



Global-Z

6 STEPS TO PREPARE YOUR CUSTOMER DATA FOR GDPR

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

Golden Master Records help companies to quickly find, edit, or delete all the information they have about any EU individual at any time, as required by GDPR.

If you have a website with a form or an email list that includes European customers, the new European General Data Protection Regulation (GDPR) will require you to be able to access, delete, and make corrections to all of your data about any EU individual, even if you don't have offices in Europe. This paper outlines six steps that will help prepare your customer record management system to handle the demand.

GDPR, which goes into effect on May 25, 2018, applies to anyone in the world who offers goods or services to European Union (EU) individuals or monitors the behavior of EU individuals. (In June 2017, the United Kingdom confirmed that GDPR will form a part of UK law, even after Brexit. Therefore, for this document, the term EU is meant to include the UK.)

Unlike previous privacy directives, GDPR applies to companies even if the company does not have a physical presence in the EU. Many companies, including Internet-only companies, that had never concerned themselves with the handling of EU personal data must do so now.

The penalty for a failure to comply could mean a fine of up to 20-million Euros (about US\$24-million) or four percent of worldwide revenues, whichever is greater.

“Personal data” is defined by the European Commission as “any information relating to an individual, whether it relates to his or her private, professional or public life. It can be anything from a name, a home address, a photo, an email address, bank details, posts on social networking websites, medical information, or a computer’s IP address.” Almost anything collected on a website, including behavioral activity, related to an EU person is regulated.

An “individual” includes both citizens and residents of any EU member state.

GDPR gives EU individuals meaningful rights over the personal data you collect, including the following:

- **The Right to Consent:** Individuals must explicitly consent for you to use their personal data. Opt-outs and pre-checked opt-in boxes are no longer sufficient. In addition, the uses for the data must be disclosed at the time of collection. Consent may be withdrawn or limited to specific purposes at any time. You must be able to prove that they have given consent.
- **The Right to Access:** Individuals may obtain a copy of the information held about them. They also have the right to know where the information is stored, who can access it, how they access it, and the reasons for access.
- **The Right to Be Forgotten:** Individuals may request that any or all information be deleted.

- **The Right to Rectify:** Individuals may require companies to correct inaccurate data and may object to any profiling that may result in discrimination.

This list covers only a small subset of the rights and regulations in the GDPR, which consists of 99 articles and 173 recitals over 261 pages in English.

Finding All The Data

To meet these GDPR requirements, it will be necessary for organizations to be able to quickly find and modify all of the information about an individual on demand.

The most effective way to do this is to create a searchable record management system with a Golden Master Record for each customer containing both the most recent information and a key that lets you find all of the source records.

However, building a Golden Master Record from multiple sources for each customer is not simple because the data in different databases or even different records in the same database rarely match exactly. (See Finding Data Example on the next page.)

For global organizations, the challenge of accurately matching a name increases when a name is written in a different character set in one country (e.g., Cyrillic or Greek) than another (e.g., Latin/Roman/Western). When an EU resident of Greece writes his or her name and address in Greece, he/she may use the

FINDING DATA EXAMPLE

If the famous footballer (soccer player) Therese Sjögran asked for a copy of all of the personal data your company maintained about her, would your system know if the names listed below belong to Ms. 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. The name she might submit on a Right To Access request might not match all of the names used in all of her transactions with your organization.

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, since transliterations vary, a matching system cannot rely on exact matches across countries.

However, a system that is flexible in one country could incorrectly match different people in countries where names are common. For example, according to The Economist, one in five South Koreans have the surname Kim.

Resolving Identity

The solution to the problem of making accurate matches is to use as many customer attributes as possible and to establish a tolerance level for each attribute to accept a match. Local knowledge is critical to do the matching properly because the significance of each attribute may vary by location.

The best practice followed by those who specialize in customer identity resolution is to take the following six steps:

1. Gathering the data
2. Data parsing
3. Postal address hygiene
4. Email address, phone number, and other data hygiene (if available)
5. Matching and merging (reconciliation)
6. Metadata management

These steps are explained below in more detail:

1. Gathering The Data

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 personal data is stored often involves large parts of an organization to identify all of the potential source files. Examples of places where personal data may be found are the following:

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

Proper data governance and data stewardship include collecting and keeping track of this source information so that when requests are made from individuals the proper actions can be taken on the source files.

2. Data Parsing

Simple approaches to matching will not work because the same address may be entered differently in different source systems. For example, “23 David Street, St. Helier” may be the same address as “Avon House, David St., Saint Helier, Jersey.”

The best practice is to parse the information into its component attributes. For example, a street address may be broken down into building number, directional (e.g., north or east), street name, and street type (e.g., street or avenue). When building a system for global use, it is

DATA PARSING EXAMPLES

ORIGINAL:

23 DAVID PLACE
ST HELIER JE2 4TE

PARSED:

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

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

important to create a consistent data structure that can be used across countries. While parsing resolution varies by geography, data usually may be parsed in the following way: Unit/Apartment/Flat, Premises, Number, Street, City, District, Town, Postal Code, and Country.

3. Address Hygiene

The same address may be entered into a database in a variety of ways, as shown in the example on the next page. 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 large urban areas with multistory buildings, such as Hong Kong, there

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

may be apartment buildings in which many people have the same name within the same premises. Sub-premises detail, such as floor numbers or building information, adds significantly to the confidence in matching. Small towns 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.

The following steps may be used to cleanse customer's email address:

- Check the email format for compliance with internet standards (RFC 2822).
- Parse email address into user and domain.
- Correct common domain errors.
- Search Domain Name System (DNS) to confirm that the domain exists.
- When even greater confidence is required, send a test message to see if the message is accepted by the server to validate the specific email address.

Phone numbers may be similarly corrected using the following steps:

- Parse phone number to identify components.

- Remove illegal characters.
- Identify or insert country code and local area code.
- Validate against reference database.
- Format phone number.

With the postal address, email address,

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 top-level domain for the United Kingdom.
- 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.
- For even higher accuracy, send a message to the email to see if the message bounces or use a third-party service to validate the user mailbox.

The standardized address would be as follows:

christiane@yahoo.co.uk

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.

5. Matching and Merging (Duplicate Reconciliation)

Simple attribute matching is not enough to retrieve all customer records because even standardized records rarely match

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

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).

Systems must be able to build a Golden Master Record even when the information is incomplete or disagrees. Without that capability, a customer could appear multiple times. Therefore, a search would not retrieve all of the customer data required during a GDPR request.

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. An example of cascading is described below.

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 those address components in Hong Kong would be higher than in a smaller city.

CASCADING: A KEY MATCHING TECHNIQUE

Cascading uses all of the customer information available across multiple records to increase the probability of a match. Consider the example of three records:

- **“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 be a match because the phone number differs and there is other useful data to match to Record 1 alone.

However, with cascading, Record 3 can be matched. Because Record 1 and Record 2 are known to match, there is knowledge about the customer’s identity gained from the combination of the records: name, address, email address, and phone number. This new combined knowledge can be used for matching. The email address in Record 3 matches the combined knowledge of Record 1 and Record 2.

As more data is compared the knowledge cascades. In addition to the name, address, email address, and phone number attributes, some valuable information used in cascading includes gender, social media handles and identification numbers.

GOLDEN MASTER RECORD AND 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

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

Because GDPR allows individuals to grant and change permissions for each attribute, it is important that the Golden Master Record contain additional information (“metadata”) for each attribute or group of attributes about these permissions. (An example of a group of attributes would be a person’s address. If an

individual grants the use of an address, the permission would apply equally to all of the address attributes: building, number, street, district, and city.)

Best practice is for the metadata for each attribute or group of attributes to include the types of permissions granted and the date of the grant. The date is critical because permissions can change and it is necessary to know the current permissions. For example, it is possible that an individual will grant the use of an address for any purpose in January, limit that purpose for credit card notifications in March, and then allow it for sale notifications in June.

As the Golden Master Record is created, it is also best practice to create and maintain a link back to the source records. This is especially important as it is necessary to change or delete a source record when requested.

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. However, an individual may grant permission to one subsidiary or brand, but not another subsidiary or brand. This may mean that the conglomerate does not have the permission to use the attribute in any place other than where the permission was granted. Master record management requires knowing the source of the permissions and the applicable entity.

Conclusion

To meet European General Data Protection Regulation (GDPR) requirements a customer identity resolution process becomes critical. While there are many other requirements for GDPR compliance, it is necessary to be able to find, edit, and potentially delete any personal data about any EU individual.

Creating Golden Master Records with links to find the original source records are critical to accomplishing this goal. Deep international experience with an understanding of data in each country greatly increases the chance of success.

If you would like to know more about Golden Master Records, this process or how they would apply in your circumstance, please contact Global-Z at gdpr_info@globalz.com.