Global Address Data Integration in SSIS:

Using Melissa's Global Verify Component

INTRODUCTION:

Internationalization is becoming a necessity for more and more businesses. As you expand your customer base, you will increasingly and inevitably start encountering data for clients outside of the United States. The data quality of address information, as most businesses are aware of already, is extremely important in terms of removing duplicates; saving in postal, shipping and product costs; as well as being able to perform effective business intelligence. This white paper will show you how to perform global address data quality with the Melissa Global Verify Component for use in Microsoft SQL Server Integration Services (SSIS).

Designing Your Input Forms

One thing to keep in mind is that U.S. address data is not the same as address data worldwide. It is better not to force international addresses into U.S. address formats, but instead start fresh. If you are not tied down already in the way your organization collects its address data, customizing the way your input forms are laid out can dramatically increase the effectiveness of the data you collect. However, one must also try to make the input easy and quick for the user. Our recommendation is to try and find a balance between accuracy and ease of use. Below is the list of all possible inputs Melissa's Global Verify Component can accept, and our recommendation for designing your input forms. This recommendation is based on a typical international address implementation. Obviously, it will not fit the needs of all organizations, but considering the points made below, should help you make the right choices.

- Our recommended fields for the input forms are based on the fact that those fields are universal to virtually every country, while the fields with an asterisk (*) are specific to a few. Most of the time, the additional asterisk field data can be put as an additional address line. So, unless you already have significant data in countries with these lesser-used fields, it should be safe to leave them off.

- Additionally, while the component can take 8 Address lines, most countries will only use a maximum of 4-5. Eight address lines might be a little overwhelming for typical users, so for typical implementations, 4 or 5 address lines should be sufficient. However, don't be too surprised if you will have to add more in the future.

Possible Inputs	Recommended for new input forms
Organization	Organization
AddressLine1 through AddressLine8	AddressLine1 through AddressLine5
Dependent Locality*	Locality
Double-Dependent Locality*	Administrative Area
Locality	Postal Code
Sub-Administrative Area*	Country
Administrative Area	
Postal Code	
Sub-National Area*	
Country	

- One very important situation to try and avoid is having non-address data in the AddressLines. For example: company names; contact names; care of information; and etc. If you expect to have that information as part of the input, it is important to have separate fields for them to be entered. While it is sometimes possible to handle and verify records with some non-address information in the address lines, it is always better to start off with cleaner data.

Output Design

The address output from the Global Verify Component is somewhat different from the input. They come in these different groupings. Each one of these groups contains the entire address, so you should not mix them for output. Doing so will result in duplicated address components.

Formatted Address	
The entire output address in one field with the organization name and destination country.	Useful for mailing labels
AddressLine1 – AddressLine8	
The entire address split into different lines.	Useful for databases as address is in format of the destination country
Parsed Components	
The entire address parsed into individual components, from all the delivery address parts to the locality, administrative area, and country.	Useful for matching and business intelligence

Further down, you will see the different output fields in the "Mapping Your Output" section.

Setup

Once you have the Global Verify Component installed on your SSIS server, you have the ability to apply data quality to international address data. You simply have to drag and drop the component onto a data flow and put it into the flow of your address data. One simple example of this is from an input table to an output table.
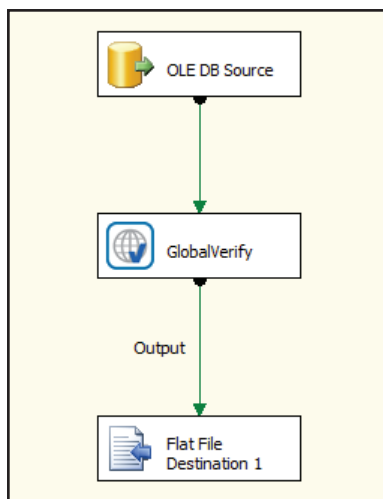
Mapping Your Input

Mapping your input in the component is very straightforward. Based on the input fields you have, map the fields to their corresponding inputs in the transform.

Mapping Your Output

The output fields are also fairly straightforward. Please remember the three types of outputs from the "Output Design." Pick which one(s) you would like and type in the column name. The AddressLines and FormattedAddress outputs are on this initial screen while the parsed components are located in the "Additional Output Columns" dialogue box.

Processing Options

The Global Verify Component has a few options to customize the output. Country of Origin and Line Separator affect only the FormattedAddress. The most important option here is the Output Script. Here, you can specify what language/character set you want the output to be. You can keep it in the language of the destination country or convert everything to Latin.
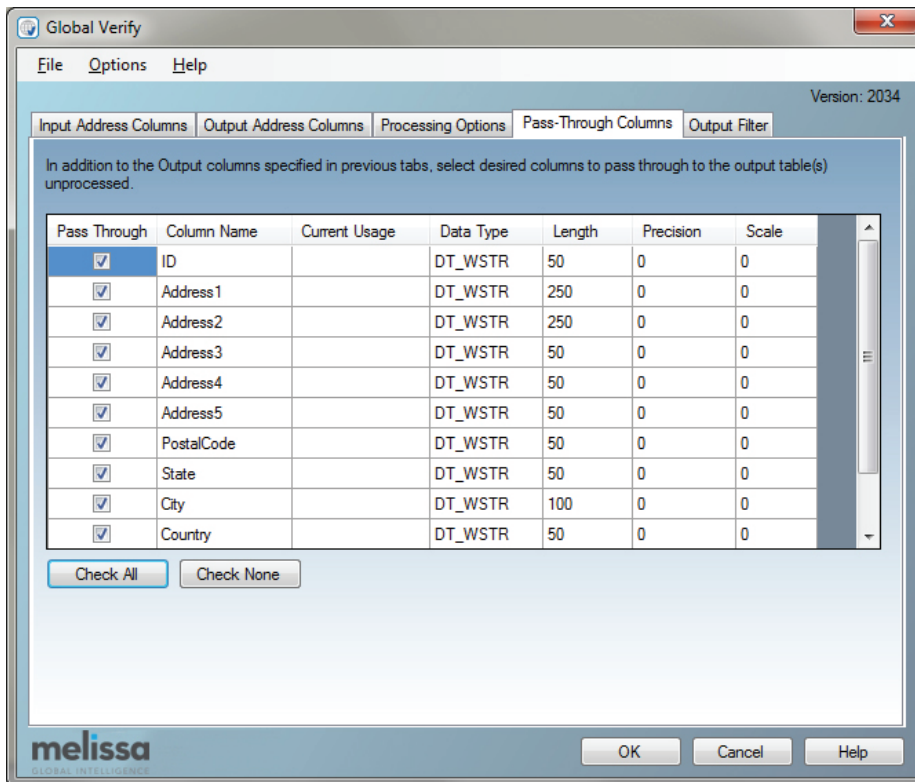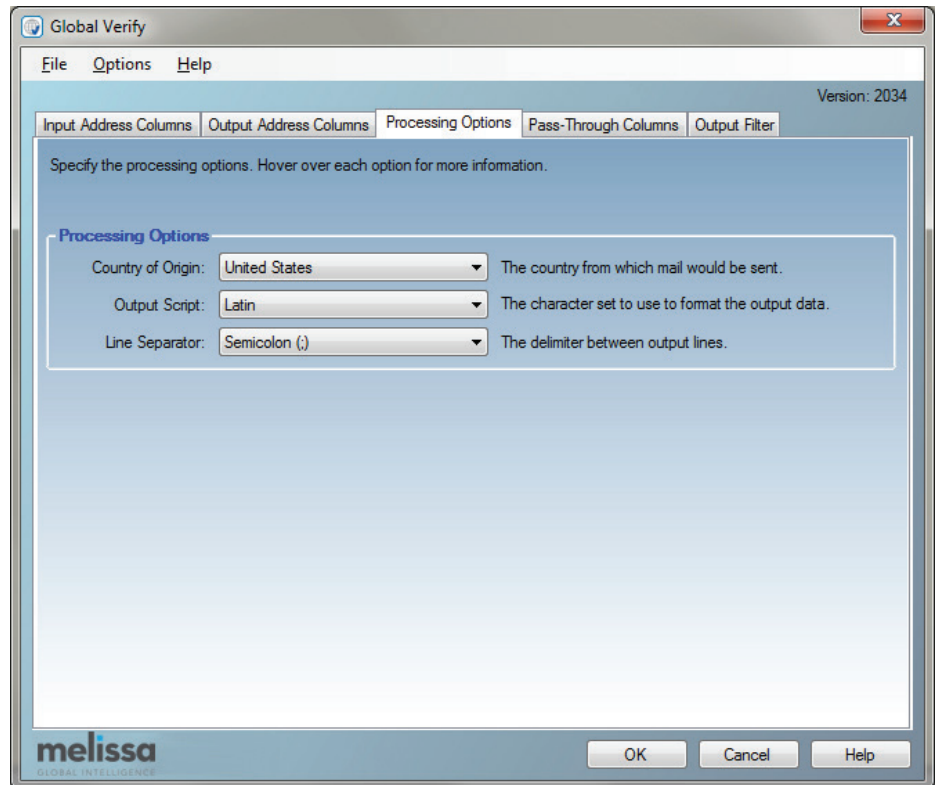
Pass-Through Columns

The Pass-through columns simply allow you to pass through from your input source any data you want to keep unchanged. This is useful for any key fields or other fields not used as input.
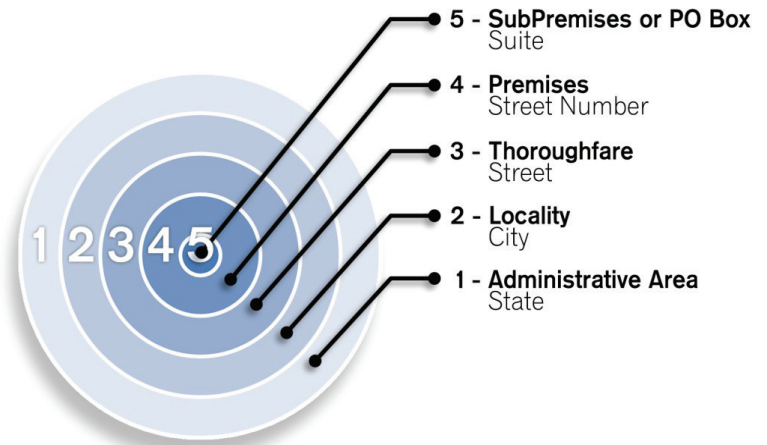
Understanding Results Codes

Results codes are the way Melissa alerts you to the status and errors for a record. There are Status Results codes, Error Results codes, and Change Results codes.

Status Results Codes:

These codes come in the format of AV2X or AV1X. AV2X denotes that the address was fully verified to the reference data available, while AV1X denotes that the address was partially verified. The 2nd digit (the X) indicates the level to which the address was verified.

Due to the fact that there are different levels of reference data available, AV2X will indicate that the address was verified to as high as level as we have data for. For example, for a US address, AV25 would be the highest level of verification because we have SubPremises reference data available for the U.S. while an address in Algeria, where Thoroughfare (3) data is all that is available, AV23 is the highest level of verification.

5 - **SubPremises or PO Box**
Suite

4 - **Premises**
Street Number

3 - **Thoroughfare**
Street

2 - **Locality**
City

1 - **Administrative Area**
State

Error Results Codes:

Error codes are returned when an address cannot be verified or corrected. They come in the form of AEXX and will denote what part of the address contains the error.

Change Results Codes:

For addresses that are corrected, change codes in the form of ACXX are returned for every part of the address that is altered from the original input.

The Global Verify Component documentation will list all the possible result codes and their meanings. Here is a sample of some.

AE*XX* Error Codes

Code	Errors
AE01	No Verification
AE02	Unknown Street
AE03	Component Error
AE05	Multiple Matches
AE08	SubPremises Number Invalid
AE09	SubPremises Number Missing
AE10	Premises Number Invalid

AC*XX* Change Codes

Code	Errors
AC01	Postal Code
AC02	Administrative Area
AC03	Locality
AC09	Dependent Locality
AC10	Thoroughfare Name
AC11	Thoroughfare Type
AC12	Thoroughfare Direction

Output Filter

The output filter allows you to split the output into different output streams based on a number of customizable conditions. Up to four filters can be specified for four different output streams. The records that satisfy a particular filter will be sent through to that output and the rest will cascade down to the next. For the following example, we have split the output into three different streams. "Fully Verified" are all the AV2X records, "Partially Verified" are all the AV1X records. The remaining records, the ones with errors, will fall to the third output.
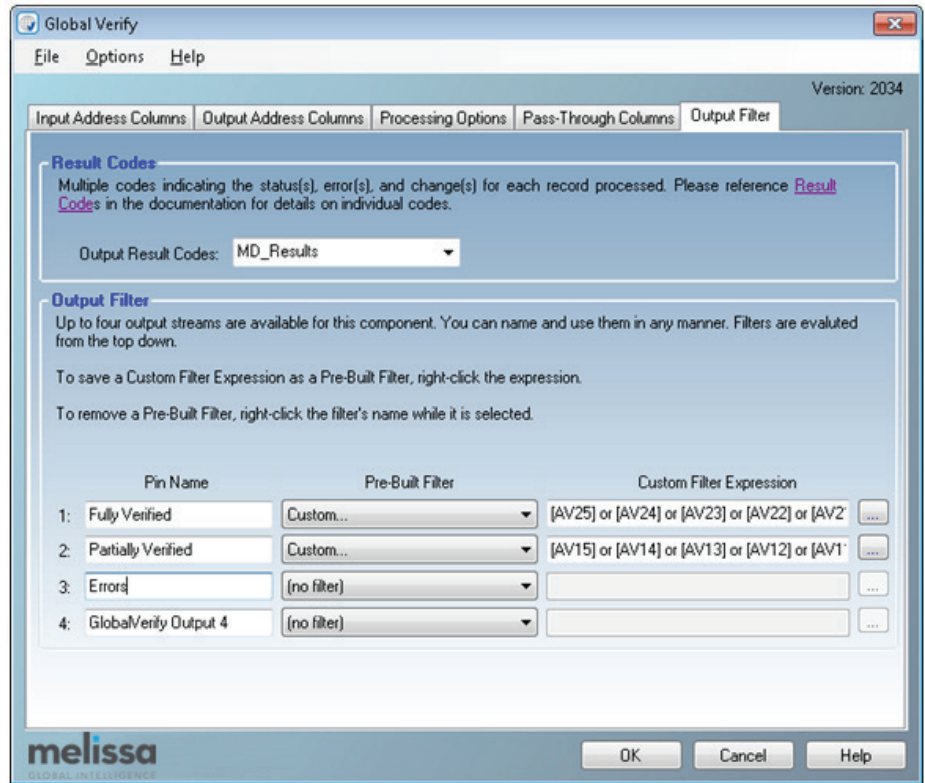
Given these output pins, we can direct our different types of verification to different destinations and handle them accordingly.
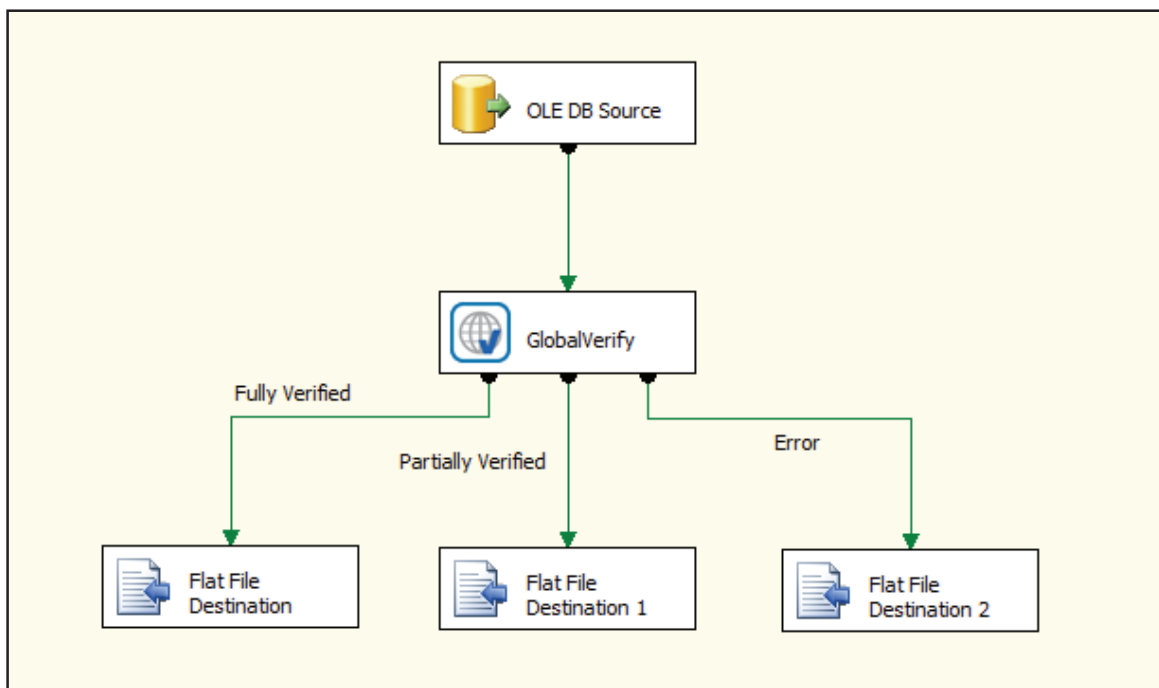
CONCLUSION

Melissa's Global Verify Component makes cleaning and validating your international addresses in SSIS simple and effective. The concepts and steps here can be applied to your organization or project, but it is by no means encompassing of every situation. We encourage you try the component yourself as well as review the documentation, samples, and videos available. If you are not already a subscriber, contact a sales representative at Melissa (1-800-635-4772 x4) for a free trial.

Resources

The Global Verify Component is part of the Data Quality Components for SSIS product. The product disc and download contains all documentation and samples. More resources can be found here:

Download a FREE Trial:
http://www.melissa.com/data-integration/microsoft-sql-server.html

SSIS Sitemap page:
http://www.melissa.com/tech/ssis/ssis-detail.html

Global Verify Documentation:
http://www.melissa.com/keywordsearch/address.htm#Global_Verify_Component.htm

SSIS Support Page:
http://wiki.melissadata.com/index.php?title=SSIS%3AData_Quality_Components

ABOUT MELISSA

Melissa is a leader in data-driven solutions that help organizations leverage Big Data and People Data (name, address, phone and email) to unite customer insights, analytics, data quality, and cross-channel marketing. We profile, cleanse, verify, enrich, and consolidate data assets, providing more than 10,000 brands in over 20 countries with accurate, reliable, and trusted information that can be utilized throughout the enterprise. For more than thirty years, our extended legacy in data quality, ID verification, and data enhancements has earned the trust of organizations from around the world.

1-800-Melissa (635-4772)

www.melissa.com