firstlogic White Papers

Data Cleansing for Improved Customer Loyalty

Published July 5th, 2022

We are in the age of personalization. Whether you call your audience customers, donors, members, constituents, or something else, you are probably looking for ways to improve the effectiveness of your messaging with a personalized approach. Accomplishing that feat requires companies to ensure the data they use during customer interactions is accurate and suitable for the communication.

Data is coming into your organization from many sources. Each department may maintain its own customer database stored in multiple formats under the control of various systems. Even more data may come from outside your organization through business affiliations or purchased lists. Your company may have captured data via consumer-completed online forms. Other information comes from telemarketers, mobile apps, or data entry operators.

In short, your data is probably a mess. When you use that information to generate personalized communications, or to segment customers and prospects, all that ugly data is likely to yield embarrassing mistakes and wasted efforts.

Data quality affects more than just documents. For example, a cruise ship company requires passengers on international cruises to present passports or other identifying information when they board the vessel. Cruise company employees at the dock compare the documents passengers bring with them to names on their reservation lists. If the names don't match, passenger boarding can be delayed or even denied, a severe downturn in the customer experience! For repeat customers, difficulties on embarkation days can be avoided by comparing the passenger-supplied names on the reservations to the legal names on file with the cruise company. By adding legal names to the company's reservation list, the company minimizes the impact of passenger name discrepancies.

In this whitepaper we'll touch on ways you can clean your data and explain why that's such a good idea. We've got plenty of experience in this area. Our clients have been making their data work for them using our data-enhancing software,

DataRight IQ® and its predecessors, for decades. DataRight IQ® is the general-purpose data cleansing component of the Firstlogic Data Quality Suite.

Besides personalization, data enhancing practices benefit businesses in other ways, including operational efficiencies, reliable reporting, and forecasting. If you rely on data to run your business (and who doesn't?), making sure that data is accurate, consistent, and relevant is vitally important. Data cleansing is necessary to produce the best results.

Communicate Like a Real Person

Addressing a customer by name seems like such a simple thing, but when data is inconsistent, accomplishing this task at scale can become a nightmare. Making it even tougher, each time you use a customer name, a different format may be suitable. In casual communications, using only a customer's first name may be acceptable. In other cases, such as when matching records among databases, data analysts must include first, last, middle, suffixes, and perhaps titles.

Working with customer name data requires expertise and patience. The variety is tremendous because each data file or application may treat this most basic and personal item completely different from the next.

Creating Personalized Greetings

Programmers can spend days writing routines to construct the first line in a letter. Everything after the word "Dear" is variable. There are so many exceptions, creating salutations is demanding work.

Unless the data has been accurately parsed and standardized, one cannot be assured the organization has stored customer names in a single, consistent format. Assuming data is formatted identically throughout the file can cause glaringly obvious errors.

A record where the last name and first name are reversed can generate salutations like “Dear Mr. Bob”. Data where two first names are stored in a single name field might generate a letter beginning “Dear Mr. John and Susan”. Missing data can cause communications to start with greetings missing the salutation entirely. Some programmers avoid the problems associated with unparsed or inconsistent names by excluding personalized salutations entirely. This approach exposes inconsistent data spectacularly and results in impersonal greetings like “Dear Customer” or “Dear John M Smith III VP Sales Central Region”. An incorrect greeting sets the tone for the rest of the communication. Customers notice their names. Failing to greet a customer appropriately weakens any further communication or call to action the communication conveys.

Gender Appropriate

Using customer names to identify gender is tricky. Many names are gender neutral, used widely by both men and women. Others are unusual and can’t be used to assign gender. First name initials are not helpful.

Considering the number of first names in common use in the United States, building your own tables of male and female names could take years to compile and still be incomplete. Even if programmers built such tables, they would encounter non-matches. Programmers would also find it necessary to develop algorithms to compute the confidence level of gender selection, making decisions based on the strength of the confidence score.

Gender identification is important. Businesses may format documents differently for male and female recipients. Text, offers, colors, and image choices may all be gender-specific. Gender mistakes can anger and alienate customers. Poorly targeted messages fail to produce desired results.

First Name? Last Name? Other?

A common condition encountered when working with customer names is the order in which names are stored. When names are not parsed into separate components, some names may be stored as First Middle Last. Others are likely listed as Last, First Middle. Commas may separate the last name from the other name components but multiple commas, such as those preceding a suffix like ", Jr." or professional designations can make it difficult to identify name components.

In most organizations securing IT support is a challenge for all but the most critical projects. It can take months before an organization allocates technical resources necessary to write nameparsing routines. Often, companies cannot wait. They forge ahead with their ugly data, consequently damaging individual customer relationships by mishandling their names.

Multiple Names

If your goal is to communicate with customers through automated methods as if they were oneon-one conversations, you probably wouldn't address them as "John Smith, Mary Smith, Ronald Smith, Caroline Smith, Eric Smith". You'd refer to this group of individuals as "The Smith family".

Data cleansing software can make this possible.

Many variations of multiple names require a comprehensive set of rules to allow the software to make the right decisions. A married couple sharing the same last name might be "Mr. and Mrs. Smith" but such treatment would be wrong for unmarried couples, same sex couples, or married couples with different last names.

Multiple name logic is important with householding strategies where organizations save materials and postage by sending only one copy of a document to a household of related individuals. Changing the format of the address appearing on the envelope or label is critical to inform recipients they are receiving just one document to share. Simply sending material to the first household member on the

list may leave the other customers wondering why they weren't contacted. Company responses to customer service calls inquiring about missing documents are expensive as customer service representatives attempt to diagnose the problem or initiate operational investigations.

Company Names

Company names can be just as challenging as person names, particularly when attempting to merge data from multiple sources. An average merging routine might not realize "General Motors", "GM", "Gen. Motors", or "General Motors, Inc." are all the same company. Great data cleansing software, however, will include extensive tables and dictionaries allowing it to connect these companies even when the actual data differs.

As organizations strive to create unified views of their customers, connecting data assembled by different departments is very important, but also very difficult. Companies build data silos with different database systems and file formats. Mergers and acquisitions introduce more data styles and make normalizing data even harder.

Use comprehensive data cleansing tools to standardize separated data sets before building a single source of customer data. Preparing the data before merging is a move sure to save time and avoid costly mistakes.

Intelligent Casing

Do you remember the first computer letters when companies printed variables in ALL CAPITAL LETTERS? The data entry devices of the day had no lower-case keyboards, so companies typically entered and stored data in all caps, resulting in that odd-looking presentation.

We've come a long way since then, but capitalization and casing are still issues. Organizations still store some data in all caps. To avoid the humiliation of creating communications that look like they produced them with forty-year-old technology, companies must convert data to upper and lower case. The rules for casing can be complicated, especially today when so many acronyms have made it into our everyday language. Company names and products may be referenced by initials or unusual casing. You wouldn't want to refer to Apple's newest product as an lPhone (with a capital "i"). Names with multiple capital letters (sometimes preceded by a space or special character and sometimes not) are difficult to case correctly using only algorithms. The best data cleansing software products use data dictionaries to handle the exceptions and offer proper casing.

Vernacular in any industry can trip up a text casing routine. Data cleansing software normally handles this challenge by providing industry-specific dictionaries and allowing users to define their own custom references.

Use What You Know

The key to success with personal customer communications is acquiring data about individuals. Use that information to create relevant messages – just as you would if speaking with a customer one at a time.

Companies can't dedicate human employees to compose individual customer communications. They can, however, make automated communications better at recognizing individual customer attributes. Somewhere in a large organization is stored information describing which products a customer has bought, how much they spent, when they made their last purchase, etc. Company data also shows which communication channels customers prefer, if they pay on time, and how they pay their bill.

Collecting customer-specific details is easy. The hard part is linking all that data and using it to correspond in meaningful and personal way.

Data cleansing software helps in this effort in two ways: 1. Parsing, standardizing, and normalizing data from disparate corporate and outside databases. 2. Presenting the normalized data in a way that seems more personal and less mechanical. Personalized messaging is definitely more effective than generic documents, but many pitfalls await companies attempting to create a personal approach without attending to the quality and relevance of their data. Personalization mistakes are easy to spot and cause customers to dismiss communications as insincere. Use data acquired about customers to strengthen the connection with them. Just make sure communications sound like they've come from a human, not a mass-communication robot.

Making It All Look the Same

Converting data stored in multiple formats to a common design and layout is a necessary task in many applications. Whether using the data to generate customer communications, statistics, reports, or operational improvements, a consistent and reliable data format is essential. Few organizations enjoy a single central data source. Most acquire data from multiple software systems designed for specific purposes. The data they collect and store reflects each application's function. Any undertaking involving the merging, matching, or comparison of data requires a preliminary step to standardize the data before the real work can begin.

Corporate databases are likely to contain differences in name formats, capitalization, abbreviation, and data elements stored. Connecting records from all the sources is impossible until someone standardizes the data.

Floating or Fixed

Conversions are easiest if the format of names, dates, and casing is known. This only occurs if data was input under strict rules and controls that enforce a consistent data file layout.

Unfortunately, firms create many files with loose rules or multiple specifications. User-supplied data from web forms, outsourced telemarketing firms, and multiple internal departments may have added or changed the data. File structure definitions are not always available or accurate, making it difficult for organizations such as outsource service providers to process the information consistently.

Search, Replace, Split

Anyone working with data files, especially files from outside their own organization, must have comprehensive tools to handle a wide variety of tasks. Every file may be different, and analysts are rarely informed of data organization nuances in advance. Analysts use search and replace functions to alter the contents of specific data elements. One may wish to change all company names in a file after a merger for example, the domain names on email addresses, or change “Occupant” to “Current Resident”.

Search and replace can also remove unwanted information. One may search for dates in a name field, for example, and replace them with blanks. Search and replace functions often support wildcards, making it possible to search for a pattern. One could define a pattern matching each vehicle manufacturer, for example, and replace vehicle identification numbers (VIN’s) with manufacturer makes and models. Split functions are especially useful for separating data fields with mixed information into discrete fields. Common uses for splits are constructing sort keys or dynamic file naming.

Another common split function use is isolating names that might appear in a single field. “Mrs. Mary and Mr. Robert Smith” might become two name fields: “Mrs. Mary Smith” and “Mr. Robert Smith”. Using a split function, analysts can reorganize the data into usable formats.

Titles and Abbreviations

Professional titles and abbreviations are some of the toughest data normalization challenges.
Failing to clean the data produces inaccurate record selection, poor segmentation, and embarrassing mistakes. This often comes into play within market segments. Medicine, for example, is overrun with abbreviations for medical conditions, facilities, treatments, drugs, and therapies. Professional designations for people working in the medical field affect the expectations patients have concerning areas of expertise.

Consider a health insurer, for example, that creates dynamic provider lists for plan members to choose physicians or facilities. Each hospital, clinic, lab, therapy provider, or individual must be identifiable by their names or professional designation. Entities recorded in the database with non-conforming abbreviations or professional titles will be excluded from the provider lists and denied access to new patients seeking their services.

Non-Party Data

Non-party data is information coming into an organization from the outside, frequently encountered with mergers and acquisitions. Merging data from two or more payroll systems, for example, can be hampered if one system uses dashes in social security numbers and tax ID numbers and another system does not. Some systems store only the digits, with an extra field used to designate whether the number is social security or tax ID. Other examples might include differences in account or policy numbers. When one system fills numeric data fields with leading zeros and another truncates them, merging the data together could cause problems for application programs expecting a certain data format.

Each industry has a set of abbreviations and terminology relevant to their business, but individual organizations may use different variations. Any application that tracks products from multiple sources has difficulty comparing or filtering data when the abbreviations are inconsistent. Imagine trying to filter computers based

on their installed optical disk drives abbreviated as CD-R/CD-RW, DVD-R/DVD-RW, DVD-RAM, or DVD+R/DVD+RW, including all the variations about where slashes, dashes, or plus signs could be used! Data cleansing software must supply or allow for custom data dictionaries to meet the specific needs of organizations dealing with unique sets of titles and abbreviations.

What if it's all mixed up?

Information in data files may float randomly among several data fields. If so, writing custom routines to find and reposition data can be tedious and time-consuming. It may take several passes of the file to accomplish the objective.

Custom-written routines may not handle unexpected conditions, creating production files with errors. Organizations facing such challenges often rely on smart tools that analyze data in context, find data elements, and reformat the information into preferred configurations. The best data parsing and cleansing tools can identify data elements by field contents and proximity to other fields, making this software very powerful.

OK, It's Clean. Now What?

Once organizations standardize and clean their data, they are ready to put it to work. All the effort to normalize data pays off in high-value communications, accurate reporting, and operational efficiencies. Comprehensive tools like data cleansing software removes much of the complexity and exception handling often required by user-written programs. Programmers won't have to include logic to construct greetings, perform case conversions, assign gender, standardize addresses, etc. The data cleansing process will do most of that work for them. As a result, custom routines become less troublesome and easier to maintain.

Dynamic Content

Perhaps the most valuable benefit of working with standardized data is the ability to generate variable- driven messaging. With better quality data, document producers are no longer forced to communicate in generalities. Data files filled with customer-specific information about account status, longevity, buying history, and more allow organizations to create true one-to-one messages bound to produce higher response and conversion rates. Companies can demonstrate customer value by addressing customers by name and acknowledging their past business transactions. They can communicate in ways that are interesting, relevant, and compelling.

Personal data such as gender and buying information, is useful for crafting messages that connect with each individual. When companies standardize and merge their data, female runners should receive ads about women's running shoes from the sporting goods store, for instance, not announcements about a sale on men's hunting boots. Marketers should expect an increase in positive response after they use clean data to create more relevant messages on an individual customer level.

Segmentation

Even when messages are not personalized, organizations can break data into groups of similar customers based on standardized and cleaned data. Examples might include charitable organizations that may segment high-dollar donors and send them more elaborate mail pieces, while modest givers receive the standard package. Or a real estate company may send families with children a brochure stressing a new neighborhood's proximity to good schools and parks while a different marketing piece tells older citizens about features like golf, dining, and condos with no yard maintenance.

Nth Record Select

Tests are an important part of direct marketing. The nth record selection function of data cleansing software allows marketers to create sample files representative of the overall database, send multiple versions of offers, and track each version's performance. Marketers can analyze results from A/B tests, assured the samples in each group are statistically similar.
Nth record select is also useful for breaking large jobs into smaller batches or drops, each containing customers sharing like characteristics.

Match-Ready

Data analysts cannot match records accurately until they standardize the data. Work performed by data cleansing software eliminates data differences and anomalies, making it possible to pair customer account data with valuable information stored in internal or external databases.
Comparing customer buying history with demographic data about home ownership, for example, may allow companies to send catalogs or ads filled with products most likely to interest the target audiences. Data quality processes can help an organization match purchases to customers, attributing products bought by Bob, Robert, Bobby, and Rob to a single individual even though they used different versions of their name in each transaction.

Operational Benefits

Standardized processes increase productivity and reduce the chance of errors. Data used to drive those operational processes must be consistent and correct. Measurement and reporting rely on consistently correct data as well.
Messy data can hamper information technology efforts, resulting in systems that take longer to design and are difficult to manage. Downtime caused by incomplete or poorly formatted data ripples through an organization and affects operations across the enterprise. Complex code developed to deal with inconsistent data

formats can challenge scalability. When organizations enhance and standardize their data, the IT department rewards them with faster IT development and implementation.
Corporate decision-making improves when executives can rely on information they access in their research. Clean data can affect decisions about staffing, inventory management, marketing, mergers and acquisitions, and more.

Why Not Write My Own Code?

Companies benefit from creating many of their systems and processes in-house. Data cleansing isn't one of them. The challenge is enormous from the beginning and the efforts never stop.
Maintenance of in-house data manipulation routines will be ongoing as new data comes into the organization or companies discover new ways to use the data. Better to allocate precious IT resources to projects centered on how to use that data than on ways to make the data usable.

Starting with Excel

Many organizations first notice data anomalies when reviewing data visually with tools like Microsoft Excel. Their eyes tell them when data is in the wrong column or information is formatted incorrectly. The initial reaction is to fix individual data records manually or use Excel functions to do so.
While this approach might be useful for small single-use data files, we can think of several reasons it's not a long-term solution for enterprise data.

First, changing live production data by hand is risky. Excel functions are powerful but applying them can cause unintentional damage to important information. Clicking the Save button is all it takes to render a data file unusable if Excel has corrupted the data unbeknownst to the user.
Second, changing files with Excel is only a temporary repair. New data coming in wouldn't be affected. Finally, manual data quality methods aren't consistent.

Employees may differ in how they apply data standardization operations. Over time, data becomes unreliable again.
In contrast, data cleansing tools like those from Firstlogic allow companies to define rules which the software applies to incoming data. The tools will consistently manipulate and compare data to achieve a reliable standard.

Parsing Engines

One of the hardest programming tasks is parsing data – especially data that departments or outside sources have collected or stored with few controls or conflicting rules. The number of exceptions analysts must handle can be overwhelming. As organizations push data submission out to their customers through online order forms and self-service portals, the problems get even worse. Separating data into distinct sections so that standardization or augmentation can take place can be a long evolutionary process. Programmers may account for data conditions or formats known at the beginning of the project, but must modify routines continuously as new circumstances arise. In the long run, writing (and rewriting) parsing code will be more expensive than using a well-established parsing engine.

Custom Dictionaries

Acronyms, naming conventions, and data-use rules are company or industry specific. Software companies write data cleansing software for general business use, so adapting the software for unique cases can be challenging. The best software addresses these hurdles by allowing users to create their own data tables and dictionaries. The software then accesses these usergenerated repositories to interpret and convert data according to the needs of individual companies or industries.
Custom data dictionaries allow you to handle specific situations in a way that works for your organization, not the way the software company decided.

Data is Always Changing

Perhaps the biggest deterrent to writing your own data cleansing code is keeping up with changing data. New data sources prompt programmers to add conversion or data standardization routines to a growing list of functions. Mergers and acquisitions, changing regulations, and emerging communication channels are creating new data at a rapid pace. Building data cleansing or data standardization programs in-house becomes resource-intensive and a long-term maintenance challenge.

Final Thoughts

The first time you communicate inappropriately with a customer or prospect, you risk losing them for good. That's why data quality and data cleansing are so important. Make a simple error like confusing someone's gender or displaying their name in ALL CAPS and recipients of your messages immediately conclude they aren't important. Even if mistakes aren't glaringly obvious, communications meant for others do not generate desired responses.
If you've failed to recognize a patron is a long-time benefactor because the old database listed her as Mrs. James Miller and the data from last year's campaign lists her contribution as James and Mildred Miller, the messaging will be all wrong. Your pitch won't acknowledge her years of support and your donation ask amounts will be too low.

Sometimes, simple name format differences can cause major headaches. We know of one example in the travel industry where operators couldn't match legal names on passports with passenger names from the reservation list. Delays and confusion while attempting to board not only affected the passengers with unmatched names, but everyone in line behind them. Public displays of poor customer experience are bad for any business!

Companies have a tough job. They want to improve customer experience and reap the benefits of customer retention, referrals, and upsells. They've got the technology and data to forge deeper and more personal customer relationships. Unfortunately, departmental databases, outside data sources, and self-serve web portals make it easy to mistake data's meaning or to use information in a way that erodes customer confidence rather than enforcing it. Data drives every business, but it's up to the organizations using that data to make sure their data is accurate, complete, and appropriate.

No-Fee Assessment

Firstlogic Solutions specializes in delivering data services solutions to data-driven companies.

Firstlogic's products set the standard for address and data quality software when first introduced in 1984. Many users of these products have been customers for more than 30 years, with good reason. Firstlogic's development and support professionals are highly acclaimed and are continuously innovating enhancements to the products, building on their stellar data parsing engine. This engine is acknowledged by many as the best in the business. To find out how software from Firstlogic Solutions can help your organization be more productive, accurate, and competitive, schedule a discovery call.

Gain insight into the health of your organization's data quality! We will process a sample set of your data using the latest data quality tools. Our system will find anomalies in your data and Identify opportunities for improvements.

When you order your no-fee data quality assessment you'll get a data profiling report showing data strengths and weaknesses, along with a custom assessment prepared by Firstlogic Solutions data experts.

Data is driving your business. Make sure you're getting the maximum benefit from the customer information you've collected and stored. Contact us today.