A MELISSA DATA PUBLICATION
01/11
Volume 6, Issue 1

Data Quality INSIDER

< IN THE SPOTLIGHT >

Introducing the Data Quality Components for SSIS

An in-depth look at the Data Enrichment and Verification Component

Data Quality Components for SSIS is a suite of custom data transformation components for SQL Server Integration Services (SSIS). SSIS is a platform for data integration and data flow applications, that allows for the extraction, transformation, and loading of data (ETL). Through SSIS, users can extract information from different database formats, run them through a transformation process, and load the output to a database in the desired destination and format.

The suite includes three sets of components to meet the requirement for total data quality as defined by Gartner©:

- Data Enrichment and Verification Component (DEVC)
- Data Matching and Cleansing Components (DMCC)
- Data Monitoring and Profiling Components (DMPC)

Data Quality Components for SSIS

What is the DEVC?

The DEVC grants you access to a spectrum of data quality benefits. Full names are parsed and genderized, email domains and phone numbers are corrected and verified, and addresses are verified, corrected and standardized.

continued on page ***3***

DATA QUALITY TOOLS

Address Object
Canadian Address Object
Contact Verification Server
< Data Quality Components for SSIS >
DQ*Plus
Email Object
GeoCoder Object
IP Locator Object
MatchUp Object
Name Object
Phone Object
Presort Object
RightFielder Object
SmartMover Web Service
Telco SmartSearch Web Service
WebSmart Services
WebSmart Global

The Insider Pro

Have you heard about our website's newly improved Support Center? Visit our FAQ section which offers more questions and in-depth answers to all our Data Quality Tools. The products, platforms and Sample Code Links featured in our Sample Source Code section are now more organized and easier to navigate.

Our new live and on-demand webinars are also picking up steam. We recently featured the County of Sacramento who demonstrated how they integrated real-time Web-service-based address verification into SAP. It's the first time we've featured a client utilizing our tools in a webinar. Be sure to check it out! To view our webinars, click here:
www.melissadata.com/shows/index.htm

Happy New Year!

Bud Walker
Director,
Data Quality Solutions

News Bytes

>> New Result Codes Provide More Details for Address Verification

In addition to accurate verification, users have often asked for a detailed description as to why an address failed to verify, or if it did, what change was made to the data. Address Object now returns an additional 19 new Result Codes – giving you increased flexibility to handle returned results. There are a total of 50 Result Codes.

New Result Codes:

- The **FindSuggestion Codes**

RESULT CODE	MEANING	DETAILS
AS21	Alternate Address Suggestion Found	One or more alternate address suggestions were found for this address since it cannot be corrected.
AS22	No Alternate Address Suggestion Found	No alternate address suggestion found for this address.
AE19	The MAX time allotted to find a suggestion has been exceeded	The time allotted to FindSuggestion has been exceeded. The process has been stopped.
AE20	FindSuggestion method cannot be used	Cannot offer suggestion. CassEnable must be set to false and DPV add-on must be enabled to use FindSuggestion.

- The **Suite Codes**

RESULT CODE	MEANING	DETAILS
AE17	Unnecessary Suite/Apt Error	The input contained suite/apt for an address that does not have suites.
AS23	Extraneous Information Found	Extra information not used in verifying address was found. They have been placed in the ParsedGarbage field.

- The **Change Codes**

RESULT CODE	MEANING	DETAILS
AC01	ZIP Code Change	The ZIP Code was changed.
AC02	State Change	The state abbreviation was changed.
AC03	City Change	The city was changed.
AC04	Base/Alternate Change	The street address was changed because it was found in the base/alternate table.
AC05	Alias Name Change	The street name was changed from a former street to the USPS preferred street name.
AC06	Address1/Address2 Swap	Address1 was swapped with Address2. (only Address2 contained a valid address)
AC07	Address1/Company Swap	Address1 was swapped with Company because only Company contained a valid address.
AC08	Plus4 Change	A non-empty Plus4 was changed.
AC09	Urbanization Change	The urbanization was changed.
AC10	Street Name Change	The street name was changed.
AC11	Street Suffix Change	The street suffix was changed.
AC12	Street Directional Change	The street predirection or postdirection was changed.
AC13	Suite/Apt Name Change	The suite/apt name was changed.

To read more on Result Codes, go to the Address Object Reference Guide, pages 62-64.
http://www.melissadata.com/tech/addressobject.htm
If you need help or more information, please contact our tech support team at 1-800-MELISSA (635-4772), opt 4.

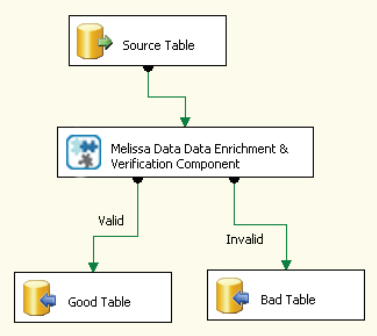
continued on page ***4***

< IN THE SPOTLIGHT >

continued from ***Page 1***

Introducing the Data Quality Components for SSIS

The DEVC also provides for useful enrichment data, such as the geographic coordinates (latitude/longitude), ZIP + 4®, delivery point, county information, census data, etc.



A typical SSIS data flow (with the inclusion of the DEVC):

Using the DEVC

The DEVC can be easily inserted into a new or currently existing data flow. A data flow is simply the visual representation of your data's ETL process.

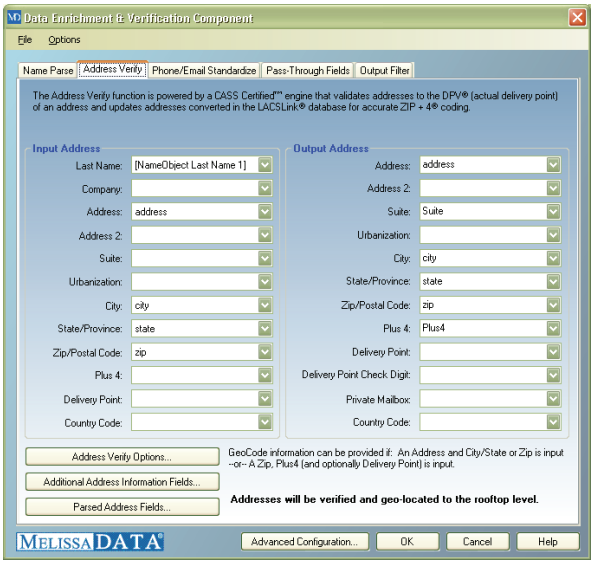
This simple data flow is divided into 3 stages: the source; the DEVC; and the destination – a basic representation of a typical ETL scenario. Let's take a closer look at these stages and how each of them works individually, and how they all come together as a whole in your data-flow process.

The Source

From flat files and XML schemas to Excel spreadsheets and OLE DB sources, SSIS provides a wide range of possible input database formats that you can connect to, automating the process of parsing out rows and individual columns.

email	FullName	address	city	state	zip	phone	recID
!!!!samarooni@gmal.coom	Worley Alex	4402 VINTON DRIVE	Omaha	NE	68105	4029912262	100000049
SAM@2fords.net	SAM PALMOROS	22382 Empressa	RANCHO SANTA MARGARITA	CA		905831##6259	100000050
m$$$elody@brightok.net	Melody Cain	rr 2 bOX 110	Marietta	OK	73448	580-276-4119	100000115
samcat0412@aol.com	gary yen^^	42 Columbiaaaaa St			02879	401 2157427	100000139
joseph21@yaho.com	Pearlie@@@ Coleman	9190 Monte Vista Avenue	Montcalir	CA		(606)2673692	100000348

A selected input database in a data flow process. This example contains unwanted characters, non-standardized casing, incomplete addresses, missing ZIP™ codes or cities, and differently formatted phone numbers.



The component is designed in a tabular step-by-step fashion, with the goal of making it as intuitive and as easy to use as possible.

DEVC Interface

All these verification and enrichment capabilities can be achieved through a single pass in your data flow process by simply mapping in the fields you wish to process through the DEVC and specifying the desired outputs. And because the DEVC makes use of a GUI, all this can be accomplished without having to write a single line of code!

The Destination

After the transform process takes place, the resulting data is stored to a destination database of your preferred format.

Through the DEVC – integrated within Microsoft SSIS – we've successfully developed a solution for your data quality needs with Microsoft's flagship data integration platform. By leveraging the SSIS environment, you can easily connect to several different database formats without having to worry about all the low level details of how to parse specific database

continued on page ***4***

continued from ***Page 3***

LastName	Address	Suite	Zip	Plus4	Latitude	Longitude	PhoneAreaCode	PhonePrefix	PhoneSuffix	CorrectedEmail
Worley	4402 Vinton St		68105	3848	41.22938	-95.97881	402	991	2262	samarooni@gmail.com
Palmoros	22382 Avenida Empresa		92688	2112	33.63723	-117.60741	905	831	6259	sam@2fords.net
Cain	RR 2 Box 110		73448	9611	33.93743	-97.121014	580	276	4119	melody@brightok.net
Yen	42 Columbia St		02879	3210	41.44199	-71.49572	401	215	7427	samcat0412@aol.com
Coleman	9190 Monte Vista Ave	Apt 210	91763	1770	34.087572	-117.698377	606	267	3692	joseph21@yahoo.com

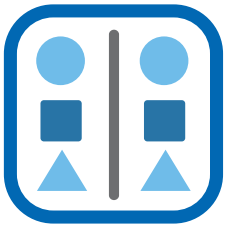
This is what our original input data looks like, after it has gone through the DEVC.

formats. But the true power of this product lies in its ease of use, to quickly provide better data quality in a database administrator-friendly ETL environment, where programming ultimately becomes unnecessary.

continued from ***Page 2***

News Bytes

>> New Fuzzy Match Component for SSIS Set for Release



Melissa Data announces the upcoming release of its new Fuzzy Match component to be included in its suite of Data Quality Components for SSIS.

Using the fuzzy matching component gives you the ability to detect, link, parse and consolidate duplicate records into a single accurate view of your customer records to streamline communications and business operations.

The new Fuzzy Match Component will debut with a variety of algorithms available for the user to pick and choose from, and with more to be added in the future. Included are Exact, Fonetex, Soundex, Jaro-Winkler, N-gram, Levenshtein distance, Dice's Coefficient, Jaccard's Similarity, Jaccard's Similarity Coefficient, and more.

The user can quickly and simply select an algorithm, set the threshold, select which columns to pass through, and run the match.

More advanced users can dive deeper and chain multiple algorithms together. The component allows you to qualify records as a match if it passes every algorithm, passes any algorithm, passes an average of the algorithms, or passes at least a specific number of algorithms.

With the ease of use, the power, and the flexibility of the new Fuzzy Match component, Melissa Data's confident it can replace any fuzzy tools you are currently using.

Call a Melissa Data sales representative today at 1-800-MELISSA (635-4772) opt 3 to request a free trial when the component is ready.

Helpful Links

Melissa DATA Resources
Link to our white papers, case studies, and insightful articles to help you get the most out of your direct mail, marketing, and data management initiatives.
www.MelissaData.com/resources

Discussion Forums
Ask questions, talk product, and get fast answers in our discussion forums.
http://forum.MelissaData.com

Free Trials

Parse full names into 5 components. Determine gender makeup. Get a free trial of Name Object.
www.MelissaData.com/dqi-name

Validate, correct email addresses. Download a free trial of Email Object.
www.MelissaData.com/dqi-email

Identify, extract, organize free form data. Get a free trial of RightFielder Object.
www.MelissaData.com/dqi-right

Director, Data Quality Solutions
Bud Walker
bud@melissadata.com
1-800-635-4772 x159

Melissa Data Technical Support
tech@melissadata.com
1-800-635-4772 opt 4 (6 am to 5 pm PST)

Editor Abby Telleria
Writers Bud Walker, Admound Chou, Tim Sidor, Joseph Vertido
Art Director Melody Yen
Graphic Designer Timothy Magoun
Contact the editor at insider@melissadata.com

Melissa Data Corp.
22382 Avenida Empresa
Rancho Santa Margarita, CA 92688
Ph 1-800-MELISSA (635-4772)
Fax 949-589-5211
www.MelissaData.com

© 2011 Melissa Data Corp. All rights reserved.

MELISSA DATA®
Your Partner in Data Quality