

Melissa Data Helps Customers Strike Gold in Data

MELISSA DATA®
Your Partner in Data Quality

New Functionality in Melissa Data's MatchUp for SSIS Component Enables Intelligent Selection of the Golden Record

WITH THE ADVENT OF BIG DATA, information is flowing into organizations from a growing array of sources. As a result, the notion of a single version of the truth can seem more elusive than ever.

To help organizations arrive at a single, trusted version of their data sets, Melissa Data has introduced Golden Record Selection, new functionality in the MatchUp Deduplication component for Microsoft SQL Server Integration Services (SSIS).

MatchUp is part of Melissa Data's data quality toolkit for SSIS that enables sophisticated match logic to find matching entities in records. MatchUp includes a new Golden Record Selection and Survivorship tool used to collapse the duplicate records into a single record for loading into the Data Warehouse. Specifically, MatchUp's Golden Record Selection tool empowers users to identify similar or duplicate records and select the most appropriate surviving record based on the scoring quality of the contact data. This provides for knowledge based—survivorship, a departure from the rules-based approach using common techniques such as most frequent or longest.

The Challenge

Often, organizations will find that they have duplicate data in their customer records, explained Bud Walker, Director of Data Quality Solutions at Melissa Data. "If you have different versions of the same name—such as Joe versus Joseph, or Bill versus William, or perhaps a slightly different or misspelled address such as 12th Street vs. Twelfth Street,

or even a completely different address but the same phone number—these duplicates would be found with our MatchUp engine, based on configured match rules," he explained. "But what do you do then? How to bring them all together? That is really the question.

Melissa Data's new Golden Record Selection and Survivorship tool selects the most appropriate surviving record based on the scoring quality of the contact data.

That is where the Golden Record and Survivorship capabilities that we have introduced come in. The new Golden Record capabilities are specifically focused on identifying which record in the duplicates is the best record—the Golden Record."

In the past, this process was performed by the data steward with a generic set of rules identified in the governance process. The rules might have been built so that the customer record with the longest phone number or the record with the highest frequency of information in the duplicate group, or the most recent entry was considered best. "But what if you have five candidate records with a mismatched 10-digit phone number in the home phone field," said Walker. "How do you know which is the best one to keep? You can't base the

decision on the most complete, longest, or most frequent—techniques that are commonly employed—as they don't really give any insight into which phone number is actually callable, or if it is even owned by the individual in question."

Uncovering the Golden Record

To enable a more intelligent approach to the selection of a Golden Record, Melissa Data brings its data quality scoring based on the veracity of a piece of data in the real world. By using real-world reference data for that individual from trusted sources and scoring the data in all the candidate records, the data steward is able to hone in on the best address, the best phone number, the best email address, and then apply the common strategies normally used to assemble the miscellaneous information to construct the surviving record.

The Golden Record Selection and Survivorship tool is applicable to all verticals—including medical, insurance, banking, and retail—all industries in which organizations are dealing with large volumes of customer data and a golden record is of critical importance, noted Walker. "We see customers come to us every day with a different problem, but similar themes. This is a game-changer. It really does change the whole paradigm of selecting a golden record and executing survivorship." ■

For a free 30-day trial of Data Quality Components for SSIS, visit www.melissadata.com/ssis or call 1-800-635-4772.