

SQL Server Edition

MELISSA DATA Magazine

Why Your Data Sucks

*The Surprisingly Short
Half-Life of Data*

Attention
SQL Server Users!

Inside:
Everything
you need to
improve
data quality
Now

Data ROI

Leading internet
tech firm shares its
data quality story



Cover Story



04 Why Your Data Sucks

The Surprisingly Short Half-Life of Data

By Abby Telleria

Features

09

Case Study

AirNet Cleans and Consolidates Data Using Melissa Data SQL Server Solution

12

Application Story

The Ultimate Guide to SQL Server Data Quality

Learn the pros and cons of various ways to integrate data quality from shared procedures

by Joseph Vertido
and Bud Walker



16

Data Quality Consulting

Solutions Spotlight

06

Data Quality Plugins

SQL Server Integration Services (SSIS) and Listware

07

Developer Tools

Data Quality Suite, MatchUp, Personator, SmartMover, IP Locator, Geocoder, and Express Entry

08

Data Quality Services

Free
Demo



Try out all our capabilities using our Listware for Microsoft Excel plugin. **Free test.**

www.MelissaData.com/sqlmag

Data Quality Tools for Developers & DBAs

A better way to build in data verification

Since 1985, Melissa Data has provided the tools developers need to enhance databases with clean, correct, and current contact data. Our powerful, yet affordable APIs and Cloud services provide maximum flexibility and ease of integration across industry-standard technologies, including .NET, Java, C, and C++. Build in a solid framework for data quality and protect your investments in data warehousing, business intelligence, and CRM.

- Verify international addresses for over 240 countries
- Enhance contact data with phone numbers and geocodes
- Find, match, and eliminate duplicate records
- Sample source code for rapid application development
- Free trials with 120-day ROI guarantee

Melissa Data.
Architecting data quality success.



MELISSA DATA®

| United States | United Kingdom | Germany | India |

Cover Story //

Why Your Data Sucks

The Surprisingly Short Half-Life of Data

By Abby Telleria, *Senior Writer*



Your data is radioactive – just like a chunk of plutonium will break down in a measured amount of time – its half-life – data also has a half-life. That's the time it takes for half of the contact information in your database to go bad.

Why? Because contact data is always in flux. Customers move, change jobs, retire, die, get married or divorced – causing the stored data to become stale and out-of-date, affecting the accuracy and usefulness of data used for communications, analytics, and compliance.

We developed a formula to determine the actual half-life of data and analyzed thousands of customers' databases. Check out the infographic on the next page for the formula and our conclusions.

The Costs of Bad Data and the 1-10-100 Rule

If left unchecked and untreated, this decayed data will contaminate your entire database – leading to increased costs, inaccurate analytics, and endangering your ability to nurture and maintain customer relationships.

In fact, poor data quality is the main reason for 40 percent of all failed business initiatives, according to a Gartner, Inc. study. Poor data quality can affect labor productivity by as much as 20 percent, and put up to 12 percent of your revenue in jeopardy. In fact, bad customer data costs U.S. businesses over \$600 billion per year.

Consider the “1-10-100 Rule” which posits it takes \$1 to verify the accuracy of a customer record at point-of-entry, \$10 to clean it in batch form, and \$100 per record if

nothing is done at all. This includes the costs associated with undeliverable shipments, low customer retention, and unsuccessful CRM initiatives. The bottom line – it'll cost you more not to have a solution in place that will verify, cleanse, and guarantee that you have valid contact data.

This highlights the importance of data quality, and why you should continuously clean and update your contact data to ensure its accuracy.

Fight Truth Decay with Data Quality Tools

Melissa Data provides smart, sharp data quality tools to help you combat bad data – and keep it clean and up-to-date for your organizational needs.

Contact data quality is all we do – and we do it better than anyone in the industry. Since 1985, our foremost goal has been helping customers achieve the highest quality data at the most affordable price.

We understand that without clean, correct, and complete contact data, it's a challenge to achieve a winning sales, marketing, or business intelligence campaign.

We offer global address, phone, email, and name verification solutions, and identity verification and data enrichment tools to add demographic and geographic data – for better business intelligence, analytics, and sales and marketing initiatives. Our flexible, scalable solutions integrate easily with Microsoft – from our add-ins for SQL Server Integration Services (SSIS) and Excel, to our powerful .NET tools.

Our commitment to data quality excellence has been the hallmark of our success for almost 30 years. We partner with technology companies at the forefront of innovation. In fact, Melissa Data was one of Microsoft's first partners to offer Data Quality Services through the Microsoft Azure Marketplace.

What sets Melissa Data apart is the multitude of reference data sources at our disposal – including mortgage and property, demographic, business, geographic, USPS®, and international data sets that propel your business forward.

Discover for yourself why more than 10,000 companies worldwide rely on Melissa Data to gain a single, accurate, and trusted view of critical information assets.

For all your data quality needs, call 1-800-MELISSA or visit www.MelissaData.com for an office near you. Our offices are located in Rancho Santa Margarita, California; Rockwall, Texas; Raleigh, North Carolina; Berlin, Germany; London, UK; and Bangalore, India.

Calculating the Half-Life of Data – Also Known as the **RATE OF DECAY**

In the U.S. annually we have a population of 316 million, according to the U.S. Census Bureau.

Changes:

Marriages = 2.3 million



Divorces = 1.2 million



Births = 4.3 million



Deaths = 2.5 million



Moves = 47 million



Total Changes per year = **57.3**

Total Changes per month = **4.8**

The half-life formula is:

$$t_{1/2} = (t \ln 1/2) / (\ln mf / mi)$$

$$t_{1/2} = \ln (1/2) / \ln ((316-4.8)/316) \\ = 45.4 \text{ months or ...}$$

After **3 years** and **9 months**
half the customer records in a
database are **incorrect**.

Solutions Spotlight

You've learned how quickly your contact data degrades. Here's a quick guide to our data quality solutions to help drive better insight into your data, improve operational efficiencies, and position your company for growth. Don't forget to check out the different ways to implement our data quality tools in our Application Story on page 12.

Test Our Tools for Free



Our tools are available as multiplatform APIs, in the cloud, or as enterprise solutions. Request a free trial of any component or easily test our capabilities in Microsoft Excel through our Listware add-in. **Get a free trial now!**

www.MelissaData.com/sqlmag



Data Quality Plug-Ins



SQL Server Integration Services (SSIS)

Data Quality Components for SSIS is a suite of data transformation components for Microsoft SSIS to standardize, verify, correct, consolidate, and update your contact data. Each component is easy to use and features an intuitive interface with drag/drop capabilities.

Components include:

Contact Verify; Personator; Global Verify; SmartMover; MatchUp; IP Locator (See developer tools section for complete description)

Featured capability: Survivorship/Golden Record Selection

Melissa Data's MatchUp® eliminates duplicates and merges multiple records into a single, accurate view of the customer – the Golden Record – through survivorship.



Listware®

Listware is an Excel® add-in that verifies, corrects, and standardizes U.S. and Canadian names, addresses, phone numbers, and email addresses – directly in Excel. Listware adds missing names, addresses, phones, and emails to complete a contact record. The add-in also enriches your data with latitude and longitude coordinates, census, county, place, tract and block codes to addresses to improve business analytics and decision-making. Listware works with Microsoft Windows Editions of Excel 2013, 2010, 2007 and 2003.

Components include:

Global Check & Verify; U.S. Check & Verify; SmartMover; MatchUp; Express Entry (See developer tools section for complete description)

Featured capability: Global Check & Verify

Listware can verify, correct, and standardize address data from all over the world. It supports data in any UTF8 language, standardizes addresses to the mailing format for each specific country, transliterates major character sets, and enables output in native or Latin characters.



Data Quality Suite (World & North America Editions)

The Data Quality Suite verifies domestic and international address, email, name, and phone information, and provides precise latitude and longitude coordinates. The suite helps maintain accurate contact data to improve analytics and key strategic initiatives. The Data Quality Suite easily integrates into Web forms, CRM, and custom applications.

APIs or Cloud



Personator®

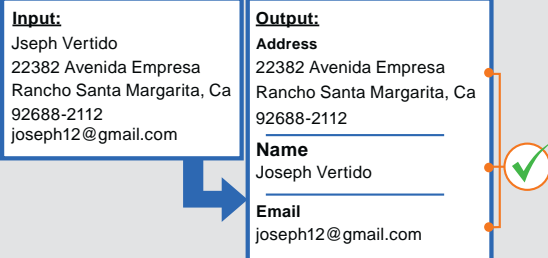
Melissa Data's flagship data quality cloud service, Personator, helps you make clear connections within and across data sets, to resolve different representations of a record, and link all touch points of contact data together to achieve a single, accurate, and complete view.

Cloud only

Personator Verify (US)

Name-to-address-verification:

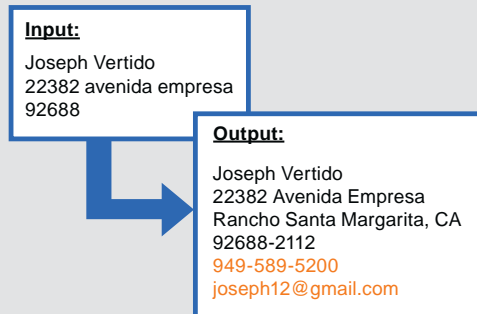
- Compares data with multisourced data sets to determine accuracy
- Verifies name, address, phone number, and email address all correlate to same individual
- Geocode address to rooftop level (lat/long)
- Move Update (10+ years)



Personator Append (US)

Completing your records:

- Add missing email, phone numbers, names, and company names to your records
- Enrich records with demographic data



Express Entry®

Express Entry utilizes a type-ahead search functionality that makes it easier for users to enter addresses with up to 50 percent fewer keystrokes. Once you type an address into the search bar, Express Entry will complete your query with a verified address, including city, state, and ZIP Code™ for U.S. addresses, locality, administrative area, and postal code for international addresses.

Cloud only

Solutions Spotlight// At a glance



MatchUp®

MatchUp is a fast, powerful advanced matching tool to help identify the most difficult-to-detect duplicate records. MatchUp uses advanced fuzzy matching algorithms, coupled with deep domain knowledge, to granularly identify matches between names and nicknames, street/alias addresses, companies, cities, states, postal codes, phone numbers, email addresses, and other data components. Consolidate your data using MatchUp to reduce costs and dramatically improve your organization's operational efficiencies.

API only



SmartMover®

Use SmartMover to get the new addresses of your contacts and reduce wasted materials and postage associated with undeliverable-as-addressed mail. Our service compares your mailing list records to more than 160 million address changes in the USPS® NCOA^{Link}® database, representing all permanent change-of-address records filed in the most recent 48 months. Melissa Data also offers Canadian National Change-of-Address processing for the most up-to-date information available on changes of address filed by Canadian households and businesses over the previous 72 months.

Cloud only



Geocoder

Use Geocoder to assign precise latitude and longitude coordinates to international, U.S., and Canadian addresses for a better understanding of your customers based on location, and to improve business analytics, logistics, and critical decision-making. Geocoding incorporates spatial data from multiple data sources – a process called conflation – to deliver the most accurate rooftop/delivery point available.

API or Cloud



IP Locator

IP Locator helps you identify a visitor's geographical location, i.e., country, region, city, latitude, longitude, ZIP Code™, using an IP address lookup technology that incorporates some 20 different techniques to determine the physical location of an IP address. The level of geo-location precision makes IP Locator ideal for e-commerce, website tracking, and delivering relevant content to users.

API or Cloud



Data Quality Services

Microsoft Azure
Marketplace

Melissa Data provides easy access to high quality datasets – from Name, Phone, IP Location, Email, Address Verify, and ZIP Data – through the Microsoft Azure Marketplace. Learn more here:
www.MelissaData.com/sql-dqservices

AirNet Cleans and Consolidates Data Using Melissa Data SQL Server Solution



About AirNet

Founded in 2002, AirNet Group Inc. offers solutions designed to give your business-sensitive information a secure home. AirNet is a leading provider of complex hosting, network services, and advanced Internet applications for enterprises that outsource their IT infrastructure. The company offers a wide array of services including Internet connectivity, disaster and data recovery services, audio and video streaming, VOIP services, Web hosting, co-location, and broadcast email services.



“The solution’s ease of integration has also expanded our ability to provide customer-specific, data-based applications – ensuring that all data entering the applications is uniform, standardized and cleansed.”

- Jeff Averbek
CEO of AirNet

Challenge

For years, AirNet relied on applications developed by external vendors for their data hygiene, enhancement, and matching of data.

“Our previous vendor of data quality tools was a large software provider for whom their data quality tools were just another product off the shelf,” said Jeff Averbek, president and CEO of AirNet.

“Integration within our proprietary applications – and even use of their product – was only done by select consultants ... there were limited resources and a shallow knowledgebase for the tools. Further compounding the problem was the instability within the ownership of the tools – the product having been purchased by three different companies (with different visions for the product) over the course of a few years,” Averbek said.

The company’s reliance on external vendors proved to be a challenge when implementing its smartCommunicator® product – an award-winning communication portal that allows individuals, businesses, and agencies to share information from diverse sources using a single, unified media platform. The company’s portal is designed to effectively deliver messages to its targeted audience.

The main challenge? Integrating disparate data sources into a data-centric hub for the purposes of consumer and constituent engagement, Averbek said.

It was essential to utilize a data quality solution to help AirNet gain a single, accurate, organized view of their customers by consolidating data across different data sources.

The company also needed a strong technical support team that would aid in the implementation of a data management solution, and provide quick responses to any issue that would arise during the transition process.

Solution

Data Quality Components for SQL Server

Key Benefits

- Easy to install, easy to configure, and easy to use in SQL Server Integration Services
- Integrate data easily across multiple systems
- Consolidate duplicate data for a 360-degree view of your customer
- Treat data as a product: validate, clean, and sell enriched data
- Increase the value and trustworthiness of your data for improved analytics and business intelligence

Results

AirNet chose to implement the Data Quality Components for SQL Server – a complete enterprise platform that works within the SQL Server's Integration Services (SSIS) data flow to deliver a wide range of data integration, transformation, cleansing, and enrichment functionality.

The solution resides at the hub of the import tool for the smartData module within the smartCommunicator application.

“Name, Address, GeoCoder, Phone, and Email verification components, or transforms, are all put into action as each individual data upload is scrubbed and merged into a single, uniform data set. This creates a single anchor record per individual from which our customers operate multi-million-record mass communications strategies,” Averbeck said.

“The solution's ease of integration has also expanded our ability to provide customer-specific, data-based applications – ensuring that all data entering the applications is uniform, standardized, and cleansed.”

Integrating the components into the organization's workflow was seamless.

“Rather than spending countless sums of money on training and specialized staff, it has been a breeze getting staffers who are already familiar with SQL's tools up and running with the new toolkit,” he said.

Utilizing Data Quality Components for SQL Server proved to be a successful venture for AirNet.

“Processing data within existing SSIS jobs – rather than relying on unstable Web-service /Command Prompt calls – has increased our productivity and greatly stabilized the application as a whole,” Averbeck said.

Averbeck added, “The solution's existence within the SQL Server Integration Services suite has enabled the company to build robust, stable ETL tools – in many cases, customizing data transforms to the specific needs of our clients.”



“

Processing data within existing SSIS jobs – rather than relying on unstable Web-service / Command Prompt calls – has increased our productivity and greatly stabilized the application as a whole.”

- Jeff Averbeck
CEO of AirNet

The *Ultimate* Guide to SQL Server Data Quality

Learn the pros and cons of various ways to integrate data quality from shared procedures

By Joseph Vertido and Bud Walker

SQL Server technology offers quite a few options for integrating third party solutions. The question of which path to take, in most cases, is determined by preference in architecture. In this article, we will look at the commonly used ways to integrate Melissa Data's flagship Data Quality Solutions within the SQL Server Environment, and point out some of the pros and cons of each type of integration.



Pros

1. Easy and straightforward integration
2. Supports all legacy versions of SQL Server

Cons

1. In-process memory tends to be constrained by surface area configuration

I. Transact-SQL OLE Automation Objects

OLE Automation Objects are Microsoft's legacy way of integrating T-SQL batches, stored procedures, and triggers with a COM Object's IDL interface. Melissa Data supports this method of integration mostly for the sake of legacy versions of SQL Server, and enables the registering of our COM Object library interfaces for use with the sp_OACreate. Melissa Data advocates this as the less-desirable form of integration due to in-process memory constraints when other, more modern approaches are available.

II. Extended Stored Procedures

Stored Procedure implementation is probably the most common way users call the Melissa Data APIs. The Stored Procedure environment allows for accessing third party DLLs and it offers a quick and easy way to connect to your SQL Database and iterate your records through the Data Cleansing regimens.

The Melissa Data APIs are bundled with SQL Stores Procedure Sample Code that demonstrates how to interface with the Standard DLLs. Wrapper functions are automatically generated for you which allow for direct access to the APIs.

Pros

1. Easy and straightforward integration
2. Edits to the SP are quick to make

Cons

1. Production on a 64-bit environment is a must
2. DLL management is not optimized within SQL Stored Procedures

III. Common Language Runtime (CLR) Integration

Probably the least common type of integration method that we have encountered from our customer base would be CLR integration. That is not to say, however, that it is the least effective.

A CLR Integration allows for calling .NET Code from within SQL Server to perform transformations to your data. This allows for performing tasks that may have been previously impossible due to the limitations of queries. A CLR Project can be created by selecting the "SQL Server Project" in your Visual Studio. After which you can deploy the project, which can then be executed through your SQL Server Management Studio.

The Melissa Data APIs can be called through the .NET project in your CLR Project as opposed to calling it through an Extended Stored Procedure. It offers a better and more stable platform when working with DLLs and is a highly recommended approach when using Melissa Data APIs.

Pros

1. Highly stable environment
2. Allows you to leverage the flexibility of .NET

Cons

1. Slightly more complex type integration

IV. SQL Server Integration Services (SSIS)

Another popular method of implementation would be through SSIS – an ETL/Data Integration Platform bundled with SQL Server. Melissa Data provides custom, GUI-based components for SSIS that are accessible through the Data Flow Task.

Melissa Data's integration with SSIS is seamless and probably the easiest to use as it makes use of an interface and does not require any type of coding. A typical SSIS project will have an ETL (Extract Transform Load) foundation and the data quality regimens are usually performed during the transform process. SSIS allows you to retrieve data from a source, clean the data using the Melissa Data Components, and output the resulting data through its interface. Melissa Data supports both 2008 and 2012 versions of SQL Server.

Pros



1. No programming required
2. Ease of connection to multiple database formats
3. Simplified multithreading

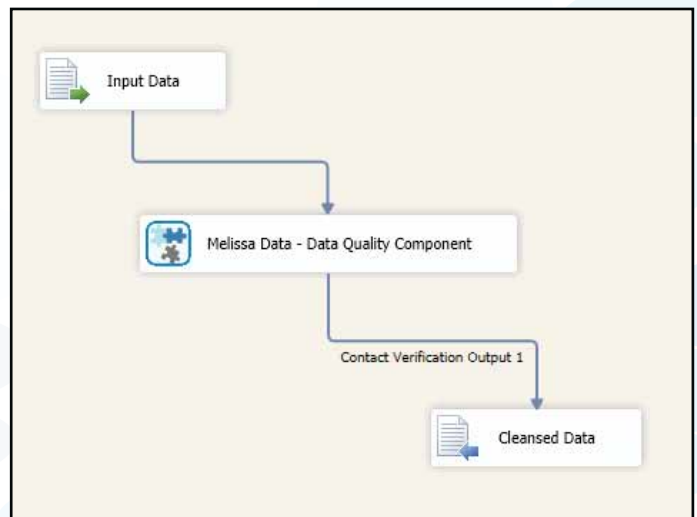
Cons



1. Data quality processing is confined within SSIS

About the Authors

Joseph Vertido is a data quality analyst for Melissa Data and Bud Walker is the company's director of data quality solutions. For more information for any of these types of integration, call us at 1-800-MELISSA or send an email to joseph@melissadata.com.



V. SQL Server Data Quality Services (DQS)

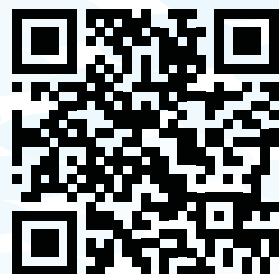
Bundled with 2012 SQL Server Enterprise Edition is the Data Quality Services (DQS) Platform – a data quality solution developed by Microsoft. DQS provides a new perspective in performing data quality by taking a Knowledge Driven approach. The use of Knowledgebase allows for cleansing data through the use of back-end reference data sets, allowing for a more data-centric approach to data quality as opposed to fuzzy correction and regular expression cleansing.

More importantly, DQS offers the use of Third Party Reference Services through the Microsoft Azure Marketplace – which offers a diverse selection of third party data quality services, including Melissa Data's flagship suite of Data Cleansing Solutions.

The Melissa Data solutions in the Microsoft Azure Marketplace can be consumed within the DQS Platform. This can be done by creating a new Knowledgebase and adding the Melissa Data service to the reference data, allowing you to run your data through the data cleansing process.

To watch a video of how to consume Melissa Data in DQS, go to:

<http://www.youtube.com/watch?v=U9GhZ2vAysw>

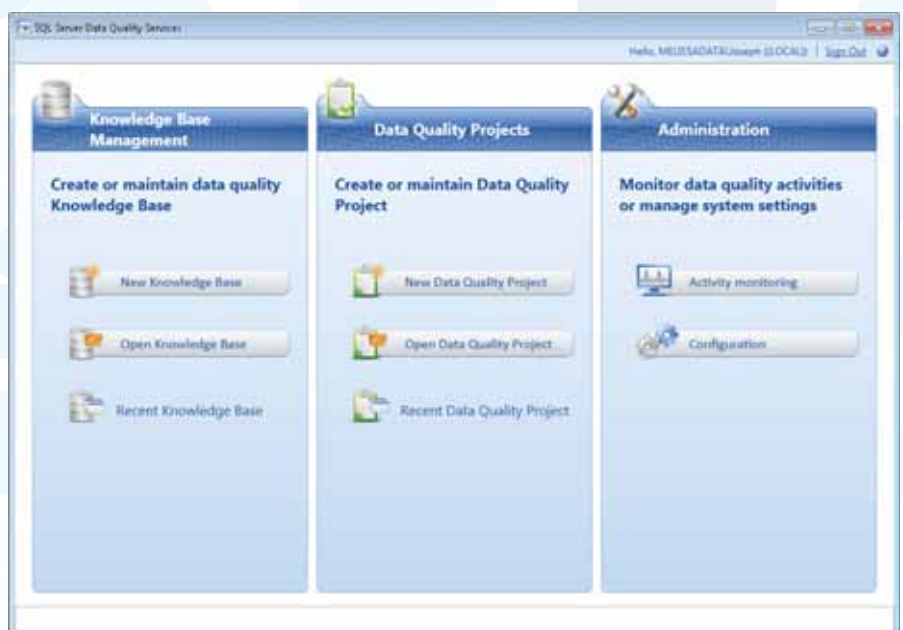


Pros

1. GUI-based implementation
2. Leverage additional DQS cleansing capabilities
3. Does not require SQL Server Knowledge for data cleansing implementation

Cons

1. Web service only access to third party solutions
2. DQS Projects cannot be executed through SQL Management Studio



MELISSA DATA CORP.
22382 Avenida Empresa
Rancho Santa Margarita
California, 92688-2112

To correct or remove your name from this list, fax your request to 949-589-5211



Unlock the Power of Your Data

with a Full Range of
Professional Services

Let us help you jump start your data quality initiatives with expert advice to get you up-and-running quickly – plus, develop an action plan that grows with your data needs. **Our Contact Data Quality Consulting service will help you:**

- **Identify your organization's data quality issues**
- **Learn how our tools clean, correct, and complete your contact data**
- **Empower better business decisions with the highest quality data**

Talk with us today about your unique technical needs and we'll work to ensure you have data you can trust for every need – every time.

Propel your business forward.

www.MelissaData.com/consult