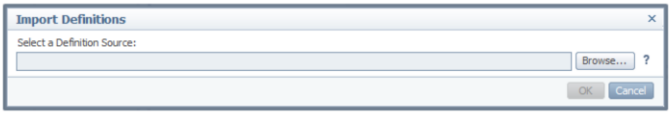
You can add classification definitions manually by using Define Values dialog, or import classification definitions from a source by using Import Definitions dialog. If more than one definition is specified for a value, you can choose which definition is active.

Lookup Tables work area

- Define Values dialog for Lookup Tables



- Import Definitions dialog for Lookup Tables



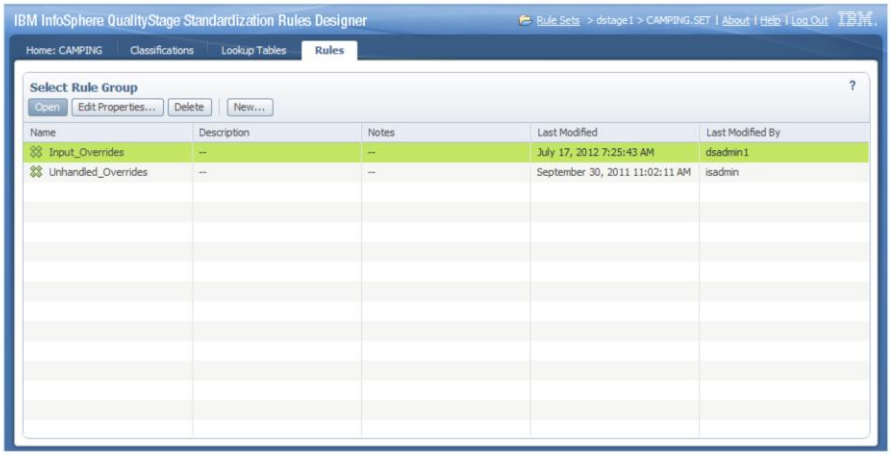
© 2013 IBM Corporation

This slide displays screen captures of the Define Values dialog and Import Definitions dialog. Both dialogs can be invoked from the pop-up menu in the Lookup Tables Values area, next to Browse Lookup Tables panel.

You can add definitions by using Define Values dialog, or import definitions from a source by using Import Definitions dialog. If more than one definition is specified for a value, you can choose which definition is active.

Rules work area (1 of 6)

- Rule Group Selection screen
  - Can add and remove rule groups
  - Can edit properties
  - Can open a rule group, which brings you to Rules work area

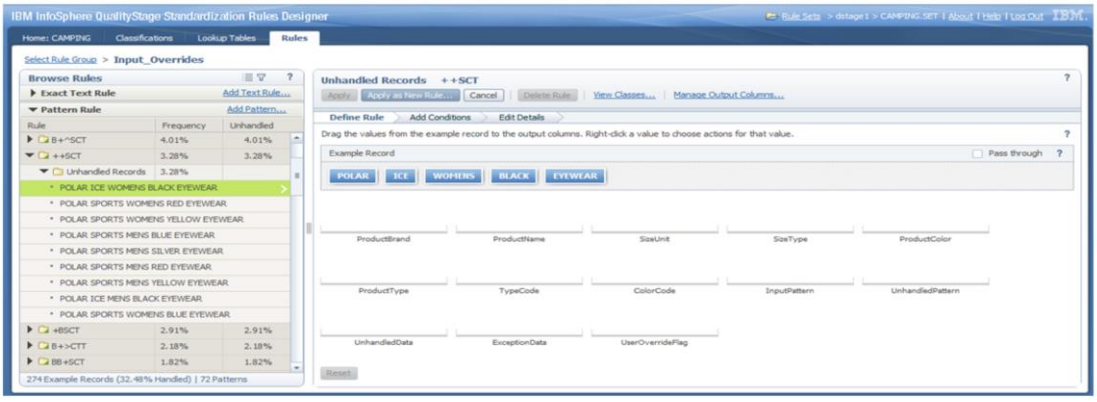


© 2013 IBM Corporation

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, organize the rules into rule groups in the SRD. Then, invoke the rule groups from the pattern-action specification, or PAT. In the SRD, you can also edit rule groups or open a rule group to view the rules in that group.

Rules work area (2 of 6)

- Browse Rules panel
  - Can add pattern rules and exact text rules to rule group
  - Can view records handled by a rule and records that are not handled
- Define Rule panel
  - Can add one or more actions that specify how a rule processes a value
- Add Conditions panel
  - Can add one or more conditions that values must meet before actions in rule are applied to record



© 2013 IBM Corporation

In the Browse Rules panel, you can add pattern rules and exact text rules to a rule group. For each pattern, you can view the records that are handled by a rule and records that are not handled.

Pattern rules apply to records that match a specified pattern. For each pattern, rules are listed in the order that they are applied. Exact text rules apply to records that consist of the text string that is specified in the rule.

The frequency is the percentage of records in the sample data that match a pattern or are handled by a rule.

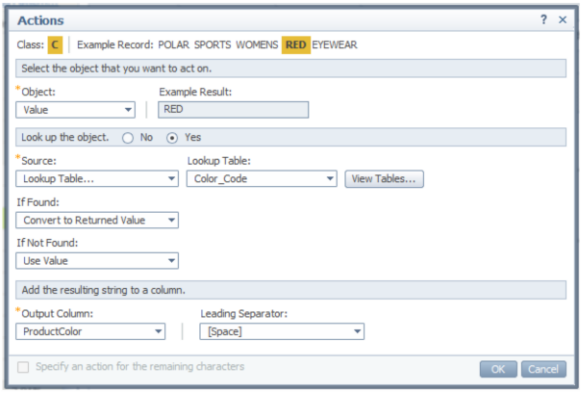
The Unhandled column shows the percentage of records in the sample data that match a pattern and are not handled by a rule.

In the Define Rule panel, a sample record is broken down into token values. You can add one or more actions that specify how a rule processes a value.

To specify an output column for a token value, for example, POLAR, the token value can be dragged from the Example Record area (the tiles area) into an output column bucket, for example ProductBrand, in the buckets area.

Rules work area (3 of 6)

- Actions Dialog
  - Specify how rule processes value
  - If action does not process whole value, can select Specify an action for the remaining characters to add separate action for characters not affected by action



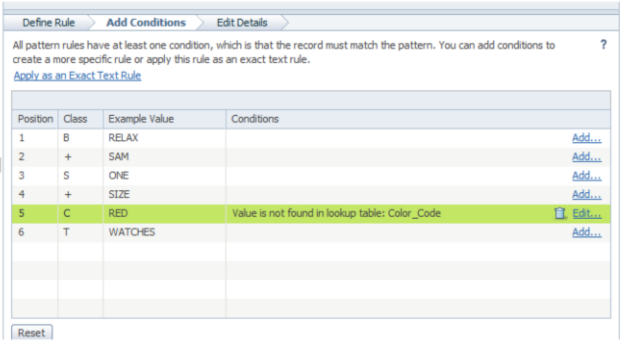
17 SRD in Information Server 9.1 © 2013 IBM Corporation

The Actions dialog allows a complete way for specifying actions to a token value. There are two ways to open the Actions dialog. One way is to create a new action to a token value by right clicking a token tile in the Example Record section inside the Define Rule panel. Then, select the New Action... menu item from the pop-up menu. A new instance of the Actions dialog is presented. Another way to open the Actions dialog is to edit an existing action for a token value. Right click a token tile in the bucketed area. A bucket represents an output column that belongs to an action. Select Edit Action... from the context menu. The Actions dialog is displayed with the previously defined actions already populated.

There are three sections in the Actions dialog. The first section allows the text for the token value to be altered explicitly, by using options like standard value, prefix, suffix, substring, and literal. The second section allows the text for the token value to be converted depending on whether the token value is found inside a lookup table or inside a user-define list of values. The third section allows an output column and leading separator character to be specified to a token value. Selecting an output column from the dropdown list produces the same effect as dragging and dropping a token value inside an output column in the Define Rule panel.

Rules work area (4 of 6)

- Add Conditions panel
  - Define conditions records must meet before actions in rule are applied to record



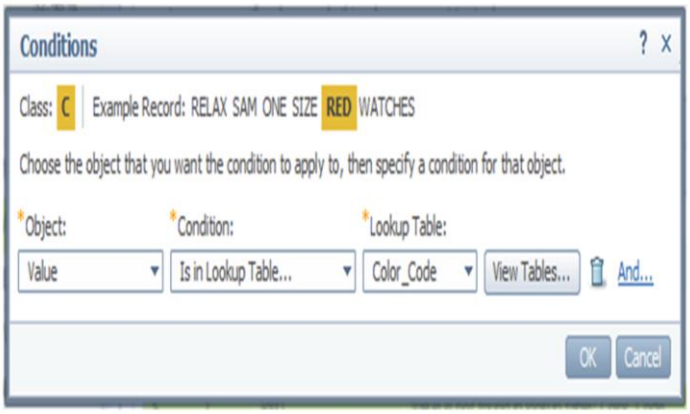
© 2013 IBM Corporation

In the Add Conditions panel, you can add one or more conditions that values must meet before the actions in a rule are applied to the record. Clicking the Add... button will bring up the Conditions dialog which allows you to define one or more conditions to a value.

Clicking the Edit... button brings up the Conditions dialog with the previously defined conditions, associated with the selected value, already populated inside the dialog.

Rules work area (5 of 6)

- Conditions dialog
- After defining condition for value, click And to add another condition for same value

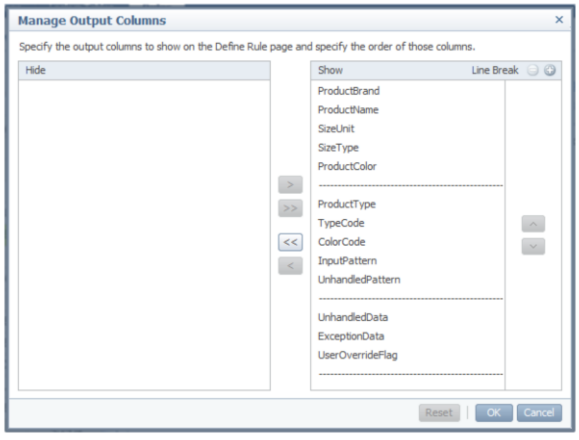


© 2013 IBM Corporation

Within the Conditions dialog, new conditions can be added by pressing the And... link button. The existing condition can be deleted by pressing the Trash Can button.

Rules work area (6 of 6)

- Manage Output Columns dialog
  - Specify output columns to show on Define Rule panel and specify order of those columns
  - Line breaks can be added/deleted/moved to organize output columns by moving a group of output columns into a new line



© 2013 IBM Corporation

This dialog can be opened by clicking the “Manage Output Columns...” link button that is displayed on slide 16.

To hide an output column, select its name inside the “Show” list box, then click the left arrow (<) button. To move all available output columns to the Hide list, press the double left arrow (<<) button.

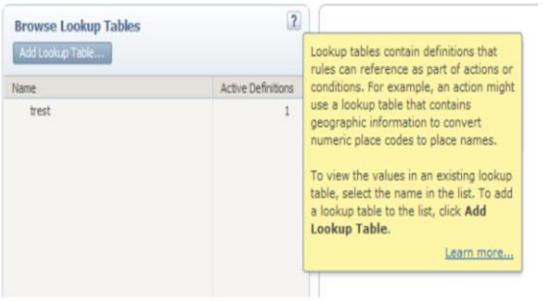
To show the hidden output column, select its name inside the “Hide” list box, then click the right arrow (>) button. To move all available output columns to the Show list, press the double right arrow (>>) button.

To specify the order of an output column that appears in the Define Rule panel, select an output column, then press the Up or Down arrow button to move it to the place that you want.

Line breaks can be added by pressing the Plus (+) button, and removed by pressing the Minus (-) button. The position of a line break can be specified by moving it up or down by using the same Up and Down arrow buttons.

Embedded assistance

- SRD features embedded assistance
  - Introduce novice users to relevant concepts
  - Define key terminology
  - Foreground potential issues to prevent errors



© 2013 IBM Corporation

The SRD features embedded provide assistance that is based on progressive disclosure principles. Progressive disclosure is an interaction design technique often used in human computer interaction to help maintain the focus of a user's attention by reducing clutter, confusion, and cognitive workload. This improves usability by presenting only the minimum data required for the task at hand.

To invoke embedded assistance, click the question mark (?) icon. The pop-up help window will appear. If the cursor still hovers above the ? icon, then the help window will stay open. If the cursor is moved outside of the ? icon, then the help window will stay open for four seconds and will disappear after that. The help window can be closed immediately by clicking anywhere on the screen outside of the ? icon.

Trademarks, disclaimer, and copyright information

THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. Other product and service names might be trademarks of IBM or other companies. A current list of other IBM trademarks is available on the web at "Copyright and trademark information" at http://www.ibm.com/legal/copytrade.shtml

THE INFORMATION CONTAINED IN THIS PRESENTATION IS IBM CONFIDENTIAL AND IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. WHILE EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, IT IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. IN ADDITION, THIS INFORMATION IS BASED ON IBM'S CURRENT PRODUCT PLANS AND STRATEGY, WHICH ARE SUBJECT TO CHANGE BY IBM WITHOUT NOTICE. IBM SHALL NOT BE RESPONSIBLE FOR ANY DAMAGES ARISING OUT OF THE USE OF, OR OTHERWISE RELATED TO, THIS PRESENTATION OR ANY OTHER DOCUMENTATION. NOTHING CONTAINED IN THIS PRESENTATION IS INTENDED TO, NOR SHALL HAVE THE EFFECT OF, CREATING ANY WARRANTIES OR REPRESENTATIONS FROM IBM (OR ITS SUPPLIERS OR LICENSORS), OR ALTERING THE TERMS AND CONDITIONS OF ANY AGREEMENT OR LICENSE GOVERNING THE USE OF IBM PRODUCTS OR SOFTWARE.

© Copyright International Business Machines Corporation 2013. All rights reserved.

© 2013 IBM Corporation