# 

Cleanse and Match Master Data by Using EIM

SQL Server Technical Article

**Writer:** Jim van de Erve

**Technical Reviewers:** Matthew Roche, Matt Masson, Gadi Peleg, Carla Sabotta

**Published:** 10/2012

**Applies to:** SQL Server 2012

**Summary:** This white paper describes how to cleanse and match master data using the Enterprise Information Management products of Microsoft SQL Server 2012 Master Data Services, SQL Server Data Quality Services, and SQL Server Integration Services. The white paper addresses scenarios for building and cleansing a new master data entity, updating one, and performing matching. It also describes some of the differences between Master Data Services and Data Quality Services.

Copyright

This document is provided “as-is”. Information and views expressed in this document, including URL and other Internet Web site references, may change without notice. You bear the risk of using it.

Some examples depicted herein are provided for illustration only and are fictitious.  No real association or connection is intended or should be inferred.

This document does not provide you with any legal rights to any intellectual property in any Microsoft product. You may copy and use this document for your internal, reference purposes.

© 2012 Microsoft. All rights reserved.

Cleanse and Match Master Data by Using EIM

# Master Data Services, Data Quality Services, and Integration Services: Better Together

SQL Server 2012 Master Data Services and Data Quality Services help ensure the integrity of a company's data. You store and manage master data using Master Data Services. You cleanse and match data using Data Quality Services. You can use them independently, maintaining master data without cleansing and matching it, or cleansing and matching source data without maintaining master data. However, Master Data Services and Data Quality Services have complementary feature sets and are best used in partnership. You can use the interactive and computer-assisted processes in both Master Data Services and Data Quality Services to create and maintain master data of high quality. And you can use them together with SQL Server Integration Services in an Enterprise Information Management (EIM) solution that runs both the data quality and the master data management processes.

After describing how Master Data Services and Data Quality Services work together, this white paper presents four scenarios in which you can cleanse and match master data using these EIM components:

* Create and build a new master data model and an entity within the model that will contain the data for a field, using Master Data Services and Data Quality Services
* Update an existing master data entity in an EIM automated lookup scenario using Integration Services, Master Data Services, and Data Quality Services
* Include checks for close matches when updating an existing master data entity, by adding fuzzy matching to the EIM automated update scenario
* Perform matching on an existing entity from within the Master Data Services Add-in for Excel.

Tutorials that provide step-by-step procedures for cleansing master data can be found in the [Tutorial: Enterprise Information Management Using SSIS, MDS, and DQS Together](http://go.microsoft.com/fwlink/?LinkId=269695) whitepaper.

## How Master Data Services and Data Quality Services work together

To describe how these features work together, here is a short description of what each of them does:

**Master Data Services** is a master data management platform that enables you to create and maintain a single, authoritative standard for your business data. Using MDS, you can maintain corrected data in a "known good" state through business rules and an Excel-based UI for managing master data. By definition, master data is non-transactional reference data that changes slowly. As such, master data empowers you to synchronize and resolve discrepancies between source data systems. Master Data Services is where you store and manage master data. From there, it is available to data stewards, information workers, and IT professionals. For more high-level information about Master Data Services, see [Master Data Services Overview](http://msdn.microsoft.com/en-us/library/ff487003.aspx).

**Data Quality Services** enables you to cleanse, enrich, and match data. DQS allows you to improve the quality of your data by tracking rules and other knowledge about how to correct problems in "dirty" data, and to apply that knowledge in interactive and automated scenarios. Data Quality Services is a knowledge-based system that performs both computer-assisted and interactive cleansing and matching processes using a knowledge base that you create. You can build a knowledge base by adding your knowledge of the data, performing computer-assisted knowledge discovery, and adding matching rules, which leveraging Data Quality Services algorithms and industry reference data that are built into each knowledge base. You can also build the knowledge gained in ongoing cleansing operations into the knowledge base, to achieve and maintain a high standard of data quality. This results in a virtuous cycle of ongoing improvement: the more you use a Data Quality Services knowledge base for cleansing, the better it works for future cleansing tasks. For more high-level information about Data Quality Services, see [Introducing Data Quality Services](http://msdn.microsoft.com/en-us/library/ff877917.aspx).

Master Data Services provides tools and processes to maintain the integrity of the master data that is stored in a Master Data Services model. You can enter changes to master data in the MDM UI, or through the MDS add-in for Excel, and you can create business rules that validate incoming master data. However, the focus in Data Quality Services on cleansing and matching data gives you much more data cleansing power than the native features in Master Data Services do. Master Data Services and Data Quality Services used together provide you with the technology that when combined with people and processes deliver a full data curation solution. The automated and interactive processes of both products empower the information worker to make the decisions that result in clean data.

Master Data Services and Data Quality Services give you multiple ways to perform cleansing and matching operations. You can perform knowledge base management, cleansing, and matching directly in the Data Quality Client. You can perform cleansing as part of an unattended batch operation by using the DQS Cleansing, Lookup, and Fuzzy Lookup transformations in an Integration Services data flow. You can perform matching by using the matching functionality in the Master Data Services Add-in for Excel, which is based on Data Quality Services. These options give you the flexibility to choose the process that fits your needs most closely.

To help you understand how Master Data Services and Data Quality Services work together to ensure high-quality master data, this paper looks at four common implementations of these tools: one for entering new master data in a new Master Data Services entity, and three for updating existing master data as new data comes in.

# Build a new master data entity by using Master Data Services and Data Quality Services

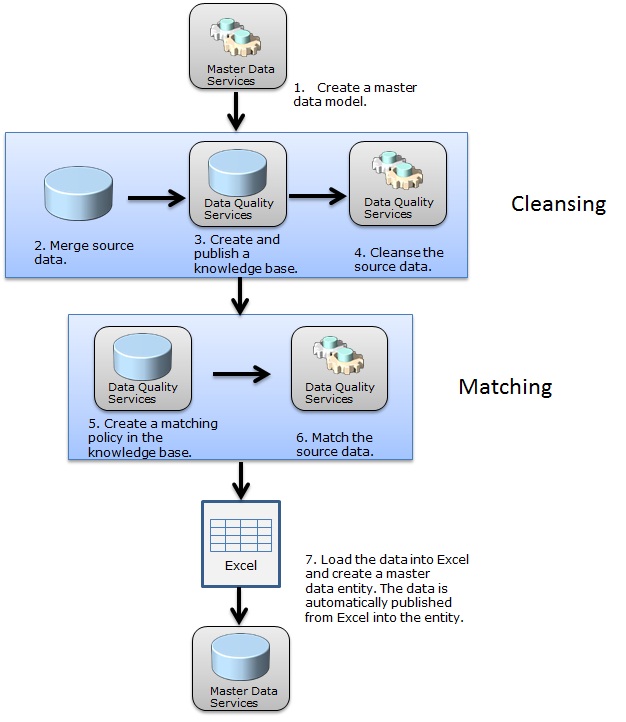
When you first create a model and an entity in Master Data Services, and populate it with master data, you want the data to be clean. Using Data Quality Services, you can create a knowledge base that is rich with cleansing knowledge and matching rules, and then run cleansing and matching projects. These steps are the best pathway to clean data. You make data decisions interactively in the Data Quality client, using the computer-assisted processes that Data Quality Services provides to aid those decisions.

When you're ensuring the quality of master data, we recommend that you cleanse all your source data before you do matching, and that you perform matching on the full set of source data. If you find and correct as many problems with as much source data as you can, getting it as standard and consistent in format as possible, you will increase your odds of finding genuine matching records. When working with pre-cleansed data, the matching process will yield fewer false positives and fewer false negatives.

After you run cleansing and matching, you can then import the data into the Master Data Manager (MDM), and going forward, use the Data Quality Services knowledge base by updating the data through automated processes. Cleansing your data before you load it into Master Data Services enables you to be fully aware of, and have full confidence in, your master data. After master data is stored inside Master Data Services, you have more control over it. You can put Master Data Services business rules in place to ensure that added or edited data conforms to your rules. You can import new data, interactively using the MDS add-in for Excel or as part of a batch process using SSIS. Similarly, you can export the master data using SSIS, or view and share master data using Excel. You can manually review and modify the master data. Master Data Services is where you manage your master data going forward so your source data systems can reliably depend upon it.

## How building a new master data entity works

The following scenario creates and populates an entity in a master data model, uses Data Quality Services to cleanse and match source data, and uses Master Data Services to publish it as master data.



To create and populate a new master data entity, and cleanse and match the master data, follow these steps:

1. In Master Data Services, create a model for the master data. You could at this point create an entity within the model that will contain master data members, but in this example scenario the entity is created later in the Master Data Services Add-in for Excel.
2. Merge source data in an Excel spreadsheet or a SQL Server table, or using an Integration Services package. If you have a lot of data to add, it's recommended that you do so in stages for better performance. Cleanse and publish a subset of your entire data, and then repeat the process for another subset. You do not have to merge all your data at once; the knowledge discovery process can be performed iteratively with multiple sources, one after the other.
3. In the Data Quality client, create and publish a knowledge base for the data to be added as master data, and create in the knowledge base one or more data quality domains each of which will represent the data in a field. To do so, you can run knowledge discovery on a sample of the data; add domain rules, term-based relations, and a composite domain; and add knowledge manually to the [knowledge base](http://msdn.microsoft.com/en-us/library/gg524799.aspx). Cover all of the data domains. Ensure that your knowledge base covers all of the source data by running knowledge discovery on as many samples as possible, importing corrections made during a cleansing process, and adding data values as needed. Your domain should contain the same values you have in your source data. For more information, see [DQS Knowledge Bases and Domains](http://msdn.microsoft.com/en-us/library/gg524799.aspx).
4. Create a [cleansing](http://msdn.microsoft.com/en-us/library/gg524800.aspx) data quality project in the Data Quality client, and then run the data quality project on the data that you want to add to the MDM. Changes with a confidence rating above the auto-suggestion threshold, but below the auto-correction threshold, are suggested by the system, but not automatically made. (Note that a dqs\_administrator can change either of these thresholds.) During the interactive cleansing process, you can manually approve or reject those suggested changes in the output, and you can manually override auto-corrected changes. You can then export the cleansed data to an Excel or .csv file or to a SQL Server table. The example scenario covered by these steps creates an Excel output file. For more information, see [Data Cleansing](http://msdn.microsoft.com/en-us/library/gg524800.aspx).
5. In the Data Quality client, create a [matching](http://msdn.microsoft.com/en-us/library/hh270290.aspx) policy in the knowledge base. This involves adding matching rules in an iterative testing process of establishing weighting, similarity, and prerequisites. For more information, see [Create a Matching Policy](http://msdn.microsoft.com/en-us/library/hh270290.aspx).
6. Run a matching [project](http://msdn.microsoft.com/en-us/library/hh270289.aspx) on the cleansed data. You can review the results and the profiling data, approve or reject the matches, and specify a survivorship rule. At the end of the process, Data Quality Services deletes records that match a survivor record, and then it exports the resulting data to a file. Note that you will not have control over which is designated as the leading record in a given cluster, but you can designate which record becomes the survivor record. For more information, see [Run a Matching Project](http://msdn.microsoft.com/en-us/library/hh270289.aspx).
7. Open the cleansed and matched data in Excel with the Master Data Services [Add-in](http://msdn.microsoft.com/en-us/library/hh231024.aspx) for Excel. In the Excel Add-in, create an entity in the master data model created in step 1. When the entity is created, the entity and the data in Excel is published to the Master Data Services model.

A tutorial with step-by-step procedures for creating a knowledge base, cleansing data, matching data, and storing supplier data in Master Data Services can be found in lessons 1 through 4 in the [Tutorial: Enterprise Information Management Using SSIS, MDS, and DQS Together](http://go.microsoft.com/fwlink/?LinkId=269695) white paper.

When you perform cleansing and matching in Data Quality Services, the source data that you run them on is not changed. The changes are saved in an Excel file, a CSV file, or a SQL Server table, which you can then import into other applications or systems such as Master Data Services. If you save the data with changes in an Excel file, you can publish the data from the Excel Add-in into the MDM model. If you save the data and changes in a Master Data Services staging table, you can use entity-based staging to import it into the master data entity by running the **stg.udp\_name\_Leaf** stored procedure. After it's in an entity in the Master Data Services model, you can make changes to the master data manually in the Master Data Services Web Application or the Master Data Services Add-in for Excel. (For more information, see [Master Data Manager Web Application](http://msdn.microsoft.com/en-us/library/hh231064.aspx).)

The Master Data Services Add-in for Excel is the easiest way to export data from, or publish data to, the master data manager. However, if your data is very large, or if you will be importing or exporting as part of an ongoing repeatable process, you may want to do this using a SQL Server table. For more information about the differences between using Excel worksheets and database tables, see [Import and Export](#_Import_and_export).

# Update master data by using automation

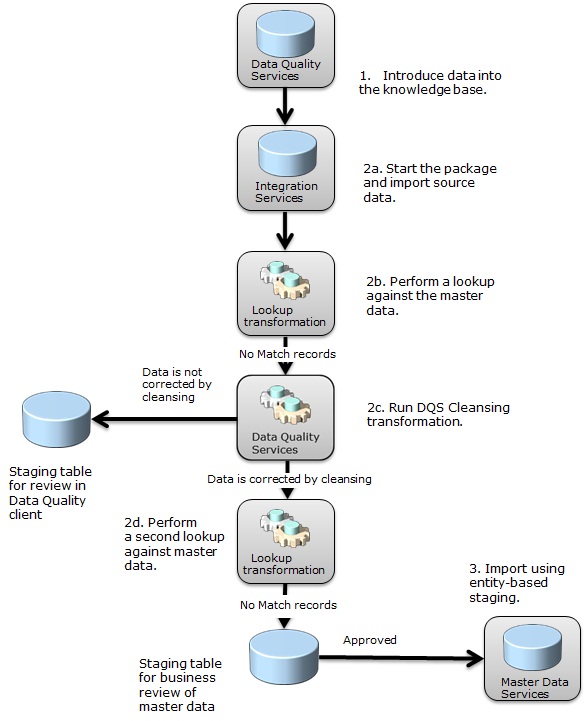
Master data needs to be managed, enhanced, cleansed, and matched over time to ensure that it is kept up to date. After master data has initially been imported into Master Data Services, you can run an Integration Services package to perform ongoing updates of the master data, rather than running the process manually in the Data Quality client. This has the advantage of limiting the interactive processes, while still leveraging the computer-assisted processes of Data Quality Services and the knowledge base.

The scenario in this section describes an Integration Services package that uses the Lookup transformation to compare the incoming updates with the existing master data. If no duplicate records are found, the updated data is cleansed. After cleansing, another lookup is run to verify that the cleansed data does not exist in the master data. The package then adds the data into a staging table for review. You can import the approved data into the Master Data Manager using entity-based staging.

The lookup in the scenario in this section looks for exact matches. It does not generate approximate matches of updates. That will be handled in the next section.

## How updating master data using automation works

The following scenario updates the master data in an entity in Master Data Services. It uses a Lookup transformation in an Integration Services package to ensure that duplicate records are not added to the master data, and a DQS Cleansing transformation that runs an automated cleansing process. The package stores new and unknown values in a separate location, making it easy for the data steward or business user to review the values.



To update master data, follow these steps:

1. Ensure that your knowledge base covers all of the source data. You can enhance your knowledge base with multiple sets of external source data by running knowledge discovery on them.
2. Run an Integration Services package that does the following:
   1. Imports source data, merging from multiple sources if applicable.
   2. Uses a Lookup transformation to do a lookup comparing the new source data to the existing master data values, using a [subscription view](http://msdn.microsoft.com/en-us/library/ee633741.aspx) as a source for the lookups and using a business key attribute that is common between the entity data and the data being imported. A subscription view is a view of the master data in an entity. This process determines which records already exist in the master data, and so do not need to be added. Being able to skip the addition of data that already exists helps improve the performance of the cleansing process. There are variations on this step of the scenario that a real-world implement would need to cover. An example is when incoming data has the same key value, but different values in other attributes, as when a record that was previously loaded into MDS has been updated in a source system so that their values are no longer the same.
   3. Runs the DQS Cleansing transformation to cleanse any records that were not found in the Master Data Manager. When the Integration Services package runs the DQS Cleansing transformation, it automatically creates and runs a Data Quality project.

If data is not corrected by Data Quality Services, the package stores it in a SQL Server database staging table for a review by a data steward in the Data Quality client. The data steward approves or rejects the data, and then can add approved data to the master data. An SSIS developer could redirect corrected rows to one output path and other rows to other output paths.

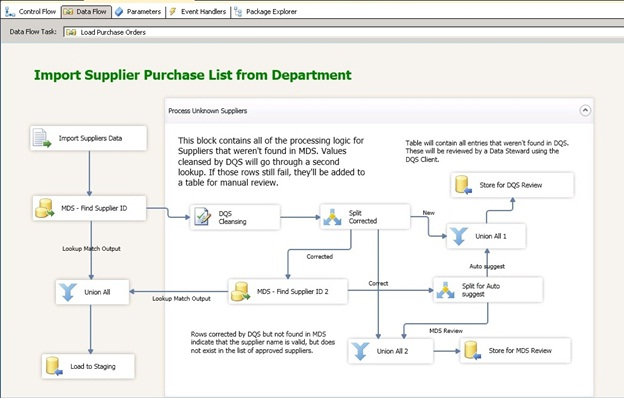
When an SSIS data flow that includes the DQS Cleansing transform is executed, a Data Quality project is automatically created. A data steward can then open the project in the Data Quality client application, review and approve (or reject) suggested corrections, and save these corrections back into the knowledge base as new domain values or synonyms. This “Cleanse to KB” process allows the data steward to be more involved in these automated scenarios.

* 1. If the data is corrected by Data Quality Services, the package runs another lookup transformation. If the updated data is not found in Master Data Services, the package stores it in a SQL Server database staging table for a Master Data Services business data review. You could load the data into Master Data Services at this point using entity-based staging, and have the data steward review it in the Web application.

1. If the business user approves a record, the data steward can add it to the master data entity by performing entity-based staging. They do so by running the **stg.udp\_name\_Leaf** stored procedure (where "name" is the entity name), which imports the data into the entity.

A tutorial with step-by-step procedures for automating the cleansing and matching of data using Integration Services can be found in Lesson 5 in the [Tutorial: Enterprise Information Management Using SSIS, MDS, and DQS Together](http://go.microsoft.com/fwlink/?LinkId=269695) white paper. Note that the scenario in the tutorial differs in some ways from the scenario outlined above.

Here's an example of an Integration Services package that updates master data using the preceding steps. (This implementation was created by Matt Masson of Microsoft.)



This package handles failed cases gracefully. It includes a Conditional split transformation to separate Corrected, Correct, and New values. Human intervention is required only if you need to decide, after the package has run, whether to add the cleansed data to the Master Data Manager based upon your business knowledge, and whether a record needs to be cleansed again interactively.

When you run the DQS Cleansing transformation, you don't have the opportunity to approve or reject a corrected or suggested change generated by the Data Quality project, as you can in the Data Quality client. You can, however, do so after the package has completed. You can review the changes in the output, as is done in the scenario in this section. And as indicated in step 2c, you can open the data quality project created by the execution of the DQS Cleansing transform in the Data Quality client just like any other data quality project, and interactively cleanse the results in the project. You can then drive the knowledge gained by running the project back into the knowledge base.

As the volume of data knowledge inside the knowledge base grows, the amount of user review that needs to take place after the Integration Services package runs may become smaller and smaller. An organization may have hundreds of thousands of records being cleansed by an Integration Services package every day, yet only need a data expert to review just a few records manually.

After you have an Integration Services package that you run to update one entity, you can create additional packages to update other entities. The new packages may require only minor configuration changes from the first package, such as pointing the DQS Cleansing transformation to a different knowledge base.

# Perform matching on updated data

The Integration Services package for updating master data that is shown in the [Update master data by using automation](#_Updating_Master_Data) section earlier includes a Lookup transformation process and data cleansing. The lookup determines whether a record already exists as an exact duplicate in the master data. However, while the Lookup component finds duplicates, it does not discover close matches for each record that is a candidate to be added to master data.

The biggest matching issues may occur at the beginning of a project. For those issues, you can run cleansing and matching, and then import the master data into a Master Data Services entity, as shown in the [Build a new master data entity by using Master Data Services and Data Quality Services](#_Build_a_new) section. However, you may still want to run matching on updates. This section shows you how.

To run matching when updating master data, you have the following options:

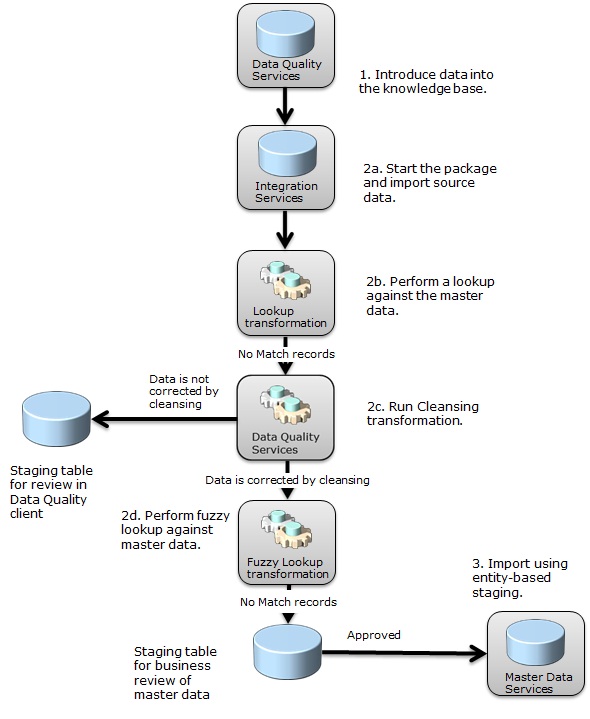
* Add a Fuzzy Lookup transformation to the Integration Services package
* Perform matching manually within the Master Data Services Add-in for Excel.

## Add a Fuzzy Lookup transformation

To add matching capability to the Integration Services package described earlier in the [Update master data using automation](#_Update_master_data) section, you can add a Fuzzy Lookup transformation. This transformation performs a matching step that indicates close matches. Add the Fuzzy Lookup transformation after the first Lookup transformation and the data cleansing transformation. If the record is not found in Master Data Services by the first transformation, the package cleanses the data and then runs a Fuzzy Lookup transformation. If an approximate match is found, the Fuzzy Lookup returns a key and the close matches. The close matches are directed to a SQL Server database table. The data steward can then decide which records to add to the entity.

### How using a Fuzzy Lookup transformation works

The following scenario runs matching in an Integration Services package using the Fuzzy Lookup transformation:



The Fuzzy Lookup transformation differs from the Lookup transformation in its use of fuzzy matching. The Lookup transformation locates duplicate records in a reference table. The Fuzzy Lookup transformation returns records with one or more closely matching records. A Fuzzy Lookup transformation frequently follows a Lookup transformation and a DQS Cleansing transformation in a package data flow. As with the cleansing process, the Lookup transformation helps improve the performance of the Fuzzy Lookup transformation by limiting the number of records that have to be matched.

For more information about the process in the workflow, see [How updating master data using automation works](#_How_Updating_Master).

The Fuzzy Lookup transformation identifies matches using matching algorithms, but does not leverage a knowledge base with a matching policy that you can interactively modify, as Data Quality Services matching does.

A tutorial with step-by-step procedures for automating the cleansing and matching of data using Integration Services can be found in Lesson 5 in the [Tutorial: Enterprise Information Management Using SSIS, MDS, and DQS Together](http://go.microsoft.com/fwlink/?LinkId=269695) white paper. Note that the scenario in the tutorial differs in some ways from the scenario outlined above.

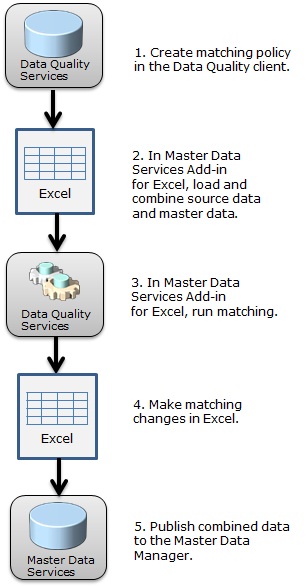
## Run matching from the Master Data Services Add-in for Excel

Master Data Services provides you with the ability to run matching in the Master Data Services Add-in for Excel. It runs matching on a dataset that can contain master data from the Master Data Manager as well as data that is not managed by Master Data Services. This functionality leverages the matching functionality in Data Quality Services, using a matching policy within a Knowledge Base to identify potential matches. For records that it identifies as matches, it adds a cluster ID to the data in the worksheet in Excel, and selects one of the records as the leading record. After matching has completed, you can publish the changes back into the entity. You can choose to delete matched records, leaving the leading record.

Using the matching capabilities in the Master Data Services Add-in for Excel is similar to running a matching project in the DQS client. The user experience is different, and some functionality isn’t implemented in the Excel Add-in, but the core matching functionality is similar. In the Master Data Services Add-in for Excel, you cannot add matching rules and iteratively test and build the matching policy, as you can in the Data Quality client. However, your particular scenario may dictate that you should perform matching in the Excel Add-in. If you are adding updated data in Excel to other data already in a master data model, it may be easier to do the [matching](http://msdn.microsoft.com/en-us/library/hh548681.aspx) inside the Excel Add-in after the data is combined. For more information, see [Data Quality Matching in the MDS Add-in for Excel](http://msdn.microsoft.com/en-us/library/hh548681.aspx).

### How running matching from Excel works

The following scenario combines master data with source data that is not managed by Master Data Services, and then it runs matching on the combined data in the Master Data Services Add-in for Excel:



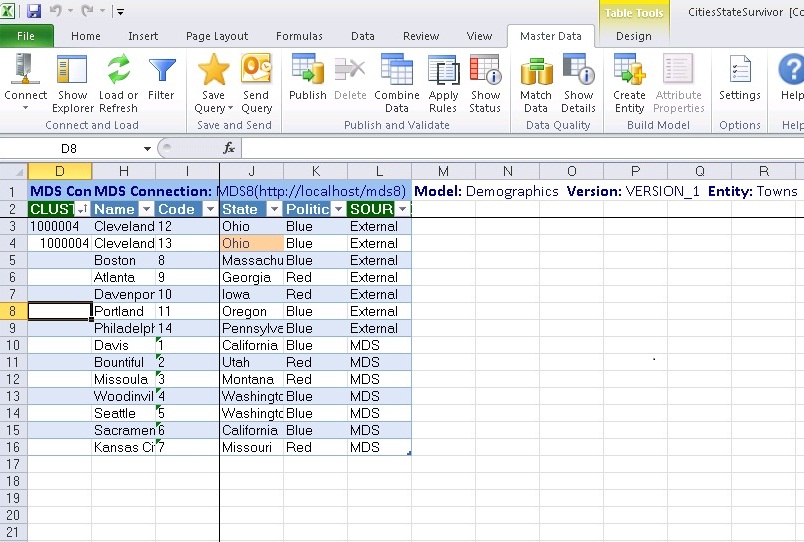
To run matching from Excel, follow these steps:

1. In the Data Quality client, build a matching policy into the knowledge base using the updated source data.
2. Copy source data that is not managed by Master Data Services into an Excel worksheet, and then use the Master Data Services Add-in for Excel to load the master data from the Master Data Manager into a second worksheet in Excel. Combine the data in the two worksheets into a single worksheet using the Combine command in the Excel Add-in.
3. Run matching from the Matching command in the Excel Add-in. A Cluster\_ID column is added to the data, and a Source column indicating MDS (managed by Master Data Services) or External (source data) is also added. You can add those new columns to the master data, and filter on them.
4. Make any matching changes required in the combined data in the Excel file worksheet.
5. Publish the data from the Excel Add-in into the master data entity.

To run the matching process in the Master Data Services Add-in for Excel, you must designate either a knowledge base that was already created or the default knowledge base. If you specify an already-created knowledge base with a matching policy, the matching process will use all the matching rule parameters specified in that policy, including similarity, weight, whether a value is a prerequisite, whether to use overlapping clusters, and so on (see [Create a Matching Policy](http://msdn.microsoft.com/en-us/library/hh270290.aspx)).

If you do not specify a knowledge base that has a relevant matching policy, the Master Data Services Add-in for Excel gives all of the fields the same weight, and uses that as a default matching policy. For example, if you have five fields, each receives a 20 percent weight. You map columns in the worksheet to domains in the knowledge base, and assign a weight value to the domains you choose. However, these weight values do not persist after the matching process is complete.

The following figure shows the results of matching in the Master Data Services Add-in for Excel.



Instead of using Excel, as in this scenario, you could use SQL Server databases to export master data from, and import it into, the master data model. You can cleanse and match master data that already resides in the model by exporting it through a Master Data Services subscription view and consuming the data in the Data Quality client. After you cleanse and match, you can import the updated data back into a staging table, [importing](http://msdn.microsoft.com/en-us/library/ee633726.aspx) it using entity-based staging. Use the ImportType field in the staging table to determine what to do when the staged data matches data that already exists in the entity.

# Compare Master Data Services and Data Quality Services

This section describes some of the ways that Master Data Services and Data Quality Services differ in their processes and some of the ways in which they are similar. Some of their functionality can overlap because they are designed to work together and they are each designed to work without one another.

Both MDS and DQS play important roles in enterprise information management scenarios. DQS emphasizes the cleansing, correction, and standardization of data through the collection of business knowledge in a DQS knowledge base, and the application of that knowledge in interactive and automated cleansing scenarios. MDS emphasizes the ongoing "care and feeding" of master data, providing a set of tools that allow business users to keep records up to date, while allowing IT departments to secure and manage the master data models, while integrating with external systems that utilize the master data.

## Import and export

Both Master Data Services and Data Quality Services have multiple ways to import and export data. They both work with SQL Server and Excel, but they do so in different ways.

You have several choices for importing data into, and exporting data from, the MDM.

* You can [load](http://msdn.microsoft.com/en-us/library/hh479636.aspx) master data from an entity into Excel, and you can [publish](http://msdn.microsoft.com/en-us/library/hh479645.aspx) data into a model from Excel, by using the Master Data Services Add-in for Excel.
* You can also [import](http://msdn.microsoft.com/en-us/library/ee633726.aspx) data from SQL Server staging tables by using entity-based staging.
* You can [export](http://msdn.microsoft.com/en-us/library/ee633741.aspx) master data from the MDM into a subscription view.

Whether to use Excel or SQL Server to import or export master data is likely to depend on who is performing the tasks, the size of the data, and whether the tasks are part of an automated process.

* Information workers are likely to find the Master Data Services Add-in for Excel useful because they are probably familiar with the product and used to performing manual tasks using it. IT pros are likely to find SQL Server processes useful given their skill sets.
* If you have a very large set of data, you are likely to want to use SQL Server processes rather than Excel. At a certain point, a large set of data becomes unmanageable in Excel.
* Entity-based staging and subscription views are easy to automate using SQL Server Integration Services and Transact-SQL. If you need to bulk-load data into an entity, use entity-based staging. Subscription views are used as a source for lookups because any data that goes into the MDM automatically shows up in a subscription view.

As with Master Data Services, Data Quality Services has integration points using both Excel and SQL Server.

* You can [cleanse](http://msdn.microsoft.com/en-us/library/gg524800.aspx) data from SQL Server or Excel source files.
* You can export cleansed data into a SQL Server, CSV, or Excel file.
* You can import domains or values into a domain of a [knowledge base](http://msdn.microsoft.com/en-us/library/hh510391.aspx) by using an Excel file.
* You can import or export a knowledge base or domain by using a .dqs file.
* You can import values from a cleansing project (such as is created when the DQS Cleansing transformation is run) back into the knowledge base used by the project.

## Master Data Services business rules and Data Quality Services domain rules

Business rules in Master Data Services and domain rules in Data Quality Services are both used to maintain the integrity of data, but they are used in different circumstances, as part of different operations, and from within different tools. Domain rules are used to identify data problems as part of a cleansing project, while business rules are used to prevent bad data from being introduced into Master Data Services. The two types of rules can apply to the same data at different places in the information management lifecycle, so the same rule can end up being defined in both places.

**Domain Rules**

A [domain rule](http://msdn.microsoft.com/en-us/library/hh510397.aspx) is used in the cleansing processes performed in Data Quality Services to identify problematic records. Domain rules are established in the domain management performed in the Data Quality client, and they are included in the applicable knowledge base.

A domain rule enables you to specify conditions for the data that you are cleansing. You can add rules to specify values that are not allowed, and what the available values are or what the range of values is. Each record you cleanse is validated against the rules that apply to the domains in the record. If a check of a value fails, the value is marked as invalid.

**Business Rules**

A [business rule](http://msdn.microsoft.com/en-us/library/ff487015.aspx) is used in Master Data Services to prevent the entry of problematic records. Business rules are created within the Master Data Manager Explorer, and they are part of the master data model.

A business rule prevents the improper entry or modification of records. It is not about identifying master data within an entity that is already incorrect. It's about keeping bad data from being added to the MDM or introduced within the MDM. Master Data Services will prevent you from entering a new record or modifying an existing value if doing so would violate a business rule. Multiple processes trigger this validation of master data. Whenever you add or change data into the MDM, whether it's through the Excel Add-in, entity-based staging, the Web UI, or the API, applicable business rules are run on the records that are added or changed. If you are manually making changes in the MDM, you can use a business rule to validate the changes.

## Conclusion

In this white paper, we have seen how you can use Master Data Services, Data Quality Services, and Integration Services to cleanse and match master data, either manually (when building a new master data model and entity) or in an automated process (using Integration Services when updating master data). We have also seen how Master Data Services and Data Quality Services work together, their differences, and their similarities.

Tutorials that providing step-by-step procedures for cleansing master data can be found the [Tutorial: Enterprise Information Management Using SSIS, MDS, and DQS Together](http://go.microsoft.com/fwlink/?LinkId=269695) whitepaper.

**For more information:**

<http://www.microsoft.com/sqlserver/>: SQL Server Web site

<http://technet.microsoft.com/en-us/sqlserver/>: SQL Server TechCenter

<http://msdn.microsoft.com/en-us/sqlserver/>: SQL Server DevCenter

See the following sections in SQL Server 2012 Books Online:

* [Master Data Services](http://msdn.microsoft.com/en-us/library/ee633763.aspx)
* [Data Quality Services](http://msdn.microsoft.com/en-us/library/ff877925.aspx)
* [Integration Services](http://msdn.microsoft.com/en-us/library/ms141026.aspx)

Did this paper help you? Please give us your feedback. Tell us on a scale of 1 (poor) to 5 (excellent), how would you rate this paper and why have you given it this rating? For example:

* Are you rating it high due to having good examples, excellent screen shots, clear writing, or another reason?
* Are you rating it low due to poor examples, fuzzy screen shots, or unclear writing?

This feedback will help us improve the quality of white papers we release.

[Send feedback](mailto:sqlfback@microsoft.com?subject=White%20Paper%20Feedback:Cleansing%20and%20Matching%20Master%20Data%20Using%20EIM).