

Specification: Automatic Customer Import

Version 1.0
2 February 2020
Karsten Silz

This is not a specification for a real system. This specification is a work sample from Karsten Silz. For more details, see: <https://ksilz.com/samples>

Table of Contents

1. Document History	3
2. Attachments	4
3. Introduction	5
4. Business Drivers	6
5. Specifications	7
5.1.Data File Delivery	7
5.2.Data File Format	7
5.3.Data File Content	8
5.4.Data File Processing	9
5.5.Result File Format & Content	10
5.6.Result File Delivery	10

1. Document History

Version	Date	Content	Authors
1.0	2 Feb 2020	<ul style="list-style-type: none">Section 5.4: Added "Postal Code" to customer ID calculation.Section 5.6: Defined mail server details.	<ul style="list-style-type: none">John Doe (client)Karsten Silz (application owner)
0.2	16 Jan 2020	<ul style="list-style-type: none">Section 5.1: Detailed data transfer through SCP.Section 5.2: Defined line ending.Section 5.3: Added field "Address Line 3".	<ul style="list-style-type: none">John Doe (client)Jane Doe (client)Karsten Silz (application owner)
0.1	7 Jan 2020	Initial version.	Karsten Silz (application owner)

2. Attachments

1. Sample data file v4.csv: A sample Data File from the client with a single customer.
2. Automatic Customer Import Results [YYYY-MM-DD HH-mm].xls: A sample result file.

3. Introduction

The application is web-based and links customer information with external events.

As part of the rollout of the application at the client and its ongoing operation, the customer data will be loaded daily into the application. This document specifies the custom solution that the application owner developed for this: the Automatic Customer Import. This specification is precise enough so that the following two tasks can be accomplished:

- The client can deliver the Data File to the Automatic Customer Import with the right format and the right content.
- The application owner can build the Automatic Customer Import to add and update applications correctly in the application with this Data File.

4. Business Drivers

There are two business needs with regards to the application inventory at the client:

- The inventory needs to be uploaded initially.
- The inventory needs to be kept up-to-date.

The application has a standard facility for these two tasks: Manual Data Import. In the Manual Data Import, users fill an Excel spreadsheet with the customer data and upload this file into the application for processing.

Unfortunately, this standard solution will not work for the client for two reasons:

- The Manual Data Import requires all customers to have a unique ID. But in the current client customer database, the customers do not have a unique ID. So the Automatic Customer Import described herein must create such a unique ID.
- The client does not want to manually upload customers every day.

5. Specifications

5.1. Data File Delivery

1. Client will make the Data File available daily before or at 6:00 AM EST.
2. The Data File name will be `automatic-customer-import_[YYYY-MM-DD].csv`, where
 - “YYYY” is the current year, expressed in four digits (e.g., “2020”),
 - “MM” is the current month, expressed in two digits and with a leading zero if necessary (e.g., “02”), and
 - “DD” is the current days, expressed in two digits and with a leading zero if necessary (e.g., “26”).
3. Two examples for file names are `automatic-customer-import_2020-02-26.csv` and `automatic-customer-import_2020-10-03.csv`.
4. Client will upload the Data File through the Secure Copy (SCP) protocol to the The application server.
 - Client will use the user called `upload` to connect to the The application server.
 - For security reasons, only remote connections for SSH version 2 through the standard port 22 will be accepted.
 - Client will provide the application owner with an SSH public key that will be used during the file upload.
 - The application owner will store that public key as an SSH authorized key for the `upload` user on the the application server
 - Client will upload the data into the `/opt/upload` directory.
5. Client may delete Data Files that were created yesterday or before yesterday.

5.2. Data File Format

1. The Data File is a plain text file in the UTF8 encoding.
2. The end of each line in the Data File is denoted by a single `LF` character (the “Unix line ending”; ASCII code 10 decimal / 0x0A hexadecimal)
3. The Data File contain no heading line.
4. All data for each application is on a single line.
5. In all lines, the fields are separated by the pipe symbol (“|”; ASCII code 124 decimal / 0x7C hexadecimal).
6. No data field is enclosed by single or double quotes.
7. Numbers use the American notation of a dot for the decimal delimiter (“.”; ASCII code 46 decimal / 0x2E hexadecimal), but do not use a thousand separator.
8. The Data File must not have an empty line as the last line.

5.3.Data File Content

1. The Data File contains a list of all customers.
2. The maximum length of text field is described in bytes. In UTF8 encoding, only the first 128 ASCII characters are encoded with a single byte; special characters (like French accent characters) use two or more bytes per character.
3. The following table describes the fields in the Data File.

Position	Name	Type	Required?	Description
1	Salutation	Text: 2-10 bytes	no	This is the of the customers, such as "Mrs." or "Mr".
2	First name	Text: 2-30 bytes	yes	This is the first name of the customer.
3	Middle name	Text: 2-30 bytes	no	This is the middle name of the customer.
4	Family name	Text: 2-30 bytes	yes	This is the family name (or last name) of the customer.
5	Birthday	Text: 10 bytes	yes	The birthday has the format "YYYY-MM-DD". There are leading zeros for both the day and the month. An example: 1989-03-04.
6	Nationality	Text: 2 bytes	yes	This is the ISO 3166-1 alpha-2 code which is a two-letter country code. Examples are US for the "United States of America" or DE for "Germany".
7	Address Line 1	Text: 2-255 bytes	yes	This is the first address line. It typically contains the street and street number.
8	Address Line 2	Text: 2-255 bytes	no	This is the second address line. It typically contains additional information, such as floor or apartment number.
9	Address Line 3	Text: 2-255 bytes	no	This is the third address line. It typically contains additional information, such as floor or apartment number.
10	City	Text: 2-50 bytes	yes	This is the name of the city.
11	Region	Text: 2-50 bytes	no	This is the name of the geographical area in which the country is divided. Examples are the state in the United States (such as California) or the county in the United Kingdom (such as Buckinghamshire).

Specification: Automatic Customer Import

Position	Name	Type	Required?	Description
12	Postal Code	Text: 4-50 bytes	yes	This is the postcode or ZIP code of the address. It can consist of letters, numbers, spaces and hyphens. Examples are 70173 or MK4 3BX.
13	Country	Text: 2 bytes	yes	This is another ISO 3166-1 alpha-2 code . Please see field #6, "Nationality", for details
14	Latitude	Number:-90 <= x <= 90	no	This is the latitude of the application, expressed in decimal notation. An example: 34.0543
15	Longitude	Number:-180 <= x <= 180	no	This is the longitude of the application, expressed in decimal notation. An example: -84.2732

5.4.Data File Processing

1. The application will read in the Data File.
2. The application will ignore the header line and process each line with application data, one at a time.
3. When matching customers between the Data File and the customers in the application, the combination of "First Name" (position 2 in the Data File), "Family Name" (position 4), "Birthday" (position 5) and "Postal Code" (position 12) will be used as the key.
4. When processing each line, the Automatic Customer Import goes through the following checks in the exact order as listed here:
 - A. **Customer Added:** The customer only exists in the Data File, not in the application. It will be added to the application.
 - B. **Customer Deleted:** The customer only exists in the application, not in the Data File. The customer is marked as deleted in the application.
 - C. **No change:** The customer exists both in the Data File and in the application, and its data is identical. No action is taken.
 - D. **Customer Updated:** The customer exists both in the Data File and in the application, but at least one field is different in the Data File and the application. The values from the Data File overwrite all differences in the application.
5. When the Automatic Customer Import attempts to add or update a customer in the application, the result will be one of the following:
 - **OK:** The operation was successful.
 - **Error:** The operation failed. This usually happens when values in the Data File do not adhere to their constraints defined in the previous section, "Data File Contents". Examples

Specification: Automatic Customer Import

are missing required fields (such as "First Name" or "Last Name") or fields with invalid data (such as malformed "Birthday" or "Nationality" fields).

6. When the Automatic Customer Import successfully adds or updates a customer in the application, then those changes are documented in a Result File (see next section).

5.5.Result File Format & Content

1. The Automatic Customer Import will create a Result File, once the Data File has been processed. That result files is available in the application in the “Data Import Results” report.
2. The Result File will be an Excel file in the Excel 97-2003 format with the following sheets in the order specified here:
 - A. **Summary:** This lists how many customers fell into each of the following categories:
 - No change
 - Customer Added
 - Customer Updated
 - Customer Deleted
 - B. **Customers Added:** The customers only existed in the Data File, not in the application.- They were added to the application.
 - C. **Customers Deleted:** The customers only existed in the application, not in the Data File. The customers were deleted in the application.
 - D. **No change:** The customers existed in both the Data File and in the application, and their data is identical. No change was made.
 - E. **Customers Updated:** The application existed both in the Data File and in the application, but at least one field differed in the Data File. The customers were updated in the application.
3. In each sheet, with the exception of the “Summary” sheet, the Result File will list the following data:
 - All fields from the Data File, as specified in section “Data File Content”.
 - **Status:** This is one of the following values:
 - **OK:** The operation was successful.
 - **Error:** The operation failed.When the “Status” is“Error”, then this field will also contain the details of the error.

5.6.Result File Delivery

1. The Result File will be name “Automatic Customer Import Results [YYYY-MM-DD HH-mm].xls”, where
 - “YYYY” is the current year, expressed in four digits (e.g., “2014”),
 - “MM” is the current month, expressed in two digits and with a leading zero if necessary (e.g., “02”),
 - “DD” is the current days, expressed in two digits and with a leading zero if necessary (e.g., “26”),

Specification: Automatic Customer Import

- “H” is the hour, based on a 24h clock and expressed in up to two digits (e.g., “5” or “15”), and
 - “mm” is the minutes, expressed in two digits and with a leading zero if necessary (e.g., “42”).
2. The Result File will be sent as a ZIPped attachment to an email, called “Automatic Customer Import Results [YYYY-MM-DD HH-mm].zip” (see point 1 for date details).
 3. The email sender will be: **TBD**
 4. The email will have the subject “The application Automatic Customer Import Report [YYYY-MM-DD]” (see point 1 for date details).
 5. The email body will have the following text:

Hello,

Please find the results of today’s Automatic Customer Import run for the application attached.

Regards,
The application owner
 4. The email recipients are: **TBD**
 5. The email will be sent using the following email server settings:
 - Server: smtp.acme.com
 - Port number: 25
 - User name: n/a
 - Password: n/a
 - Authenticate: no
 - Use TLS: no
 - Use SSL: no
 6. If sending the email fails, the Automatic Customer Import will write error details to a log file.