



# How Quality Brands Improve Customer Data Quality

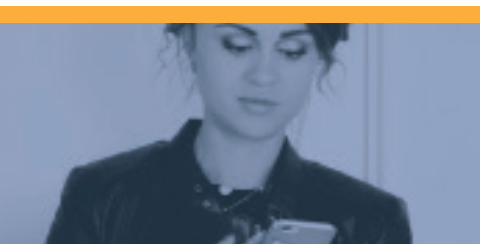
By Roger Matus, Vice President of Global Marketing, and Ed Barbeau, Vice President of Product Management

*It takes a long time for a brand to build a reputation ... losing a reputation can be accomplished in moments*

It takes a long time for a brand to build a reputation for quality through features, performance, service, reliability, and/or durability. Unfortunately, losing a quality brand reputation can be accomplished in moments.

One of the fastest ways to lose a quality reputation is to use poor data when communicating with a customer or prospect. Here are some examples of seemingly simple data-related issues that can cause a relationship to go sour:

1. Confuse one customer with another customer.
2. Use the wrong name for a customer, such as using their family name as their given name.
3. Using the wrong gender for a customer (e.g., he instead of she).
4. Have incomplete or erroneous information about a customer.
5. Send multiple offers to the same person, especially if they are slightly different offers based on past purchases.
6. Send a personal greeting that contains a default value, such as one that begins "Dear (Insert Name Here)."
7. Send documents or packages with an address or phone number that uses the incorrect format for the local market.



*Garbage in / garbage out erodes customer satisfaction.*

Bad customer data can also lead to a bad analysis and forecasting. If you are trying to analyze your company's ideal buyers, you will get the wrong answer if you cannot match a customer to all of his or her transactions.

"Garbage in/garbage out erodes customer satisfaction," stated Forrester Research in a report entitled *Poor Data Quality: An Often Overlooked Cause Of Poor Customer Satisfaction Scores*. "Customer service agents need the right data about their customers, purchases, and prior service history at the right point in the service cycle to deliver the right answers. But when their tool sets pull data from low-quality data sources, agents don't have the right information to answer their customers."

Sixty-nine percent (69%) of organizations believe that inaccurate data will undermine their ability to deliver an excellent customer experience, according to a 2018 study conducted by Insight Avenue for Experian Data Systems.

A poor customer experience has a direct impact on a brand's reputation:

"Overall, 71% of respondents said their typical response to a bad experience is to stop doing business with the company. A slight majority (55%) typically tell friends and family about it in person or by email, while 42% said they complain to the

company and 26% post a comment on social media," according to a study released by the Economist Intelligence Unit (EIU).

The impact is that poor customer data costs organizations 6% of annual revenues, according to a Royal Mail study of businesses in the U.K.

How bad is the problem? A recent study of the more than 10-million customer records by Global-Z showed the following:

- 45% of the records contained invalid information.
- 38% had terms such as "Need New Address" in the address field.
- 12% had duplicate information in multiple fields.
- 5.5% had the identical values in the given name and surname fields.

Many of remaining 55% valid records contained significant problems for creating a quality customer interaction. Different records about the same customer often did not match exactly. For example, a customer may enter his or her name and address differently on different documents. (See Name Matching Example on the next page.)

### **Name Matching Internationally**

A name can be misused simply because of data entry. For example, in some cultures, such as those in

Japan, Korea, and Hungary, it is common to write the family name before the given name. A software program that displays a “first name” as the given name could cause the agent or letter to unintentionally offend a customer. At a minimum, it shows that the brand does not understand the local market.

For global organizations, the challenge of matching a name increases when a name is written in a different character set in one country (e.g., Cyrillic, Greek, Chinese, Japanese, Korean) than in another (e.g., Latin, Roman, Western).

For example, when a Greek resident writes his or her address in Greece, he/she is likely to use the Greek character set, which is different from the way it would be written in the rest of Europe.

Generally, matching systems use transliterations of names for matching when character sets are mixed. However, as transliterations vary, a matching system cannot rely on exact matches across countries.

In addition, a flexible system for one country may be too flexible for another country with resulting false positive errors. For example, according to *The Economist*, one in five South Koreans have the surname Kim. The likelihood of a mismatch

using names in Korea is higher than in many other nations.

Unfortunately, basic data quality tools and fully automated cleansing systems are prone to make matching errors, especially for international customers.

Proper cleansing of customer data incorporates a detailed understanding of the unique properties of the particular customer database. Items to consider when matching include the country, language, quality level, number and types of attributes, as well as the tolerance for false positive vs. false negative matches (usually driven by use case).

### Best Practices for International Data

The best practice followed by those who specialize in international data quality fall into the following categories:

1. Data gathering
2. Data parsing
3. Postal address hygiene
4. Email address, phone number, and other data hygiene (if available)
5. Matching and merging
6. Managing metadata

### Name Matching Example

Would your customer information system consider records containing names below as belonging to the famous footballer (soccer player) Therese Sjögran?

Therese Sjögran  
Thérèse Sjögran  
Teresa Sjogren  
Terry Sjogran  
T. Sjögran  
KIT Sjögran  
Theresa Sjogrann  
Kirstin Sjögran  
Kerstin Ingrid Therese Sjögran

The answer is that any of these could be Ms. Sjögran, whose real name is Kerstin Ingrid Therese Sjögran.



## 1. Data Gathering

The first step in the process of resolving a person's identity is to gather all of the person's data. Taking an inventory to determine where data is stored often involves many parts of an organization to identify potential source files.

Examples of places where data may be found are the following:

- **Marketing:** Website forms, email lists, trade show leads, etc.
- **Accounting:** Billing, receiving, credit reports, etc.
- **Shipping / fulfillment records**
- **Sales:** CRM systems, contact lists, etc.

## 2. Data Parsing

Once the data is collected, the best practice is to parse information into its component attributes. Examples of attributes used for addresses are as follows: Unit/Apartment/Flat, Premises, Number, Street, City, District, Town, Postal Code, and Country.

The selection of attributes ought to be defined with international markets in mind. Conventions for addresses and phone numbers vary considerably by market. Consistent attributes make it possible to compare records and to display them properly.

### Data Parsing Example

#### ORIGINAL:

23 DAVID PLACE  
ST HELIER JE2 4TE

#### PARSED:

Number:	23
Street:	David Place
City:	St Helier
Postal Code:	JE2 4TE

### Data Parsing Example with Non-Roman Character Set

An individual may write an address for billing in Japan as follows:

北海道札幌市東区北二十四条東3-3-1

The source information in the Japanese system would be parsed as follows::

Block Sequence:	3-3-1
Area Name:	北二十四条東
District:	東区
City:	札幌市
State:	北海道
Postal code:	065-0024

The same address would Romanized for a European copy of the database:

Block Sequence:	3-3-1
Area Name:	Kita-24 Johigashi
District:	Higashi-Ku
City:	Sapporo-Shi
State:	Hokkaido
Postal Code:	065-0024

### 3. Address Hygiene

The same address may be entered into a database in a variety of ways, as shown in the example below. Therefore, the next step is to validate and standardize addresses.

Validation involves comparing a parsed address with official records and making corrections. In some countries, a location may have multiple designations or even names. Best practice would be to perform the following operations:

- Determine or validate the country
- Match to Postcode Address File (PAF) reference data for the

specific country, which includes officially licensed sources such as the country's postal authority, other government agencies, and third parties.

- Correct and standardize identified components
- Append/insert missing components

While licensed and up-to-date PAF data may cost more than other sources, they can vastly improve accuracy and quality. It is a good idea to ask vendors whether they use officially licensed sources.

In dense urban areas with multistory buildings, such as Hong Kong, many

*While licensed and up-to-date PAF data may cost more than other sources, they can vastly improve accuracy and quality.*

#### Address Hygiene Example

The original records appear as the user entered them into various on-line forms

FirstName	LastName	Address1	Address2	Address3	City	Postcode
CHRISTIANE	GELLESCH	AVON HOUSE	23 DAVID PLACE, ST. HELIER		JERSEY	JE2 4TE
CHRISTIANE	GELLESCH	AVON HOUSE	23 DAVID PLACE	ST HELIER JERSEY	CHANNEL ISLANDS	JE2 2TE
CHRISTIANE	GELLESCH	AVON HOUSE	23 DAVID PLACE,	ST. HELIER	JERSEY	JE2 4TE
CHRISTIANE	GELLESCH	43 DAVID PLACE	AVON HOUSE		ST HELIER	JE2 4TE

The parsed address is validated and standardized as follows:

Building: Avon House  
 Number: 23  
 Street: David Place  
 District: St. Helier  
 City: Jersey  
 Postcode: JE2 4TE

*Validating using known patterns for a particular country adds confidence that it was correctly entered into the system and parsed.*

### Phone Number Hygiene Example

The same phone number may be entered into a database in the following ways:

+44 7509312345  
07509 312345  
0 7509312345  
75.09.31.23.45

A check of the address record would determine that the phone number is likely to be in the Channel Islands, which is part of the UK. If the incoming number has a +44 at the beginning, it can be matched to the country code for the UK. If it does not begin with a +44, the country can be postulated from the country field of the address. The remaining digits can then be parsed for a known UK pattern. Once properly parsed, the area code can be used to identify the phone type, as follows:

Country code: +44  
NDD Prefix: 0  
Area code: 7509  
Local number: 312345  
Phone type: Mobile

The final result can then be shown in the correct formats. The domestic format would be 07509 312345. The international number format would be +44 7509 312345.

Validating using known patterns for a particular country adds confidence that it was correctly entered into the system and parsed. Here are some ways in which invalid phone numbers may be identified:

Country	Data Found	Problem
Australia	26621764	"26621" numbers must have exactly nine digits
Germany	5188547213	Numbers in Germany may not start with "5188"
Germany	0114989250	Numbers may not start with "498" and "89250" numbers must have seven digits
France	3.33600E+11	Exponential notation caused by corruption

people may have the same name within the same premises. Sub-premises detail, such as floor numbers or block information, adds to the confidence in matching. Other less dense locations may use building names or route numbers without more specific address components.

#### 4. Email Address and Phone Number Hygiene

Email addresses and phone numbers may add valuable information when used with names to resolve an individual's identity. As with postal addresses, a critical step is hygiene.

Phone numbers may be corrected using the following steps:

- Parse to identify components.
- Remove illegal characters.
- Identify or insert country code and local area code.
- Validate against reference database.
- Format phone number.
- The following steps may be used to cleanse a customer's email address:
- Check the email format for compliance with internet standards (RFC 2822).
- Parse address into user and domain.
- Correct common domain errors.

- Search Domain Name System (DNS) to confirm that the domain exists.

With the postal address, email address, and phone number fields cleansed, standardized, and verified, it is now possible to compare records with each other for matching and reconciliation, which is the key to identity resolution and building the Golden Master Record for the Single Customer View.

#### 5. Matching and Merging (Duplicate Reconciliation)

Standardized records rarely match exactly. Personal data is always changing. People move. Individuals may provide a work phone number during one interaction and a mobile number during another interaction. Many people also have multiple addresses (e.g., primary and vacation homes), multiple phone numbers (e.g., fixed landline and mobile), and multiple email addresses (e.g., personal and work).

The objective of matching systems is to build a Golden Master Record that references the data of matched records. Without that capability, a customer could appear multiple times.

An important technique known as "cascading" helps to confidently build a complete Golden Master Record even when some attributes disagree and to fill in incomplete information.

#### Email Hygiene Example

In this example, the customer's email address in the database is as follows:

**Christiane@yahoo.couk**

- Recognize that ".couk" is not a proper top-level domain.
- Parse the top-level domain and identify that "uk" is a proper country code.
- Alter the domain to "@yahoo.co.uk".
- Check that the domain "yahoo.co.uk" exists using the available DNS records. In this case, it does not.
- Recognize common errors, such as typing "yaho" instead of "yahoo".
- Alter the domain to "@yahoo.co.uk"
- Validate the domain "yahoo.co.uk" using the available DNS records.

The standardized address would be as follows:

**christiane@yahoo.co.uk**

*The art in the process of matching is to understand the importance of each attribute in a particular geography.*

For cascading to work, matching tolerances must be established individually for each attribute. The match tolerance for any given attribute may differ based on the geography. As mentioned earlier, in some densely populated locations, such as Hong Kong, sub-premises detail is important. The weight placed upon address components in Hong Kong would be higher than in a smaller city.

The art in the process of matching is to understand the importance of each attribute in a particular geography and to set the match threshold accordingly. Significant local knowledge and experience are required to get this right.

## 6. Metadata

It is important that the Golden Master Record contain additional information (“metadata”) for each attribute or group of attributes in order to know what is the most current information, the quality of the source, and more.

The source databases must be considered when building the metadata for the Golden Master Record as well. A large conglomerate may want to have a Golden Master Record that covers the entire conglomerate. In other cases, subsidiaries may be considered as separate entities.

### Cascading : A Key Technique

Cascading uses information across multiple records to improve the accuracy of a match. Consider the following example:

- **“Record 1”** contains a name, postal address, and phone number.
- **“Record 2”** contains a name, postal address, phone number, and email.
- **“Record 3”** contains a name, email, and a different phone number.

Record 1 and Record 2 match because the name, address and phone number match.

Record 1 and Record 3 considered in isolation would not match because the phone number differs and there is no other information in common for comparison.

However, Record 2 and Record 3 would be a likely match because the name and email match, even though the phone number is different.

Since Record 1 and Record 2 match, and since Record 2 and Record 3 match, cascading allows us to consider Record 1 and Record 3 a match.



## Golden Master Record Using Cascading Example

The following illustration shows a Golden Master Record with the most accurate consolidated information about the customer after successful matching with cascading:

Attribute	Record 1	Record 2	Record 3	Record 4	Golden Master
First Name	Christiane	Christiane	Chris	Christiane	Christiane
Last Name	Gellesch	Gellesch	Gellesch	Gellesch	Gellesch
Building		Avon House		Avon House	Avon House
Number	23	43	23		23
Street	David Place	David Place	David Place		David Place
District	St. Helier	St. Helier	St. Helier	St. Helier	St. Helier
City		Jersey	Jersey	Jersey	Jersey
Postcode	JE2 4TE	JE2 4TE	JE2 4TE	JE2 4TE	JE2 4TE
Country	Channel Islands	Channel Islands	Channel Islands	Channel Islands	Channel Islands
Country Code		+44	+44	+44	+44
Local Area Code		0	0	0	0
Area Code		7509	7509	7509	7509
Local Number		312345	312345	312345	312345
Email	christiane@yahoo.co.uk	christiane@yahoo.co.uk		christiane@yahoo.co.uk	christiane@yahoo.co.uk

*Golden Master Records with links to original sources is critical to accomplishing the goal of brand quality.*

### Conclusion

While creating a quality reputation is a long-term and often expensive process for a brand, losing a reputation can be accomplished in moments.

Bad data quality can offend a customer and lead to lost revenue.

There are steps that can be taken that enable companies to present the right information in the right format. Creating Golden Master Records with links to original sources is critical to accomplishing the goal of

brand quality. Deep international experience with an understanding of data in each country increases the chance of success.

If you would like to know more about improving data quality for building better brands, this process, or how our deep experience would apply in your circumstance, please contact Global-Z at [info@globalz.com](mailto:info@globalz.com).



**Global-Z**

v5/18