melissa

A **MELISSA** WHITEPAPER

Golden Records are Key to Solid Data Quality

by Stephen Wynkoop
SQL Server MVP, Founder and Editor at SSWUG.org

TABLE OF CONTENTS

INTRODUCTION

With the masses of information flowing through applications, from other applications, in addition to end-users, getting "clean" information that is correct and valid is more important than ever.

The costs associated when working with incorrect or incomplete information are substantial. From transaction validation to analysis and identification of trends and important milestones, not having correct and valid information is a risk that can cause major headaches.

Data Quality works to resolve this issue, but how you go about that process of identifying and correcting information is extremely important. You can just as quickly invalidate a piece of information by applying incorrect approaches to "fixing" it. Conversely, proper approaches to information management significantly increases the value of that information.

Data Quality refers to the reliability, completeness, and accuracy of information flowing through your systems. This often applies to the personally identifiable information elements, like address, email, and phone. You may also find that data quality checks need to be applied to transactional and application-to-application information, making sure transaction values are valid and complete.

Data Quality is critical because, for example, it creates a basis for working more effectively with your customers. This may be in response to a transaction that has occurred, or the follow-up contact. In any case, having a proper email address, physical address, and other contact information is crucial. Mail may be delayed or even rejected if addressed incorrectly.

One challenge is the correction of, and use of, international addresses. The formatting, the specifics of the address, along with simply having the right address for the right person are all key to success. Not having these elements means costly returned mail or failed transactions.

THE WORLD IS WATCHING

As online systems continue to be the norm, correct address information becomes more challenging to support. This, added to the fact that it's typical for a customer to be entering their own information, rather than having an interim operator entering the information, means that the address may have odd abbreviations, typos, or missing elements.

Customers and other end-users and stakeholders expect you to have the correct information stored and used, even if it's entered in a manner that may be less than optimal. It's up to you to clean and correct information, preferably with confirmation from the customer/end-user, prior to storing and using that information.

From an application standpoint, this presents a real challenge. How can an application possibly know which, if any, of the addresses on file for a customer are correct? What changes are needed? How should corrections be suggested? These are key questions and challenges that your approach to Data Quality should carefully address.

BROAD APPROACHES TO DATA QUALITY

One of the first approaches to making sure you have the right information, and don't have conflicting information across data sources, is to use a single-source approach to information. As customers enter their information, for example, you store it in a single location. Later, if the customer updates their information, they are updating the master record, and all uses of that information are updated.

This works, as long as the initial data entry is correct. You can also provide filters and structured input as well. For example - a list box of states and countries, formatting on phone numbers, checking email addresses for obvious mistakes – each of these is a key component to a user-friendly input form. They don't, however, provide the correction of improper formatting or an outright invalid address.

Typical Data Quality is done by validating against either internal or external systems. These systems will consider the data you've received, provide correction, or issue notifications to your application or flow. With Melissa's specific implementation for Microsoft® SQL Server™, you have options to build in a data validation process using your SQL Server Integration Services (SSIS) project flow. This validation will work through the data presented and correct it automatically or provide exception reporting and suggestions.

It's important to understand that different systems work with data at varying levels of comprehensiveness. There are several key decision points that will help determine the type of solutions you could consider. Here are a few questions that can help determine a best approach for your systems:

- Do you want to do your validation in real-time as the user submits their form or other information?
- Do you want to integrate your validation routines at the server (SQL Server) level?
- Will you have both domestic (US) and International addresses?
- Is there a chance of conflict between different versions in your systems for the same addressee?

As you build your applications and solutions, you will need to determine your tolerance for information differences.

What's the Risk?

Perhaps one of the most costly mistakes is to ignore the need for managing contact information. The risks run from the obvious and clear costs of returned mail, invalid contact information that renders accounts meaningless, and the labor and processing costs to work with these invalid accounts, to the customer impression of your business and its ability to work with this information.

When customers perceive that information is less valuable to you, doesn't require your attention and management, and is accepted when invalid, they tend to have a less complimentary impression of your operations, your service, your product. Add to this the different regulatory compliance requirements and you can quickly see that the benefits of a solid, comprehensive data quality approach are significant, while the costs of not managing your information can be substantial.

GOLDEN RECORDS – NOT JUST VALUABLE, INVALUABLE

The concept of Golden Records encompasses the art and science of selecting the "true" data and using it to correct incomplete or incorrect information. This process is one of finding records that are clearly referencing the same entity, but have (for example) differing addresses.

A Golden Record is the record that is deemed to be the master correct record. Using the Melissa approach, the Golden Record is determined by a data quality score that takes into consideration the level of quality of the information provided and uses that as a basis of survivorship. Once found and validated against reference databases in real-time, the information is used to update other instances. This assures you have consistent information across all instances of the data. At the same time, it corrects all of the different occurrences, making all of the elements actionable and correct.

Golden Records are determined by looking at online databases of address information, known information for individuals, companies, and other entities. This, along with other transactional information, can be used to determine issues and provide updates and corrections. The SSIS process can update and correct the other rows and "clean up" data as it flows into your processes.

There are many different ways of determining Golden Records – it's important to understand the rules that are applied, how the data is manipulated, and what types of control you have over how the rules are applied. In some systems, simple filtering is used. In other systems, it's assumed that the most recently entered information is correct. In still other cases, issues can arise with international addresses or mixed addresses. There are many options when it comes to automating your solutions. It's critical that you're able to determine how Golden Records are established, validated, and applied to your own information.

92%

of businesses admit they have duplicates in their systems

More than

70%

of businesses believe a single view of the customer would lead to cost savings

Only

15%

of businesses have an effective plan in place

Melissa provides a comprehensive set of capabilities for the cleansing and deduplication of contact data. Full integration with SQL Server SSIS helps to leverage your investment in Microsoft technologies. Here's a look at the key features and functions included:

Global Verify

- Verify, standardize, & parse addresses for 240+ countries & territories
- Supports data in any UTF8 Language

Contact Verify

- U.S./Canadian address verification, standardization, parsing, & enrichment
- Rooftop-level geocoding
- Email & phone verification
- Name parsing & genderization

SmartMover℠

- Update addresses of individuals, families, & businesses that have moved
- Search up to 48 months of move-update records

Personator®

- Authenticate an identity
- Correlate verified contact data to a single individual
- Append accurate, current, missing contact info

MatchUp®

- Find & consolidate duplicate records within or across multiple datasets
- Match using multiple contact-specific or fuzzy matching algorithms
- Advanced matching capabilities like householding & geographic proximity
- Automated Golden Record Selection

Fuzzy Match

- Perform fuzzy string matching with over 15 algorithms
- Industry-accepted plus Melissa proprietary algorithms
- Profile & Monitor
- Identify data quality issues for immediate attention
- Monitor conformance of source data to specified requirements of pre-set limits

For a free trial of Data Quality Tools for SQL Server, visit:
www.Melissa.com/sql or call 1-800-635-4772

About Stephen Wynkoop

Stephen Wynkoop is the founder and editor of SSWUG.ORG – the SQL Server Worldwide User's Group where he writes a column and maintains the site. SSWUG features weekly video programs about the database and IT world, webcasts, articles, online virtual community events, and virtual conferences several times a year. Stephen is a Microsoft SQL Server MVP and the author of more than 10 books, translated into at least 7 languages. Stephen has been working with SQL Server since the very first version, with a prior experience in database platforms that included dBase and Btrieve.

Stephen can be contacted at **swynk@sswug.org**.

melissa

www.melissa.com

About Melissa

Since 1985, Melissa has specialized in global intelligence solutions to help organizations unlock accurate data for a more compelling customer view. Our breadth of data and flexible API technology integrates with numerous third-party platforms, so it works for you and makes sense for your business. More than 10,000 clients worldwide in key industries like insurance, finance, healthcare, retail, education, and government, rely on Melissa for full spectrum data quality and identity verification software, including data profiling, cleansing, matching, and enhancement services, to gain critical insight and drive meaningful customer relationships.

US

22382 Avenida Empresa
Rancho Santa Margarita, CA 92688-2112

800.MELISSA (635.4772)

UK

+442030510140

GERMANY

+493079788829

INDIA

+918048540142

AUSTRALIA

+61280917500