

## DATA QUALITY COMPONENTS FOR SQL SERVER®

Your Full Spectrum  
Data Quality Solution



800.MELISSA (635.4772)  
[www.Melissa.com](http://www.Melissa.com)

**melissa**<sup>TM</sup>

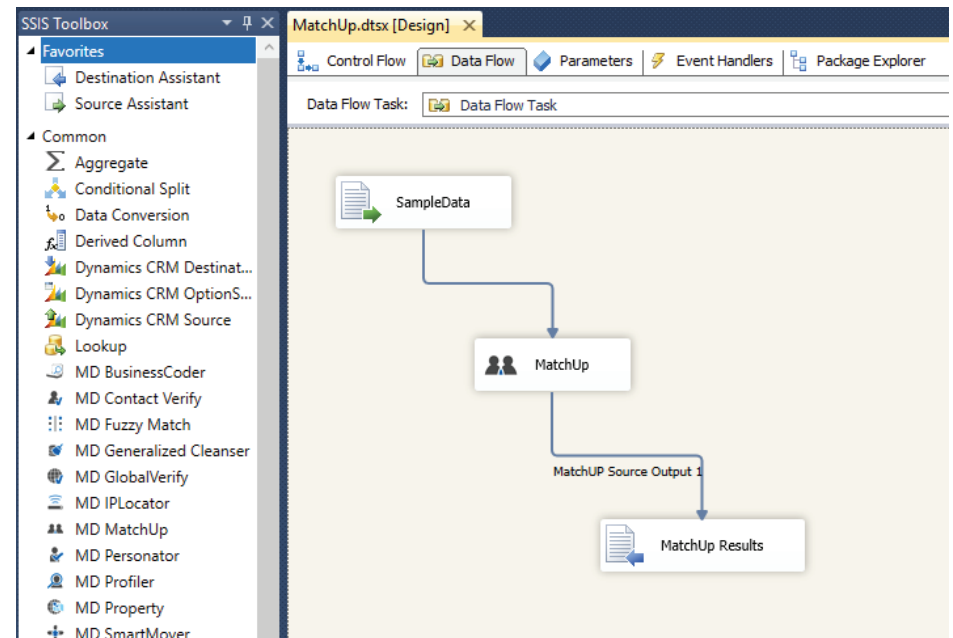
Did you know that 91% of businesses suffer from common data errors? The most common errors include incorrect and inaccurate data, outdated data, missing information, and duplicate data. Without accurate data, it's hard to find new customers, better understand customer needs, and increase the value of each customer.

That's where the Data Quality Components for SQL Server® come into play. This unique set of customer data management tools leverages the integration power of SQL Server and Melissa's suite of global data quality solutions to empower organizations to collect data from any source, cleanse and transform it, and gain immediate insight for actionable intelligence.

## Robust ETL & Customer Data Integration

Melissa's Data Quality Components for SQL Server reads from and writes to relational databases that exist in SQL Server, fixed or delimited text files, and other file formats supported by third party connectors. Go beyond standard ETL tools to scalable and flexible management of end-to-end data flows with:

- Dynamic and reusable data integration templates to save time and create transformation on the fly.
- Robust administration features like performance monitoring, job roll-back and restart, and operations made for usage auditing.
- Enterprise-grade security including access and version controls as well as a LDAP and Active Directory integration.



## Full Spectrum Data Quality

Our Data Quality Components for SQL Server leverage the full spectrum of data quality solutions Melissa offers, so you can:

**PROFILE YOUR DATA** to identify weak points in the data collection process.

**CLEANSE DATA** with a variety of programmatic or regular expressions to correct data inconsistencies.

**VALIDATE DATA** using powerful global address, name, phone, and email verification processes, and match name-to-address to verify ID.

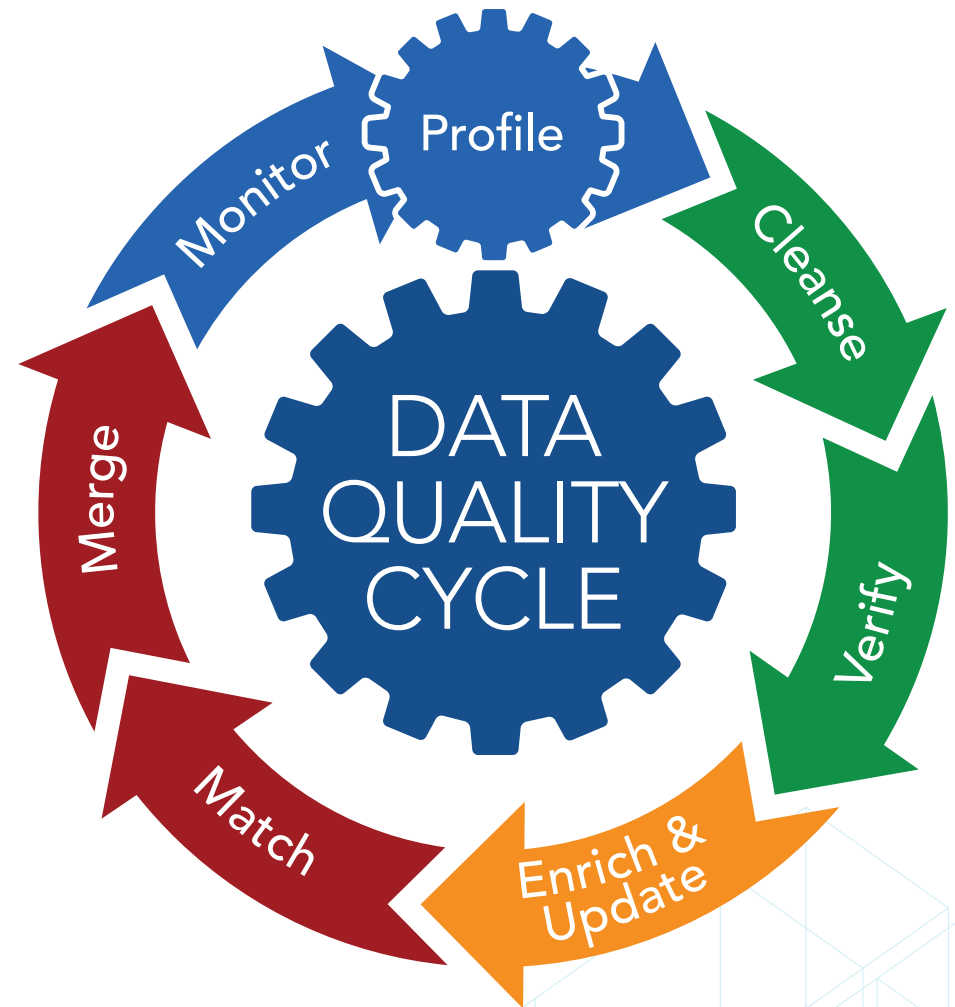
**ENRICH YOUR DATA** by adding valuable demographics, IP location, and property data for greater insights.

**UPDATE DATA** by accessing up to 20 years of change-of-address records to get the most current address of your customers.

**MATCH DATA** with advanced fuzzy matching algorithms and survivorship rules to eliminate duplicates.

The Data Quality Components for SQL Server are perfect for organizations that need consistent, trusted, accurate data across the enterprise for better customer data management and integration. With our full suite of easy-to-implement solutions, customer data management becomes simply better customer relationships.

- Achieve better insight with clean, accurate customer data
- Easily onboard, apply standards, enforce rules, cleanse and enrich Big Data
- Run data quality on-premise and in the cloud



*The full spectrum of data quality delivers accurate, trusted, actionable data for data integration, MDM, enterprise data warehousing and Big Data success.*



## Profiling Component

The first step in improving data quality is to profile your data to gain valuable insight into its accessibility and usability. The Profiling component generates simple-to-advanced profiling information, including basic data statistics (mean, median, frequency, variation, etc.) and details (structure, content, classifications, etc.) to identify data errors at their source. You can then monitor the performance of your source data over time to specified requirements of pre-set limits.

- Enhance data management and data warehousing efforts by identifying weak points in your data
- Optimize data quality over time by continuously monitoring data
- Enforce business rules on incoming records to maintain data standardization

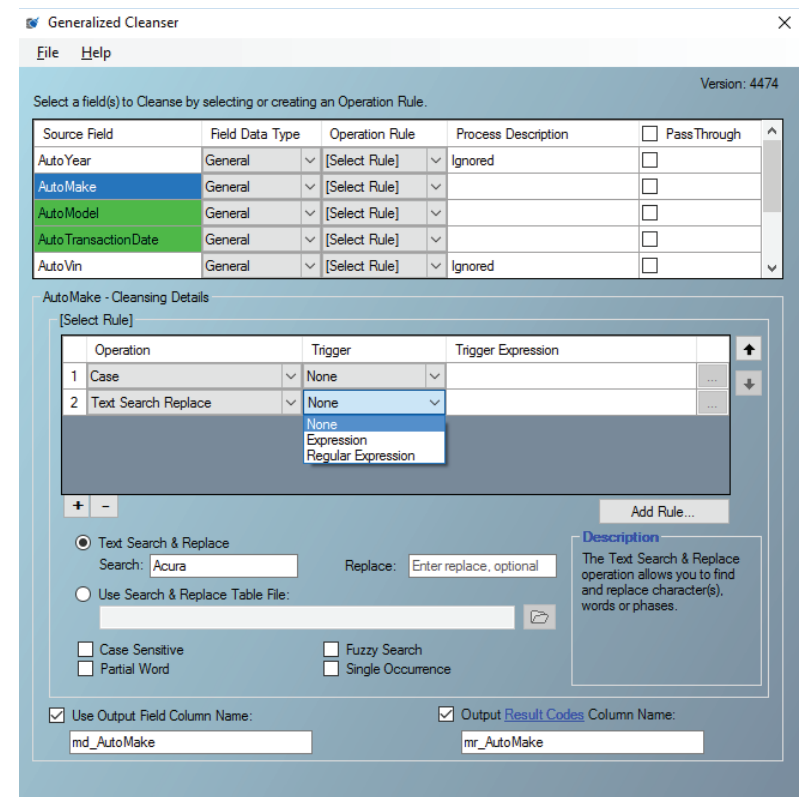
The screenshot shows the 'Profiler' application window. It has a menu bar with 'File' and 'Help'. Below the menu bar are three tabs: 'Input Field', 'Analysis Options', and 'Output'. The 'Analysis Options' tab is active. The main content area has a blue header with the text: 'Analysis Options have an impact on how quickly processing completes. For faster processing, turn off unneeded options.' Below this, there is a list of analysis options with checkboxes: 'Sort Analysis' (unchecked), 'Matchup Analysis' (checked), 'RightFieldier Analysis' (checked), and 'Data Aggregation' (checked). To the right of this list is a section titled 'Setup Options' with four input fields: 'Table Name:', 'User Name:', 'Job Name:', and 'Job Description:'. At the bottom right of the window, the version number 'Version: 1.0.0.4485' is displayed.



## Generalized Cleansing Component

Build data cleansing scripts for a wide range of data errors and inconsistencies with the Generalized Cleansing component. It combines six operations that allow you to cleanse data and save operations (simple or complex) for future projects. The component can be beneficial to users in many different industries, giving them the ability to standardize and validate inventory lists, correctly format automobile data, and much more.

- Cleanse any type of data and achieve a higher standard of data quality for integration, warehousing, and analytics
- Gain greater control of your data when optimized, and save your business time and resources
- Customize and create rules (triggering) to standardize data



## CLEANSING OPERATIONS AVAILABLE:

### CASING

Change the casing of data from capital to lowercase, etc.

### ABBREVIATION

Expand or contract abbreviations, for example: CA to California

### EXPRESSIONS

Create programmatic expressions to make sense of data values

### PUNCTUATION

Add or remove punctuation

### SEARCH AND REPLACE

Replace portions of a string

### REGEX

Use regular expressions to extract, validate, etc.

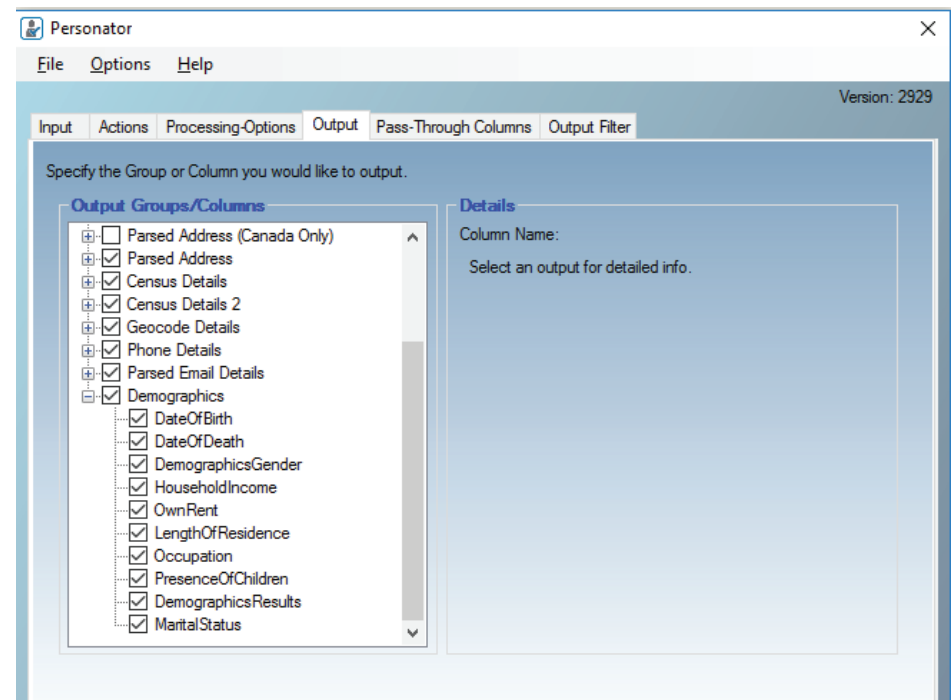
## Personator® Component

The Personator component taps a multisourced dataset containing billions of records to validate each element of a U.S. or Canadian contact record: name; address; phone; and/or email – and match name-to-address to verify identity.

Personator also enriches your contact records by:

- Filling in missing contact information like business/consumer emails, business/consumer addresses, business/consumer phones, and business/consumer names
- Adding current addresses for customers and prospects that have moved in the U.S. and Canada – matching your records against a propriety database of address changes going back 20+ years.\*
- Adding detailed consumer demographics including date of birth, gender, presence of children, marital status, occupation, income, and more
- Appending records with the most accurate Geopoints (lat/long coordinates) at the rooftop level for 95% of U.S. physical addresses (residences and businesses). Includes information on the County Name, FIPS Code, Census Tract, Block Groups, Block Numbers, and "Core Based Statistical Area (CBSA), and Metropolitan/Micropolitan areas for location intelligence

*\*Note: This process does not satisfy the USPS® Move-Update requirement for discounted postage rates. The SmartMover<sup>SM</sup> component does.*

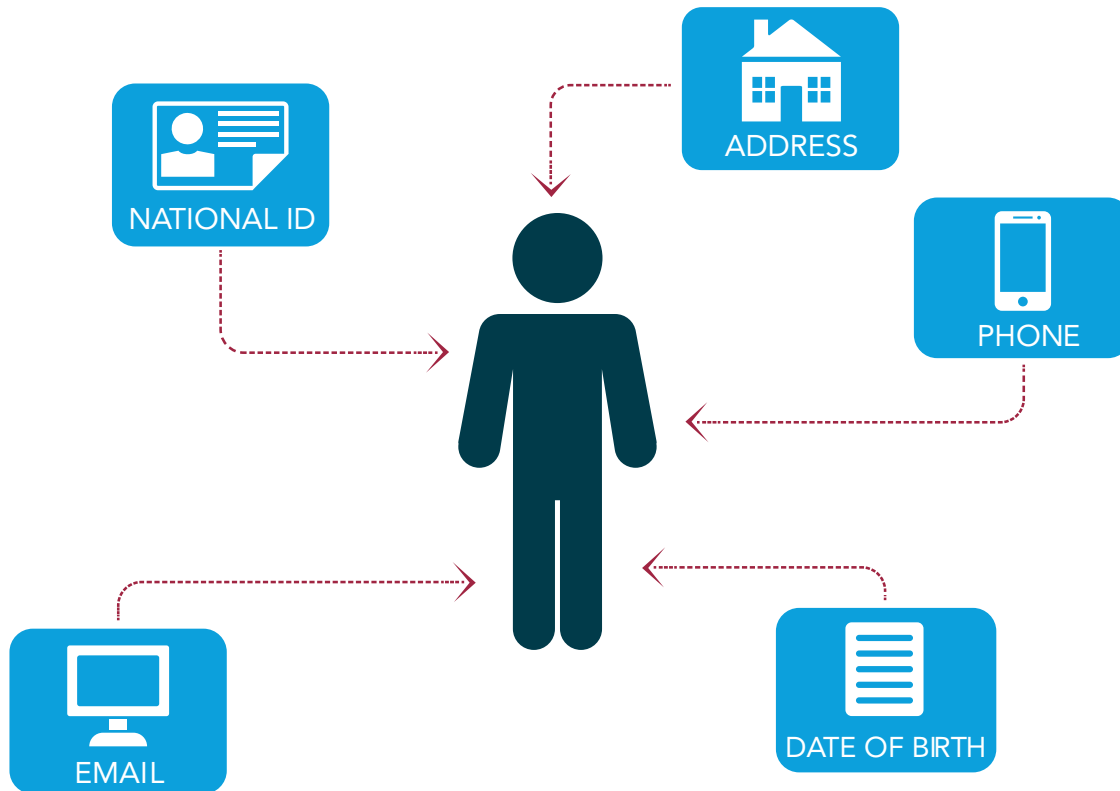


## Need the speed and security of on-prem deployment?

If you are looking for greater speed and the security of on-premises processing, our Contact Verify Component is available as an on-prem API for Address, Name, Email and Phone Verification that meets your security, privacy, compliance and/or bandwidth requirements.

# Personator World Component

Personator World offers premium identity verification for smarter, simpler ecommerce, fraud prevention, Know-Your-Customer (KYC), Card Not Present, and Anti-Money Laundering compliance.



## Age Verification and National ID

Verify a customer's national ID, such as a social security number or driver's license, as well as date of birth.



## Name-Address Verification

Leverage a comprehensive dataset containing billions of records to confirm the name and address match and are current - with a high degree of accuracy.



## Address Correction and Formatting

Add missing street suffixes, state/province/administrative area info and standardize addresses to specific country formats.



## Contact Data Validation

Determine that the given postal address is deliverable, the email address exists, the name associated with a mobile device, and whether the phone number is active and callable, and the given name is in a valid format.

## Global Verify Component

The Global Verify component cleans, verifies, standardizes, and formats all your global people data – name, address, email, and phone – to ensure your database contains accurate, trusted, and actionable information.

### Global Address Verification

The component verifies addresses from 240+ countries, standardizes addresses to local country formats, and adds missing postal codes, regions, etc. It also supports many different language sets and can transform non-Latin writing systems into Latin characters so addresses in foreign languages like Chinese, Russian, Greek, Japanese, and others can easily be validated.

### Global Phone Verification

The component's phone verification function includes real-time lookup technology to distinguish between global landlines and mobile numbers, and validates the number is accurate, live, and callable. Caller ID will return the name and address associated with the billing contact.

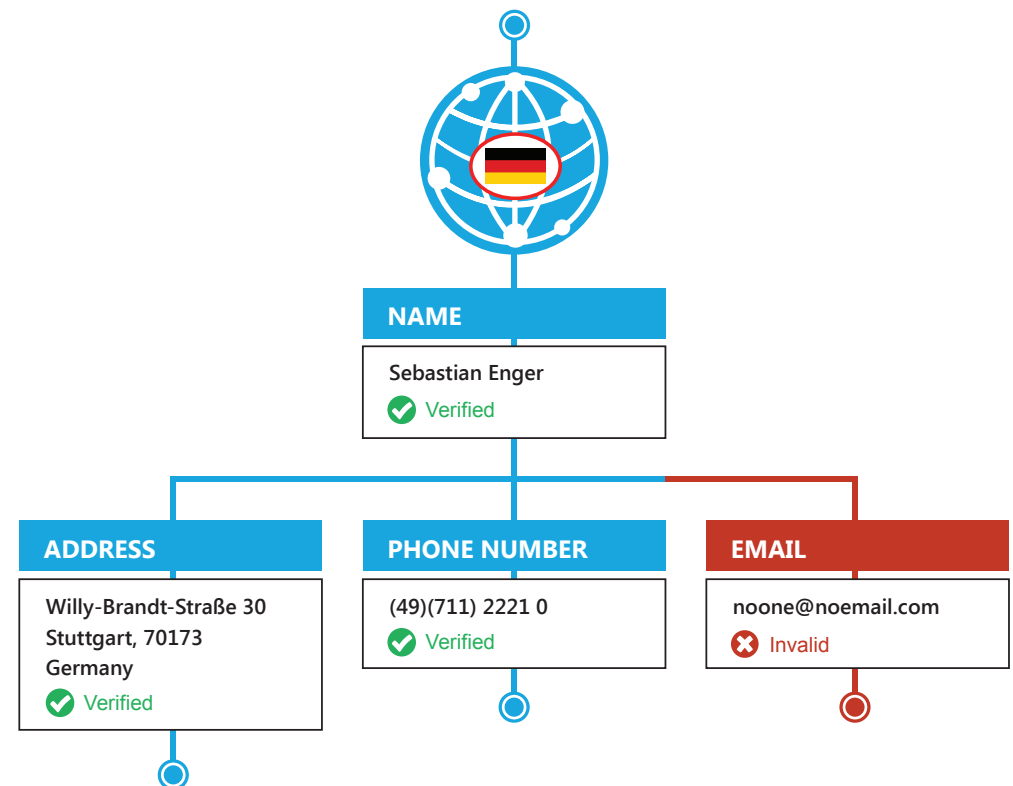
### Global Email Verification

Email verification capabilities include: real-time email mailbox verification to ensure an inbox is live; email syntax and domain correction; and FCC Mobile Domain Detection to ensure CAN-SPAM compliance.

## Global Geocoding

The Global Verify component geocodes international addresses for 40+ countries, providing precise (rooftop) latitude/longitude coordinates for an address to power better business intelligence, market segmentation, sales clustering, logistics, and risk exposure. Find customers around the world to increase ROI, more efficiently plan investments, and go deeper with global market analysis.

*\* Global Contact Verify is also available as a U.S. and Canadian Contact Verify Component.*





# 20%

of the data in a company's database is incorrect or outdated, resulting in returned mail, address correction fees and missed opportunities.

## SmartMover<sup>SM</sup> Component

The SmartMover component provides change-of-address processing for U.S. and Canadian addresses. This helps you stay in touch with consumers and businesses that have moved, eliminate the costs associated with undeliverable-as-addressed mail, and qualify for applicable postal discounts.

For U.S. addresses, the SmartMover component matches your customer records against the USPS® full NCOA<sup>Link</sup>® data file of 160 million moves going back 48-months. Processing your records utilizing the SmartMover component will satisfy the USPS Move-Update requirement for a period of 95 days from

the date of processing to qualify for First-Class<sup>TM</sup>, and USPS Marketing Mail<sup>TM</sup> (Standard Mail<sup>®</sup>) discounts. Melissa is an NCOA<sup>Link</sup> Full Service Provider licensee of the USPS. NCOA<sup>Link</sup> processing requires a Processing Acknowledgement Form (PAF).

For Canadian addresses, the SmartMover component will match your address file against the Canada Post NCOA<sup>®</sup> database containing over 11 million records from the last 72 months. Melissa is one of only a handful of Canada Post licensees. Canada Post NCOA processing requires a signed copy of the Canadian NCOA Acknowledgement Form.



## MatchUp® Component

On average, a database contains 8 - 10% duplicate records. These duplicates result in waste and inefficiencies and cloud your ability to get a single, accurate view of the customer. The MatchUp component uses over 16 advanced fuzzy matching algorithms and deep domain knowledge to find even the hardest-to-detect U.S. and international (Canada, Europe, South America, Australia and more) duplicate records.

The MatchUp component also includes other matching capabilities for specific scenarios including:

### Survivorship/Golden Record

MatchUp's most unique method for determining the most accurate view of the customer – the Golden Record – is called survivorship. Most traditional rules-based approaches to survivorship use techniques involving (1) the most recent, (2) the most frequent, or (3) the most complete record. MatchUp also utilizes these techniques, as well as determines and consolidates the best possible record based on the actual quality and validity of the data in each field to easily rank and select the surviving record. This technique for golden record selection offers the most effective and logical approach when it comes to

survivorship, since it's based on a relevant data quality score derived from the validity of address, name, phone, and email data. Duplicate entries are collapsed into a single customer record, while still retaining any information that may be applicable.

### Proximity Matching

MatchUp's patented distance algorithm uses lat/long coordinates and proximity thresholds to identify duplicate data.

### Householding

MatchUp can identify and consolidate records that are members of the same household. This is useful in being able to evaluate the total sales relationship and by eliminating unnecessary multiple mailings for cost savings.

### List Intersection/Suppression

MatchUp finds all the common data elements between multiple lists and/or use suppression to find just the data unique to each individual list.

NAME	ADDRESS	CITY	STATE	ZIP	PHONE	LAST_VISIT	SALE_AMOUNT
John Doe	22382 Avenida Emp.	RSM	CA	92688		4/16/2013	45
John Doe	22382 Avenida Empresa	RSM	CA	92688	800-800-6245	10/22/2008	
John Doe	22382 Ave. Empresa	RSM	CA	92688	949-2007	8/4/2011	560

The following three contact records are duplicates. The Golden Record selection criteria uses a data quality score to determine that the second record is the Most Complete record.

NAME	ADDRESS	CITY	STATE	ZIP	PHONE	LAST_VISIT	SALE_AMOUNT
John Doe	22382 Avenida Empresa	RSM	CA	92688	800-800-6245	4/16/2013	560

This graphic shows how the survivorship process works. Incomplete matching records, like the most recent sale date (Last Visit) and the highest purchase amount (Sale Amount) are collapsed to fill in the blanks - to create a complete, accurate, single record - the Golden Record.

## Fuzzy Match Component

The Fuzzy Match Component leverages a toolbox of state-of-the-art fuzzy matching algorithms and allows user-specified granular control on match thresholds and even fine tuning of the algorithms. Fuzzy Match can use any of the following matching algorithms on any column in your database to identify “non-exact matching” duplicate records:

Phonetex	MD Keyboard
Soundex	Jaro
Containment	Jaro-Winkler
Frequency	Levenshtein
Fast Near	n-Gram
Accurate Near	Needleman-Wunch
Frequency Near	Smith-Waterman-Gotoh
UTF-8 Near	Dice's Coefficient
Vowels Only	Jaccard Similarity Coefficient
Consonants Only	Overlap Coefficient
Alphas Only	Longest Common Substring
Numerics Only	Double MetaPhone

For a complete description of each matching algorithm, visit our wiki at [wiki.melissadata.com](http://wiki.melissadata.com).

## Levenshtein Algorithm

JOHNSON	JHNSN	71%
JOHNSON	JOHNSNO	71%
JOHNSON	JAMESON	57%

		J	O	H	N	S	O	N
	0	1	2	3	4	5	6	7
J	1	0	1	2	3	4	5	6
H	2	1	2	1	2	3	4	5
N	3	2	3	2	1	2	3	4
S	4	3	4	3	2	1	2	3
N	5	4	5	4	3	2	3	2

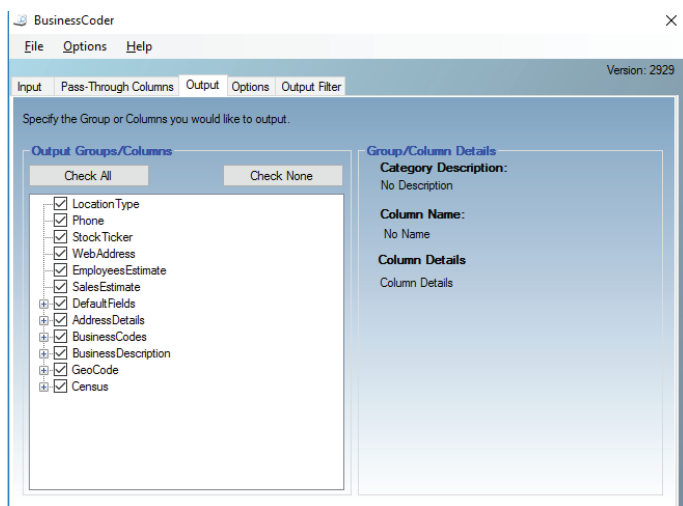
The Fuzzy Match component performs a fuzzy match on the non-exact matching records to calculate a match score between pairs of records. If the match score of the two records meets the match score threshold, then Fuzzy Match considers the two records a match. One example of the fuzzy matching algorithms used is Levenshtein shown above.



## BusinessCoder® Component

Add detailed firmographic data to your business records including company and contact information (name, address, phone, and email address), job title, SIC/NAICs codes, employee size, stockticker abbreviation, sales estimate and more, with the BusinessCoder component. Improve lead scoring, analytics, plus sales and marketing efforts by adding firmographic elements to your prospect and account records.

Company Name	City
Job Title	State
Phone Number	Postal Code Plus4
Email	Delivery Type Indicator
Address	And more!
Suite	

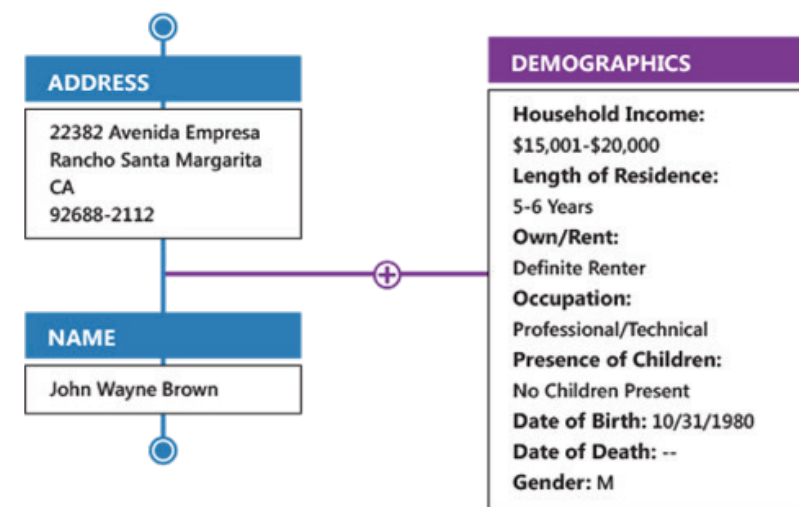


BusinessCoder returns firmographic data from over 25 million businesses.

## Demographic Component

The Demographic Component provides detailed consumer demographics from over 2 billion records containing 250 million individuals and 170 million households in the U.S. This valuable information offers you deeper insight into your customers, their behavior, and trends. Armed with this information, you can also craft personalized messages to help increase response rates.

Date of Birth	Home Owner/Renter
Deceased Information	Household Income (range)
Gender	Length of Residence
Presence of Children	Dwelling Type
Number of Adults	Occupation
Marital Status	

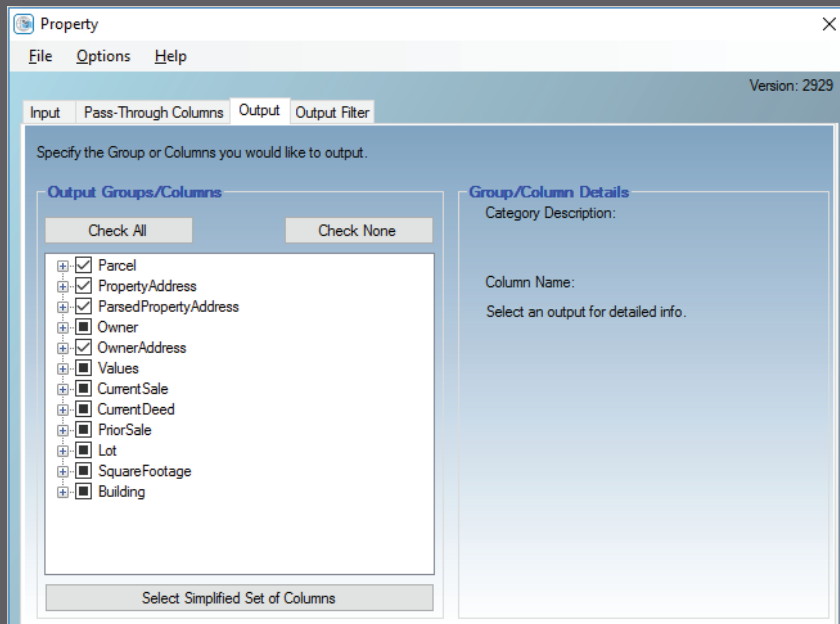


Variables for 11 selects like gender, occupation, and income to Know Your Customer (KYC).



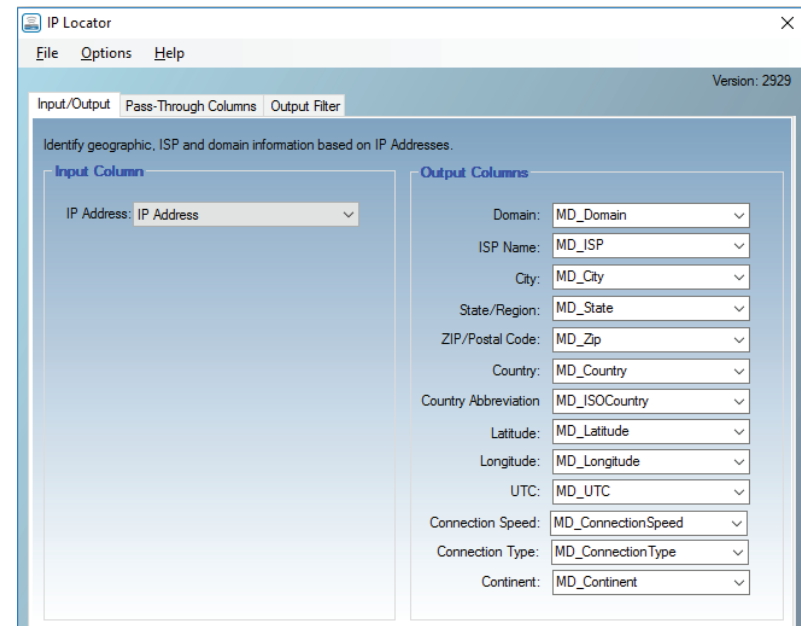
## Property Component

The Property component provides valuable property and mortgage data from over 140 million records for the U.S. Access variables from over 165 fields in categories such as: parcel; property address; owner; owner mail address; property values; current sale; current trust deed; prior sale; lot/land; square footage; building; and more!



## IP Locator Component

The IP Locator component provides the geographic location of an Internet Protocol (IP) address, including latitude, longitude, city, state, ZIP Code™/Postal Code, region, hosting Internet Service Provider (ISP), connection speed and type, and domain name. This information helps you identify where your web visitors are coming from – very useful for ecommerce sites looking to prepopulate country code on forms, display a different language, filter access from countries you don't do business with, and/or reduce credit card fraud based on geographic location. IP Location can also help fight illegal spamming and hacking by identifying the location of the problem.



# Community and Full Edition – Data Quality Components for SQL Server

Melissa is focused on delivering the industry's most complete solution for data quality and data integration. We offer enterprise-class capabilities with features superior to or not available in other products.

If you are interested in testing out our toolkit, we offer a free trial of the Full Edition or a free downloadable Community Edition. The Community Edition includes:

- Address Parsing
- Name Parsing (last name only)
- Email Correction
- Phone Formatting
- Matching (50,000 record limit with 9 built-in matchcodes)
- Profiling (50,000 record limit)
- Generalized Cleanser (100,000 record limit)

To request either version, visit [www.melissa.com/sqlserver](http://www.melissa.com/sqlserver).



## Customer Testimonials & Reviews

"Initial POC was up and running in just a few days with no training needed. The plug-in into our ETL tool was seamless and fully integrated into our existing processes. (MatchUp for SSIS) de-duplicates our customer data in an effective way so that we are able to reduce marketing costs and increase the quality of communication with customers. We are continually impressed by its sophistication and ease of use."

- Gary M. | Data Architect

"We simply passed basic details of our customers like Name, Address, and Phone Number and the Personator application was able to append Emails, New Address if moved, Phone number, Geocodes. Personator also standardized our existing customer information. The match process provided reliable single view of data."

- Mohan P. | Director of Business Intelligence

"We are cleansing our data so that we do not mail invalid records. We've saved hundreds of thousands of dollars a year and decreased cost per acquisition."

-Nicole M | Marketing Manager

"We loved the simple and straight-forward integration with our existing website order flow. We especially appreciate the up-to-date address information (CASS certified) and the real-time access to that data. In addition to the rich web services restful API offerings we also found great value in their Excel plugin for one-time use projects. I have used Melissa services now for nearly a decade and have never experienced any downtime issues. What a wonderful professional partner!"

- Ed B. | Sr. IT Project Manager

"The technology works seamlessly with SQL Server Integration Services, and provides another way for my clients to achieve highest data quality information. It can support high volume data matching, reporting, and it is helpful for especially small to medium sized business information."

- Mark T. | Microsoft Solution Architect

"The SSIS on-premise components give us:

1. More flexibility for creating custom matching rules
2. Our version of the "golden record" criteria
3. Control over how the data is processed
4. Ease of updating the monthly address data by copy/paste the data file onto server in the appropriate directory without need for updating the SSIS package or re-deploy."

- Erin L. | Sr. Project Manager



## WHY MELISSA?

Data is one of the most valuable assets you have—accurate, high quality customer data empowers you to do business with anyone, anywhere in the world and deliver exactly the experience necessary not only to help drive sales, but to improve loyalty, trade, business intelligence, technology, logistics, and more. Since 1985, over 10,000 companies around the world have counted on Melissa to harness the value of their Big Data, legacy data, and people data (names, addresses, phone numbers, and emails) to drive insight, maintain data quality, and support global intelligence.

Here's what sets us apart:

- Free unlimited worldwide tech support
- 99% uptime on our servers
- Free trials with absolutely no risk

Time Inc.

BAE SYSTEMS

SIEMENS

NORDSTROM

Bank of America



FedEx



All-in-One Data Quality Components for SQL Server – Try it for FREE!

US

22382 Avenida Empressa  
Rancho Santa Margarita, CA 92688-2112

800.MELISSA (635.4772)

[www.Melissa.com](http://www.Melissa.com)

UK

+442030510140

INDIA

+918026632431

GERMANY

+493079788829

AUSTRALIA

+61280917500