

Evaluating Merge/Purge Systems: Part Five

*By Jim Wheaton and Cynthia Baughan Wheaton
Principals, Wheaton Group*

Original version of an article appeared in the November 1987 issue of "Direct Magazine"

[Note: Despite dramatic increases in raw computing power and a proliferation of end-user software tools since the publication of this series of six articles, virtually all of the content remains highly relevant. The occasional obsolete point is highlighted.]

Statement of Purpose

In a series of six articles, we explain a number of the key concepts that mailers should understand about merge/purge, as well as reviewing (in the first article) a methodology that could be helpful in evaluating the effectiveness of either present or prospective merge/purge systems. While our comments are primarily addressed to mailers, merge/purge vendors can benefit by measuring themselves against the criteria that we have identified as important.

Our objective is to describe new and specific tools that can be used to evaluate and improve the performance of the merge/purge process. Through commentary and examples, we will attempt to translate into layman's terms the technical jargon that baffles many mailers. In the process, practical applications should become apparent.

This month's article, Part Five, focuses on run-time, ZIP correction, output reports, vendor service, and client responsibilities.

Run-Time: A Hidden Cost

An important issue for large volume mailers is merge/purge run-time, which can impact turnaround on major jobs. The best software in the world is of limited value if deadlines cannot be met. It is important to recognize, however, that there is often a trade-off between processing speed and quality.

- Some of the least sophisticated systems on the market, those that rely on match-codes, are also some of the swiftest. This is because they use simple matching procedures as discussed in the second part of this series (*Direct Marketing*, July 1987). If all other factors are equal, it simply takes less time to analyze just part of a record.

- Some leading vendors have developed single-pass software for jobs that have previously required two passes, which is often the case with business-to-business unduplication. This obviously cuts run-time significantly, and counterbalances the additional time required by sophisticated systems to process the entire record.
- Also, as you will see, the ZIP correction method that is chosen can reduce run-time.

ZIP Correction

ZIP correction is another important element of the merge/purge process. Software has been developed by a number of vendors to ensure that the ZIP code is consistent with the remaining elements of the address. When an inconsistency is perceived to exist, corrective action is taken.

A problem with some ZIP correction software is a tendency to render undeliverable a technically incorrect but still-deliverable address. This happens when only the street, city and state elements of an address are evaluated. For single-ZIP cities, however, only the city and state, and not the street, are evaluated.

The following example shows how, in a single-ZIP situation, focusing only on the city and state can lead to undeliverability:

Record #1	Record #2
Brad Kraft 14 Main Street Sandy Hill, PA 19401	Brad Kraft 14 Main Street Sand Hill, PA 17042

- In record number one, Brad Kraft really lives in Sandy Hill, Pennsylvania, not Sand Hill, Pennsylvania.
- Due to an input error, the “y” was left off the city name.
- Even though these two cities have similar names, the ZIP Codes are quite different:
 - Sandy Hill is 19401.
 - Sand Hill is 17042.
- Record #1 is technically an incorrect address because it contains a misspelled city name. It is, however, deliverable. This is because the postal carrier within ZIP 19401 is familiar with the Brad Kraft household. He may even know him personally, and have been delivering his mail for years. The postal carrier will get that package to Brad Kraft!

- ZIP correction software, however, does not know Brad Kraft. All it knows is to compare the city name in the record to a look-up table of valid cities within Pennsylvania. When a match is found, the software will ensure that the record contains the corresponding ZIP Code. The implicit assumption is that the city in the record, not the ZIP, is correct.
- In our example, the software will change the ZIP to 17042, the one that corresponds to Sand Hill, as we can see in record number two. Because no one in Sand Hill has ever heard of Brad Kraft, he will never receive his package. The record is now undeliverable.

This is how some ZIP correction software can take a technically incorrect, but deliverable, record and render it undeliverable.

Just to give you some perspective, the following are other examples of similar-sounding single ZIP cities with very different ZIP Codes. All are within Pennsylvania and begin with the letter “S”:

Town	ZIP	Town	ZIP
Smithton	15479	Stoufferstown	17201
Smithtown	15840	Stoverstown	17362
Spring Hill	15946	Stringtown	15320
Spring Mill	19428	Strongtown	15957
Spring Mills	16875		
Springmont	19609	Summerhill	15921
Spring Mount	19478	Summer Hill	18603
Stateline	14775		
State Line	17263		

We are not suggesting that slight input errors would cause all of these to be incorrectly handled by some of the ZIP correction software that is currently available. Instead, they are meant to give you some perspective of the potential magnitude of the problem.

ZIP Correction Methodologies

[Subsequent, August 5, 2003 comment: Due to changes in U.S.P.S regulations and corresponding standardization of industry practices, this section is obsolete.]

There are two basic ZIP correction methods, both of which are performed in the edit process. The primary difference between the two is the percentage of total names that are ZIP corrected.

The first method is to pass all incoming records through ZIP correction.

- Proponents contend that this increases the number of matches in the unduplication phase.
- However, care must be taken to correctly handle a deliverable address that contains an incorrect city, as in the Brad Kraft example above.

The second method is to ZIP correct only the input records that cannot be Carrier Route coded, since Carrier Coding virtually ensures that an address will be deliverable. With this method, approximately 10 percent of total names will undergo ZIP correction.

- Carrier Route coding examines the street and ZIP, but does not force changes that can result in undeliverability.
- ZIP correction run-time can be reduced by as much as 90 percent, and costs are lowered proportionately.
- However, match rates may be adversely affected by the retention of incorrect city names.

Because our study did not involve a quantitative analysis of these two ZIP correction methods, we are unable to recommend one more than the other. You should be familiar with the two, however, and discuss them with your vendor.

Reporting

Valuable, but frequently overlooked, byproducts of the merge/purge process are the summary and control reports. Analysis of these reports can result in valuable marketing insights.

Reporting quality varies significantly among vendors, ranging from handwritten notes on computer paper (which we actually received), to concise and understandable representations of all the key processes that have transpired (which we also received).

- The best reports are those that are written with the user in mind and do not require a great deal of “translation” by the vendor.
- Always review the standard reports of any vendor under evaluation or currently in use. It is important to be able to understand how each field is defined, as well as how different fields relate to each other. A vendor should not inundate you with partially organized data but rather should put that data together in a meaningful fashion. You should be able to clearly account for each name eliminated in both the edit and unduplication phases of the merge/purge.
- It might be helpful to design your own “ideal” documents, at least in very rough form, in order to evaluate how well a vendor’s standard report meet your needs.

- Be sure to include a review of the format of output reports, such as the duplicate and final output listings. These are the documents that need to be periodically reviewed on a “micro” level in order to assist in monitoring the appropriateness of the parameters applied to the mailing. Some vendors’ reports make it easier to spot overkill and underkill than others.

Information requirements can vary significantly for different mailers.

- They may be tied to senior management reporting needs, or to the level of maturity of the direct marketing venture.
- Special reports are sometimes necessary. Some vendors offer flexible reporting software at a reasonable cost, while others can produce special reports only at a significant expense to the client. You have to decide how much you are willing to spend for the flexibility you require.

Vendor Service

Merge/purge is a complex process. Vendor service can be a critical factor in maximizing the quality of names available for mailing. The best vendors attempt to build a close, long-term relationship with the client.

- Time should be spent by the vendor in translating technical jargon into English and defining the marketing needs of the client in systems terms.
- Parameters must be fine-tuned by the vendor over time, with approval from the mailer, as new needs evolve.
- Jobs need to be turned around on a timely and predictable basis.

These are all criteria you should consider in evaluating a vendor.

Client Responsibilities

To be fair, the vendor is not the only party with responsibility in this process. To maximize the effectiveness of the merge/purge, each company should designate an employee to be responsible for the details of execution:

- The employee would need to understand the total range of vendor capabilities and act as the primary vendor contact. This will minimize contradictory and poorly directed communications.

- This person would also provide comprehensive, timely, written instructions (we cannot emphasize that enough!) for each job, as well as analyze merge/purge output for the four types of errors reviewed earlier.

With the client and vendor working together, the merge/purge process can become a very powerful business tool.

Jim Wheaton and Cynthia Baughan Wheaton are Principals at Wheaton Group, and can be reached at 919-969-8859 or jim.wheaton@wheatongroup.com. The firm specializes in direct marketing consulting and data mining, data quality assessment and assurance, and the delivery of cost-effective data warehouses and marts. Jim is also a Co-Founder of Data University www.datauniversity.org.