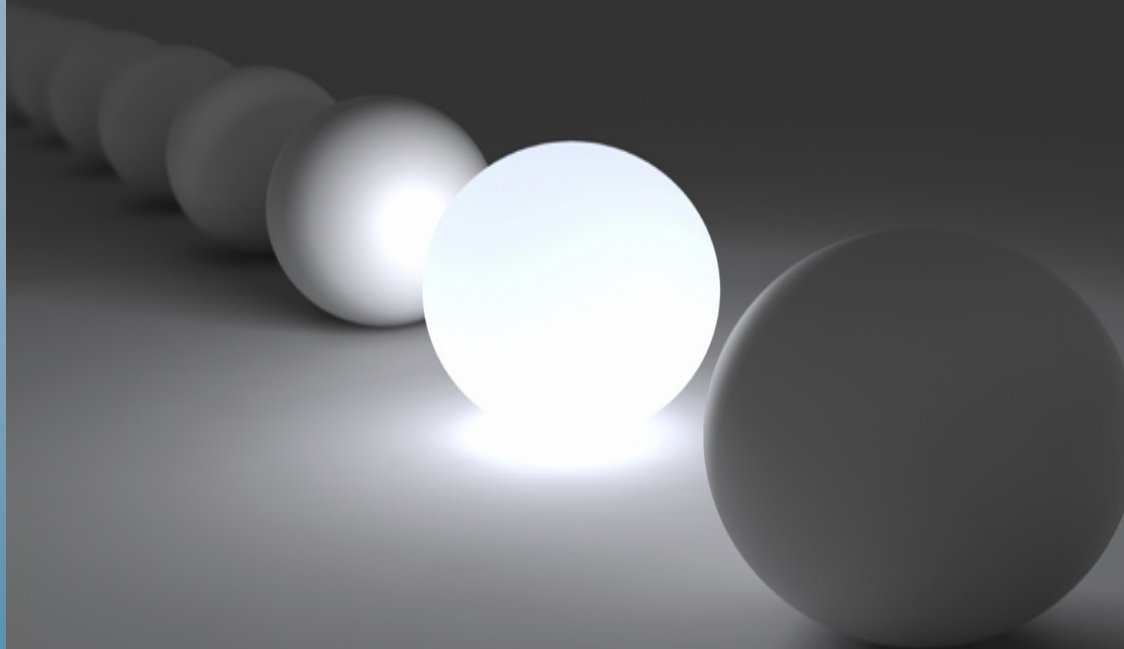


Scalable Data Quality:

A Seven Step Plan For Any Size Organization



Data Quality

Scalable Data Quality: A Seven Step Plan for Any Size Organization

The term Data Quality can mean different things depending upon the nature of one's organization. When applied to customer address records, data quality can be summed up by the following requirements:

- **The data is accurate.** The address actually exists within the city, state and ZIP Code™ given. In addition, if a person or business is associated with the address in the record, that person or business listed is actually located at that address.
- **The data is up to date.** The name and address in any given record reflect the most current information on that person and business.
- **The data is complete.** Each address contains all of the necessary information for mailing, including apartment or suite number, ZIP Code™ and, if needed, carrier route and walk sequence.
- **The data is not redundant.** There is only one record per contact for every address in a mailing list.
- **The data is standardized.** Each record follows a recognized standard for names, punctuation and abbreviations.

Every record that fails to meet the above standards of quality can lead to either lost revenue or unnecessary costs. This is true regardless of the size of the enterprise; from a local florist to and multi-national conglomerate. In fact, data quality is probably even more crucial for the small to medium-sized business or organization than it is to the large corporation.

Not only does each customer potentially represent a much larger percentage of a small business's sales volume but smaller businesses are generally expected to deliver a higher degree of personal service. Therefore, every misdirected or undelivered piece of mail has a greater impact on that business's bottom line than it would for a larger enterprise.

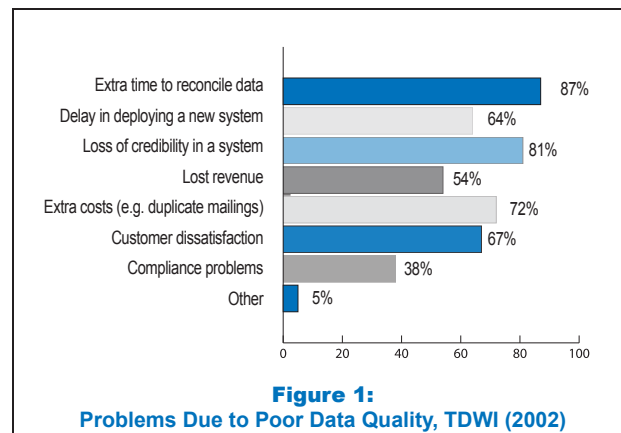
Given these factors, organizations of virtually any size can benefit from a strong commitment to a data quality initiative, one that addresses immediate needs and provides flexibility to meet changing business requirements.

The Scope of the Problem

Undeliverable as Addressed Mail

According to a recent study undertaken by PricewaterhouseCoopers and the United States Postal Service® (USPS), on average, approximately 23.6% of all mail is incorrectly addressed and requires correction of some kind. An additional 2.7% is completely undeliverable.

The USPS currently charges a minimum of \$0.21 per mail piece for its address correction service. For a



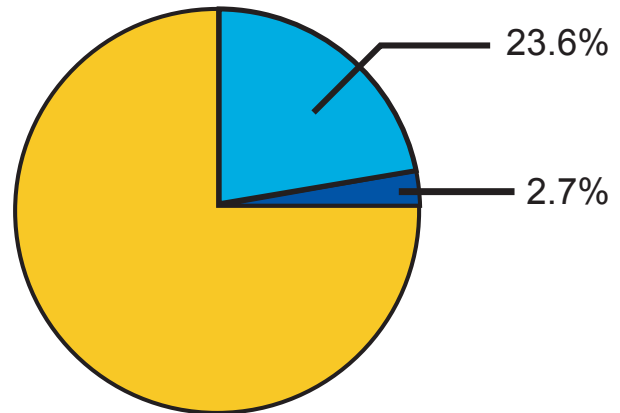
one-time mailing of 10,000 pieces, this could potentially add another \$500 to the cost of the mailing. Manual correction more than triples this cost.

Bulk parcels returned as undeliverable cost nearly \$2.00 per item. For items sent via delivery services like UPS and FedEx, the cost is \$5.00 per item. It isn't difficult to see how these costs can add up in a very short time, diluting profit margins and potentially damaging customer relations.

If the address data is used for billing, incorrect addresses cannot only lead to unnecessary expenses, but also delay collections.

Figure 2:
23.6% incorrectly addressed,
2.7% undeliverable

USPS and PricewaterhouseCoopers Report (2002)



Points of Entry

Bad address data has multiple points of entry in any organization. If an organization collects sales leads over the web, the customer can either mistype their address or deliberately provide false information. Even if employees of the organization collect the addresses, the possibility for errors still exists.

If an organization buys address lists from a vendor or other third-party, there are also multiple entry points for error. The list may contain errors due to poor quality control by the vendor. One simple component of a data quality initiative is to only patronize list vendors that consistently deliver a quality product. But unlike wine, address data does not improve with age and even an error-free list will not stay that way forever. Customers move, companies merge and go out of business.

Whatever the source of the data, these risk factors create the need for a data quality firewall - a set of tools and regular procedures that protect your data from errors at the point of entry to ensure only valid, usable addresses enter the database. This data quality firewall should also scan, correct and update your previously acquired data in batch, preventing any existing errors from "escaping" into the real world and affecting the organization.

Unnecessary Duplication

A database that is relatively error-free and up-to-date can still overlap with existing data. This creates duplicate records, leading to unnecessary expense when a potential customer is contacted more than once.

Selecting the Right Approach to Data Quality

Selecting the right method for ensuring data quality depends on many factors: how the data is acquired, how it is to be used and the amount of data that needs to be processed at one time.

No matter what data quality solution is ultimately chosen, at the very minimum it should accomplish the following:

- Trap inaccurate addresses before they enter your database.
- Process existing records to flag undeliverable mailing addresses for correction or deletion.
- Update current records when information changes (address change, etc).
- Enhance records by appending related mailing information to the record, (ie. ZIP + 4[®] codes and Carrier Route) for faster processing and discounts.
- Standardize addresses using preferred USPS[®] spelling, abbreviation and punctuation.

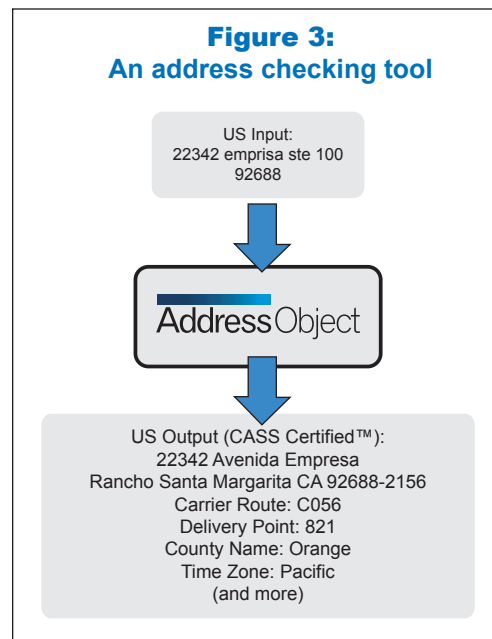
Using an API

If your organization acquires individual sales leads or accepts orders via a web site, you would probably be best served by an address checking component that can be incorporated into a web application. One possible approach is to buy programming tools to be integrated into the application. The advantages of this approach include speed of execution as well as data security (the customer's information never leaves your site after it has been entered). To use the address data for mailings, the address checking logic should also be CASS Certified™ to qualify for postal discounts.

Using a Web Service

Another option is to use a web service that offers the same functionality as an API. This option has two advantages: it may be a less expensive option if you process a relatively low number of addresses and the web service vendor maintains and updates the postal databases for you.

Another point of entry your own data entry personnel can also be a source of inadvertent errors. The ability to check and standardize an address “on the fly,” before it is even stored, will serve to minimize bad addresses. To build an address checking logic into your data entry or CRM systems, consult the address verification tools mentioned above. If you don't have the resources for that sort of development or your systems do not allow that level of customization, you can also purchase a standalone application that accomplishes a similar function.



Using Standalone Software

If you work with lists acquired from an outside source, such as a vendor, or if you choose to process your address data in bulk, there is still more than one option available to you.

The address checking components mentioned earlier can also be used to build in-house applications. This option requires that you have the necessary resources on hand to create, develop and maintain this application. However, if you are already using the same object to support a web site, your developers' familiarity with the component's interface can be extended to developing other applications as well.

If your organization is not large enough to support this kind of effort, you will have the option of purchasing a standalone application that handles the same functions. While this is something of a "one-size-fits-all" solution, it will fill many of the same needs and be ready to run almost from the moment you receive the software.

Another important criterion is that the solution scales well with the growth of your organization. If your approach to data quality works the same with a hundred thousand or a million records as it does with ten thousand, then your operations can grow without worrying about hitting an artificial "ceiling" imposed by a data quality system that you have outgrown.

SEVEN STEPS TO SCALABLE DATA QUALITY

Data quality does not happen by accident. It requires a plan that addresses any current weak points in your data, delivers the desired results and can be accomplished with the available resources. It should also attempt to anticipate future needs when possible, and allow the data quality solution to grow with the organization.

The following section outlines a good framework for creating a plan for your own data quality initiative. It is not the only approach, of course, but it represents a good starting point.

Step 1: Acknowledge that there is a problem

Before anything can happen, your company must be aware that it has a problem with poor data quality and needs to find a solution that addresses this problem. This requires a reasonably rigorous audit of your mailing operations and the costs incurred by inaccurate data.

It should include expenses directly related to bad address data (such as USPS or parcel carrier fees), lost revenue and employee hours spent correcting problems that could be prevented by a data quality initiative.

Step 2: Assign a "point person"

Assign a person or a team that will be responsible for executing your data quality initiative. This person will be in charge of identifying the problem, the possible solutions and bringing these recommendations back to management. After an approach is decided upon, the "point person" will also be in charge of putting the initiative into effect.

Step 3: Identify the problem

The first responsibility of the point person will be to identify the root cause of any data quality problems in your organization. This will require a detailed examination of how data quality problems enter the “pipeline” and an evaluation of what would be the most effective point to enforce data quality.

Step 4: Technology assessment

Deciding upon an approach to data quality requires a thorough and realistic appraisal of the technology and resources available within the organization. Ask yourself the following questions:

1. If data quality is to be enforced via a web site, is it feasible to integrate the technology into the web site or will the site have to be retooled to make it possible?
2. Is the volume of addresses to be verified small enough to make a web service a more economical solution?
3. Does the organization have the programming resources to integrate a data quality tool into an existing application?
4. Does the software used by the organization allow customization or is it a closed system? If it's a closed system, will a standalone address verification program fill the needs of the organization?
5. What would be the best way to ensure that all address data passes through a data quality process before it is used in any way? What changes to the network might be necessary to make this happen?

Step 5: Evaluate vendors

Once the extent of the problem and the available resources has been identified, the point person then identifies possible data quality solutions vendors and evaluates their products. This evaluation should include downloading and using trial versions of their software (if these are available), as well as possibly contacting other users of the same software or organizations similar to your own.

After evaluating as many possible solutions as is feasible, the point person then makes his or her recommendations to management regarding the best solution for the organization.

CASE STUDY - A

Shari's Berries

Sacramento-based Shari's Berries ships chocolate-covered strawberries nationwide via FedEx, guaranteeing on-time delivery of its perishable product line. They accept as many as 200 orders per hour during peak holiday times.

Prior to deploying a data quality solution, the company relied upon their FedEx package scanner to detect bad addresses. A misaddressed package would prompt a call to the customer to verify the address before shipping. During peak times, this process could easily place a strain on their customer service department.

After implementing their data quality initiative, the rate of address errors was less than one percent, reducing the number of customer service verification calls by more than ninety percent.

Step 6: Implementation

At this stage, the recommended data quality solution is put into practice, possibly on a limited basis, for a trial period. The point person tracks costs and manpower usage to compare with those of a similar period before the data quality initiative.

Step 7: Validation

After a trial period, the results are collected and reported back to management. At this point, the effectiveness of the data quality initiative can be evaluated and any adjustments, if necessary, can be made before putting the solution into full-scale use throughout the organization.

Where to Go For More Information

A well-executed data quality initiative is not difficult to accomplish, but it is crucial to getting the maximum value out of your data. While a larger organization could probably absorb the costs of bad data quality, there is no reason why it should when, with the proper tools and a well-planned initiative, a solution is within easy reach.

In a smaller organization, for which every sales lead, order or potential customer is proportionately more valuable, poor data quality is really not an option at all. If your organization survives on its ability to reach its customers, implementing a thorough data quality solution is crucial to your success.

To learn more about practical and affordable data quality solutions that can scale to your growing business needs, visit <http://www.MelissaData.com> or call 1-800-MELISSA.

CASE STUDY - B

Simon & Schuster

Each week, the corporate communications department of Simon & Schuster mails out thousands of books to reviewers. Book reviewers are housed in a database of 100,000 records of journalists, editors, educators and others. Review books are sent out daily to individuals or to groups of people. All contacts are entered into the system by the publicity department.

Prior to implementing a data quality solution to clean, verify and standardize addresses, Simon & Schuster was relying on UPS to make the correction. ZIP Code™ errors were the most common address problem. But UPS charged \$5 per package to correct in the field, plus returns as undeliverable also cost \$5, not to mention labor time at S&S to process the return.

All in all, bad data was costing S&S about \$250,000 per year. Now Simon & Schuster corrects addresses before the packages are handed over to UPS for delivery, preventing any charges for in-the-field correction and re-routing and eliminating approximately ninety-percent of the problems.