Dialog Insight

Smart Marketing Catalyst

Scheduled Imports

CSV and Excel File Formats

12/23/2016

Table of Contents

1 Introduction

This document first presents the CSV and Excel file formats supported by Dialog Insight, as well as the data formats used for the different types of data used and exchanged in these files (numeric data, dates, etc.).

This document also describes how to properly prepare a CSV or Excel file so it complies with the format required by Dialog Insight's scheduled import system.

Please note that Dialog Insight can work with files that do not comply with this format but that any deviation from the supported format will require custom development to import the file. Contact your account manager for more information.

2 Valid formats - CSV and Excel

Dialog Insight's schedule import system accepts files in both Excel and CSV (Comma Separated Values) formats.

Although there are many variations of CSV formats for data exchange, Dialog Insight uses the CSV format described by the RFC 4180 document called, "Common Format and MIME Type for CSV Files", available at the following address: (http://www.ietf.org/rfc/rfc4180.txt).

3 Rules for importing CSV and Excel files

The following rules apply to both CSV and Excel files and must be respected when import files. Note that for importing CSV files, additional rules apply; these are presented under section 4 of this guide.

3.1 Datatype formatting

CSV and Excel files processed by Dialog Insight can only contain text. So, some types of data in these files must be formatted in a particular way to be recognized by Dialog Insight's import system and used for exports. This is the case for numerical or decimal values, true/false fields (boolean), and date and time fields.

3.1.1 Integers

Integer values should be represented as a sequence of numbers, without any thousands separator (space or comma) (ex: 12345678). Negative values must use a leading "-" character, with no space.

For example: **1234567** or **-9876543**

3.1.2 Decimals

Decimal values should be represented by an integer, followed by a point (".") and a series of numbers. Here also, there are no thousands separator.

For example: **12345.67** or **-98765.43**

3.1.3 True/False fields (boolean)

A true/false field is represented by a numerical value that must be zero (0) to indicate false, or one (1) to indicate true. No textual representation (true/false, yes/no, etc.) is accepted.

3.1.4 Dates and times

Date and time fields require a specific data format, to be selected amongst the following options:

- ✓ YYYY.MM.DD HH:mm:ss
- ✓ YYYY.MM.DD HH:mm
- ✓ YYYY.MM.DD

Where "YYYY" represents the year, (ex: 2010), "MM" the month (ex: 06), "DD" the day, "HH" the time, in 24-hour format, "mm" minutes, and "ss" seconds. The exact number of characters must be present, which means that the month of June must be represented by "06" and not just "6".

3.1.5 Non-standard dates

The scheduled import system in Dialog Insight supports the use of certain date formats that vary from those mentioned above. In such cases however, the file header must be properly pre-formatted to describe this format. If you ever encounter a situation where you cannot format a date in any of the mentioned formats, contact your account manager or project manager to discuss alternative solutions.

3.1.6 Maximum values

Note that the value shown in a date or numerical field cannot exceed the maximum values storable in the field associated to this column in the file. The maximum is defined by the SQL storage system used by Dialog Insight's databases.

As an example, integer type fields added to a contact's profile are stored in a 32-bit SQL integer field that supports values between -2147483648 and 2147483647. Another example is that date fields used in certain situations do not accept dates prior to 1753.

3.2 CSV or Excel file preparation for automatic mapping

Dialog Insight's scheduled import system lets you import data in a list of contacts or in a relational table in the application.

This section describes how to prepare a CSV or Excel file so that is can be import automatically in Dialog Insight, without having to manually map columns to fields.

3.2.1 Header line

In addition to complying with the format described in this document, your file must also provide the import system with the required data to help mapping the right column to the right field in the database, and indicate, if possible, how this data should be formatted.

The **first line** of the file must be a **header** that describes the data contained in file columns.

For each column, the header line must contain one of the following values:

3.2.2 "Normal" fields

Standard fields in the **project** (columns to be imported in a normal field in the database) must simply show the field code as the name, prefixed with "f_". For example, the First Name field, whose code is "FirstName", will be indicated in the header as: "f_FirstName".

For normal fields in **relational tables**, you do not have to add the prefix "f_", and should only use the field code. For example: "FirstName".

3.2.3 Key fields

Fields that are part of the primary key of the **project** must be prefixed with both "key_" and "f_". For example, if the project key is the email address in the field Email, (code: EMail), then the column must be named "key_f_EMail". If you project uses more than one key field in its primary key, then all these fields must be prefixed with "key_f_".

For key fields in **relational tables**, there is no need to add the "f_" prefix; just use the "key_" prefix. Example: "key_Email".

3.2.4 Mandatory fields

Note that all the fields that are used in your project's primary key MUST also be present in your file. For example, if your project uses the combination of two fields to create the primary key, then you must import these two fields, and not only one of them.

3.2.5 Fields to ignore

If there is a column in your file that you do not want to import, you simply have to leave the column name in the header empty. This way, the column will be ignored by the import.

Note: This option is only offered to ease the import when files are generated with tools leaving very little control; otherwise, it is strongly recommended NOT to use this option. It is always better to use a file that contains exactly what you want to import, as the volume of data transferred, as well as the required work to process the file, will be greatly reduced, thus improving its performance.

3.3 File validation and import setup

When setting up a scheduled import, Dialog Insight' service team will provide you with a list of all the exact field codes for your project. You can also see these codes in the field configuration pages of your project.

Once you have generated your first files, you can submit them to Dialog Insight's development team, who will validate format, encoding and data and setup the proper automated processes for your imports.

4 Additional rules for importing CSV files

The following rules apply only to CSV files, and complement those rules already mentioned for importing Excel and CSV files.

4.1 Data formatting

4.1.1 File preparation

The CSV format requires specific rules to be followed when preparing a file to be imported:

- ✓ The file must contain one record per line.
- ✓ Each line must end with a line ending "CRLF".
- ✓ The comma is used as field separator, and the quotation mark as field delimiter.
- ✓ The field delimiter is optional for fields that do not contain special characters.
  Example: 123,456,"789",ABC,"DEF"
- ✓ The field delimiter is mandatory for fields containing commas, quotation marks, or any form of line break or vertical spacing (CR, LF, VT, etc.).
  Example: a field containing **ab"c** becomes **"ab""c"**
  Therefore: 123,456,"789","ab,c","DEF"

  In addition, quotation marks in a delimited field must be doubled.
  Example : **ab"c** becomes **"ab""c"**
  Therefore: 123,456,"789","ab""c","DEF"

4.1.2 File reading

Reading the file must be done in stream mode, which means that it is better to analyze one character at a time than one line at a time. The reason for this is that a line break could appear within a delimited field, without corresponding to the end of the record.

- ✓ If the file starts with a quotation mark, then:
  - The field will end only if it encounters a quotation mark *that is not part of a doubled quotation mark*, regardless of any other characters shown before this field end marker.
  - Both start and end quotation marks will be ignored.
  - Any pair of 2 consecutive quotation marks in a field will need to be replaced by one single quotation mark.

- ✓ The first `CRLF` pair encountered outside of a delimited filed identifies the end of the record.

- ✓ The last record of a file can end with `EOF` instead of `CRLF`, but if the end of the file is reached and the delimited field is not closed, then this will cause an error.

- ✓ No single `CR` or `LF` (outside of a `CRLF` pair), or vertical spacing character (`VT`) is allowed outside of a delimited field and will cause an error.

4.2 File encoding

CSV files can, in theory, use any type of encoding, as long as both parties have agreed in advance on the encoding to use.

Dialog Insight's scheduled import system can use a variety of encodings, except for manual imports, which are for now limited to importing files using utf-8 or iso-8859-1 encoding.

4.2.1 Detection of utf-8 encoding during imports

The Dialog Insight application allows two types of file imports: manual (launched manually by a user) and scheduled (recurrent process that retrieves a file and imports it based on a defined schedule).

For history purposes, when importing a file manually, the file is deemed to be encoded using iso-8859-1, unless it contains a marker that indicates it as being encoded in utf-8. This utf-8 Bite-Order-Mark is represented by a chain of 3 characters (0xEF, 0xBB, 0xBF) that is injected at the beginning of the file and serves as a signature for the utf-8 format. For more details, visit: http://en.wikipedia.org/wiki/Byte_order_mark.

Scheduled imports can define precisely the encoding used for each imported file. The possible encodings include, in addition to iso-8859-1 and utf-8, utf-32 and utf-16 in both byte orders (Big-endian and Little-endian).

4.2.2 Files generated by Dialog Insight

Since Dialog Insight's databases accept UNICODE data, files that are exported use by default the utf-8 format and the Byte Order Mark will be included at the beginning of the file. This ensures that all characters contained in the database can be represented in the file. Some manual export operations in the application accept the iso-8859-1 format, but be aware that the export will fail if there are characters in the data to export that cannot be represented in iso-8859-1.

4.2.3 Choosing the right encoding

Although Dialog Insight supports iso-8859-1 encoding, it is limited to characters that are specific to the North American and European alphabets, and we do not recommend its use in modern databases. A simple reason for not using it is that this encoding does not support styled apostrophes and quotation marks often used (’ instead of '). However, using iso-8859-1 is accepted if you are exporting data from older databases that do not support UNICODE.

Contact
Canada: 1 866 529-6214
France: 01 84 88 40 66
Russia: +7 (495) 226-04-11

Email: info@dialoginsight.com
Website: www.dialoginsight.com
Blog: www.dialoginsight.com/en/resources/academy

@DialogInsight
Dialog Insight