

Ensuring Data-Quality When Managing Large Address Databases



Why postal address quality has such a great impact.
Use scenarios of address verification and correction software systems.

All company names, product names and services mentioned in this publication are trade names and/or registered trademarks of the respective companies.



ADDRESS DATA QUALITY

One of the major reasons why CRM projects or dialog concepts fail or do not yield the expected benefits is poor data quality.

Inaccurate, missing or outdated address information causes a high percentage of defective data, especially in large customer and prospect databases. Directly or indirectly, these defects are responsible for most of the problems, a fact that has momentous consequences. Address inaccuracies have tremendous negative potential with immediate impact on all communication processes and enterprise-wide activities for which address data is drawn on or used.

The following factors can cause the quality of postal information to suffer if no systems or inadequate systems are used for postal validation and correction of address data:

- Hearing, reading, spelling or transmission errors and different abbreviations when recording or updating addresses
- Changes in city and street names and in house numbers when municipal authorities rename or renumber streets or when municipalities are annexed
- Changes in the postcode by the postal authorities for organizational reasons

Establishing the basis

To ensure the accuracy of address data, a solution for integrated postal validation and correction is essential as basis. Special software tools run an error-tolerant comparison of the address data against the directories of the postal authorities (reference files), and whenever necessary and possible, correct the addresses. It is important in this context not only to check the individual fields but to validate the integrity of the entire address (postcode, city, street, and house or PO Box number). Another important requirement for comprehensive verification is the availability of complete street directories for all cities and municipalities. Often, inadvertently, only partial directories – similar to online postcode books – are used for the verification check, allowing refined verification only within cities that have more than one postcode. However, with these partial directories, address data of the highest quality cannot be achieved. Even the house number should be included in the address verification check.

To ensure high data quality in large databases, the use of expert systems for postal validation and correction provides the essential basis without which all other measures of quality assurance and enhancement would not succeed.

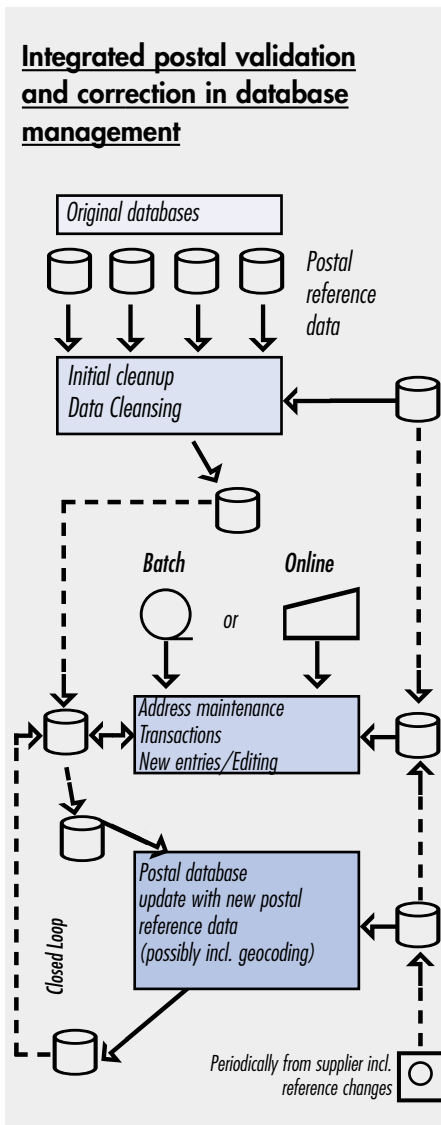
If no software system is used to validate and correct postal addresses, within a very short time the output of the address database will have an error rate of 20 to 35 percent. The consequences are fatal since they affect business communication and customer interaction in their entirety. The direct effects, such as when a mailing is delayed or under certain circumstances when it cannot be delivered at all, are only the tip of the iceberg. Just as detrimental are the indirect problems resulting from addresses of poor quality, such as when performing duplicate detection procedures, when clustering transaction data on the basis of addresses or when enhancing the address database with data for micromarketing.



Typical process flows

An integrated solution runs as a "closed loop" and can typically be divided into the following three areas:

- Initial cleanup
- Maintenance process
- Periodic updating



Initial cleanup

In the initial cleanup, address databases originating from the most diverse sources are uniformly structured, standardized, checked and corrected before they are loaded into the new database. The next step in the data cleansing process is postal verification, which should take place before the database is checked for duplicates and prior to clustering on the transaction level. Doing the steps in this order positively influences the recognition quota in the duplicate check, or rather makes duplicate recognition possible in the framework of incorporations or street renamings.

Twenty to thirty-five percent of the addresses that have not previously undergone postal validation require changes that can be corrected automatically. The percentage of addresses that are recognized by the system as containing errors but which cannot be corrected automatically is usually between 0.5 and 3 percent. One should bear in mind, however, that while a lowest possible error rate is generally desirable, it should not be "bought" at the cost of false or uncertain changes. Data records that are still incorrect are usually transferred to the new database with a distinguishing mark, so that a subsequent clarification, e.g. by the field sales force, can be initiated and monitored.

Maintenance process

All new addresses to be entered or all addresses in which changes must be made should undergo postal validation and correction, independent of whether these change transactions take place via the Internet, online or in a batch process. In Internet or online applications, the validation check should not be optionally activated by the user but rather automatically executed as a matter of principle. In cases of doubt or when the entry is not clear-cut, the system should offer instructions or choice dialogs. A general recommendation cannot be made as to whether an incorrect address should not be accepted into the database as a matter of principle or only with a corresponding error mark. The decision depends among other factors on the kind of business process, the user group and the technical conditions in that context.



Periodic updating

Periodic updating of the address database should always be carried out whenever new postal reference data is issued. In Germany the name coined for that is "quarterly update", since the Deutsche Post AG makes new reference data available every three months. Other countries have different updating cycles. The main objective of a regular postal update is to record the changes in city and street names and house numbers made necessary because of renaming initiatives within the municipality, when there has been an annexation or when the postcode is arbitrarily changed. The user need not activate the changes in the database manually. For that purpose it is absolutely necessary that the reference files not only contain information relating to the reference date, but also that historical information is available and linked to the current valid data. First when periodical updating is an integral component of the system can data of the highest quality be guaranteed on a continuing basis. Experience has shown that clients almost never (and prospects even less so) explicitly send notice of changes in the address because of renaming, municipality annexation or arbitrary changes in the postcode. Periodic updating is also advised when the intention is to update geocoding information in the same process.

Good to know

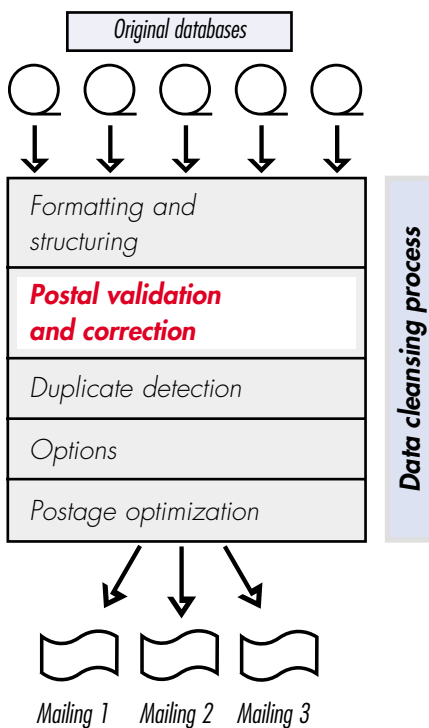
All records that are to be accepted into the database despite postal errors automatically receive a special status, which is readily apparent. Thus, these in correct addresses can be selected by the field sales force for clarification, or they can be omitted in certain mailing campaigns. Generally, the monitoring of these cases has been simplified.



Promotion-specific option for postal verification

As an alternative to an integrated solution for postal cleansing and correction, you can also choose "promotion-specific postal verification". This is often utilized for mailing campaigns in which third-party lists are deployed as well. The validation and correction procedure is exclusively intended for this specific sales promotion, and the changed address data is not transferred back into the databases. The validation procedure is activated solely to improve output quality in order to check for duplicates, to optimize postage rates and to facilitate and ensure delivery as far as possible. Following the postal validation and correction procedure, a decision should be made whether to further process the still remaining inaccurate records or not, depending on the application. If in the later selection process the total amount is limited anyway, the remaining incorrect postal addresses ought not be retained.

Promotion-specific postal verification and correction



Standard solutions made by Uniserv

Uniserv markets its line of postal validation and correction products under the product family name *post*. These systems are expert systems that integrate into the database management, thus ensuring the highest quality of address data. The *post* product line can be utilized in batch processing, in a "classic" online environment as well as in combination with web applications. These standard solutions are available for all generally used system platforms, including special solutions for SAP and Oracle.



Checklist for postal verification and correction systems to safeguard and maintain the quality of address data

- Are all processes required for the closed loop (initial cleansing, ongoing maintenance, quarterly updating) supported by all the necessary functions?
- Are all incorporations, street renamings and postal code changes implemented completely and automatically?
- Can the systems be applied in batch mode, in classic online environments as well as in Web applications?
- Is matching to reference data accomplished using fault tolerant methods, enabling the software to automatically correct typical reading, audio or spelling errors as well as different abbreviations?
- Is the reference data really complete? (Particularly in larger cities, partial street segments are often misleadingly utilized.)
- Are ambiguous cases protooled in the batch log? Are online application users given clear choices and assistance in making decisions?
- Is integration into the applications easy via support of the various interface architectures? Are the systems applicable in all generally used IT platforms?

