



Catalyseur d'intelligence marketing

## Imports automatisés

Formats de fichiers CSV et Excel

23/12/2016



Canada • France • Russie

[dialoginsight.com](http://dialoginsight.com)

## Table des matières

1	Introduction.....	2
2	Formats valides – CSV et Excel.....	3
3	Règles pour l’import de fichiers Excel et CSV.....	3
3.1	Format des différents types de données.....	3
3.1.1	Nombres entiers.....	3
3.1.2	Nombres décimaux.....	3
3.1.3	Champs vrai/faux (booléen).....	3
3.1.4	Dates et heures.....	4
3.1.5	Dates non-standard.....	4
3.1.6	Valeurs limites.....	4
3.2	Préparation d’un fichier Excel ou CSV pour une assignation automatique.....	5
3.2.1	Ligne d’en-tête.....	5
3.2.2	Champs « ordinaires ».....	5
3.2.3	Champs clés.....	5
3.2.4	Champs obligatoires.....	6
3.2.5	Champs à ignorer.....	6
3.3	Validation du fichier et mise en place.....	6
4	Règles additionnelles pour l’import de fichiers CSV.....	7
4.1	Formatage des données.....	7
4.1.1	Préparation d’un fichier.....	7
4.1.2	Lecture d’un fichier.....	8
4.2	Encodage des fichiers.....	8
4.2.1	Détection de l’encodage utf-8 lors des imports.....	8
4.2.2	Fichiers produits par Dialog Insight.....	9
4.2.3	Choisir le bon encodage.....	9



## 1 Introduction

Ce document décrit en un premier temps les formats de fichier CSV et Excel, tel que supportés par Dialog Insight, ainsi que les formats de données utilisés pour les différents types de données qui seront échangés dans ces fichiers (données numériques, dates, etc.).

Le document décrit ensuite comment préparer adéquatement un fichier CSV ou Excel pour qu'il soit conforme aux attentes du système d'imports automatisés de Dialog Insight.

Il est à noter que Dialog Insight peut travailler avec des fichiers qui ne respectent pas ces normes, mais que toute déviation de ce standard entraînera possiblement des travaux sur mesure pour le traitement des fichiers. Contactez votre directeur de compte pour plus de détails.



## 2 Formats valides – CSV et Excel

Le système d'imports automatisés de Dialog Insight acceptent comme fichier à importer les formats Excel et « CSV » (acronyme de « Comma Separated Values », c'est-à-dire que les valeurs sont séparées par des virgules).

Bien qu'il existe historiquement plusieurs façons distinctes de préparer des fichiers CSV, Dialog Insight utilise le format CSV tel que standardisé dans le document RFC 4180 disponible à l'adresse suivante : <http://www.ietf.org/rfc/rfc4180.txt>

## 3 Règles pour l'import de fichiers Excel et CSV

Les règles suivantes s'appliquent tout autant pour l'import de fichiers CSV, que pour l'import de fichiers en format Excel. Notez que pour importer des fichiers en format CSV, des règles additionnelles s'appliquent, et sont présentées à la section 4 du présent guide.

### 3.1 Format des différents types de données

Les fichiers Excel et CSV traités par Dialog Insight ne doivent contenir que du texte. Certains types de données qui seront échangées dans ces fichiers doivent donc être formatés en texte et le système Dialog Insight supporte des formats précis qui sont requis lors de l'import et utilisés à l'export. C'est le cas pour les valeurs numériques ou décimales, les champs « vrai/faux » (booléen) et les champs de type « date & heure ».

#### 3.1.1 Nombres entiers

Les nombres entiers doivent être représentés par une séquence de chiffres, sans aucun séparateur de milliers (espaces ou virgule). Un trait d'union court (« - ») doit préfixer immédiatement le nombre pour représenter une valeur négative.

Par exemple : **1234567** ou **-9876543**

#### 3.1.2 Nombres décimaux

Un nombre décimal est représenté comme un nombre entier, suivi d'un point (« . ») et d'une série de chiffres. Encore une fois, aucun séparateur de millier n'est permis.

Par exemple : **12345.67** ou **-98765.43**

#### 3.1.3 Champs vrai/faux (booléen)

Un champ vrai/faux est représenté par une valeur numérique qui doit être zéro (0) pour indiquer faux, ou un (1) pour indiquer vrai. Aucune représentation textuelle (vrai/faux, oui/non, on/off, etc.) n'est acceptée.



### 3.1.4 Dates et heures

Les champs Date et Heure requièrent un format de données précis choisi parmi les options suivantes :

- ✓ AAAA.MM.JJ HH:mm:ss
- ✓ AAAA.MM.JJ HH:mm
- ✓ AAAA.MM.JJ

Où « AAAA » représente l'année (ex : 2010), « MM » le mois (ex : 06), « JJ » jour, « HH » l'heure en format 24 heures, « mm » les minutes et « ss » les secondes. Le nombre de caractères exacts doivent être présents, c'est-à-dire par exemple que Juin (le 6e mois) est représenté par « 06 » et non « 6 ».

### 3.1.5 Dates non-standard

Le système d'imports automatisés de l'application Dialog Insight supporte l'utilisation de certains formats de date différents de ceux énoncés ci-dessus. Dans ce contexte cependant, l'en-tête du fichier doit être pré-formatée adéquatement pour décrire ce format. Si vous vous retrouvez dans une situation où il vous est impossible de formater une date dans l'un des formats ci-dessus, communiquez avec votre directeur de compte ou chargé de projet pour discuter des alternatives.

### 3.1.6 Valeurs limites

Il est à noter que la valeur illustrée dans un champ numérique ou date ne peut pas dépasser les valeurs maximales stockables dans le champ qui sera associé à cette colonne du fichier. Il s'agit ici des limites définies par le système de stockage SQL utilisé par les bases de données de Dialog Insight.

À titre d'exemple, les champs de type « nombre entier » ajoutés à la fiche d'un contact sont stockés dans un champ SQL de type « nombre entier » 32 bits et supportent donc des valeurs entre -2147483648 et 2147483647. Un autre exemple est que les champs de type « date » utilisés dans certaines situations n'acceptent pas de date inférieure à l'an 1753.



## 3.2 Préparation d'un fichier Excel ou CSV pour une assignation automatique

Dialog Insight offre un système d'imports automatisés qui permet d'importer des données dans une liste de contacts, ou dans une table relationnelle, de l'application Dialog Insight.

Cette section décrit comment préparer un fichier Excel ou CSV pour qu'il puisse être importé automatiquement dans l'application Dialog Insight, sans devoir effectuer d'assignation manuelle des colonnes

### 3.2.1 Ligne d'en-tête

En plus de respecter le format décrit dans ce document, votre fichier doit aussi transmettre au système d'import les données nécessaires qui l'aideront à assigner les bonnes données de votre fichier aux bons champs de la base de données, en indiquant optionnellement comment ces données sont formatées.

La **première ligne** du fichier doit donc être un **en-tête** qui décrit les données contenues dans les colonnes du fichier.

Pour chaque colonne, la ligne d'en-tête doit contenir l'une des valeurs suivantes :

### 3.2.2 Champs « ordinaires »

Les champs standards du **projet** (colonnes à être importées dans un champ ordinaire de la base de données) doivent simplement porter comme nom le code du champ, précédé du préfixe « f\_ ». Par exemple, le champ Prénom (code : FirstName) sera indiqué comme suit : f\_FirstName.

Pour les champs ordinaires de **tables relationnelles**, vous ne devez pas ajouter le préfixe « f\_ », mais utilisé seulement le code du champ. Exemple : « FirstName ».

### 3.2.3 Champs clés

Les champs qui font partie de la clé du **projet** doivent, en plus d'être préfixés par « key\_ », être précédé du préfixe « f\_ ». Donc, par exemple, si la clé de votre projet est le champ courriel (code : EMail), alors la colonne sera nommée « key\_f\_EMail ». Si votre projet utilise plus d'un champ dans sa clé primaire, tous ces champs doivent inclure le préfixe « key\_f\_ ».

Pour les champs clés de **tables relationnelles**, vous ne devez pas ajouter le « f\_ », mais uniquement le préfixe « key\_ ». Exemple : « key\_EMail ».



### 3.2.4 Champs obligatoires

Il est à noter que tous les champs faisant partie de la clé unique de votre projet doivent obligatoirement faire partie de votre fichier. Ainsi, si par exemple votre projet utilise la combinaison de deux champs pour définir la clé unique de votre projet, alors il est interdit de faire un import utilisant un seul (ou aucun) de ces champs.

### 3.2.5 Champs à ignorer

Si une colonne présente dans votre fichier doit être ignorée par l'import, il suffit de laisser complètement vide le nom de la colonne dans la ligne d'en-tête. Le système d'import ignorera cette colonne à l'import.

Note : Cette option existe pour faciliter l'import lorsque les fichiers sont produits avec des outils sur lesquels le client a peu de contrôle; autrement, il est fortement recommandé de ne PAS utiliser cette option. Il est préférable de toujours produire un fichier qui contient exactement ce qui doit être importé, ainsi le volume de données transférées, ainsi que la charge de travail requise pour traiter le fichier, seront réduites, et par conséquent, améliorera la performance.

## 3.3 Validation du fichier et mise en place

Lors de la mise en place d'un import automatisé, notre équipe de service vous fournira la liste des codes de champs exacts correspondant à votre projet. Vous pouvez aussi consulter la liste de ces codes dans l'interface de configuration des champs de votre projet.

Une fois vos premiers fichiers produits, vous pourrez les remettre à notre équipe de développement, qui fera une validation du format, de l'encodage et des données, et mettra alors en place les procédures automatisées appropriées.



## 4 Règles additionnelles pour l'import de fichiers CSV

Les règles suivantes s'appliquent uniquement aux fichiers CSV, et s'ajoutent à celles mentionnées précédemment pour l'import de fichiers Excel et CSV.

### 4.1 Formatage des données

#### 4.1.1 Préparation d'un fichier

Le format CSV exige une façon précise d'écrire les données dans le fichier :

- ✓ Le fichier doit contenir un enregistrement par ligne.
- ✓ Chaque ligne doit se terminer par une fin de ligne « CRLF ».
- ✓ Le séparateur de champ est la virgule, et le délimiteur de champ est le guillemet.
- ✓ Le délimiteur de champ est optionnel pour les champs ne contenant pas de caractères spéciaux.  
Exemple : 123,456,"789",ABC,"DEF"
- ✓ Le délimiteur de champ est obligatoire si le champ contient une virgule, un guillemet, ou toute forme de saut de ligne ou espacement vertical (CR, LF, VT, etc.).  
Exemple : un champ contenant **ab,c** devient obligatoirement **"ab,c"**  
Donc : 123,456,"789","ab,c","DEF"

De plus, tout guillemet contenu dans le champ délimité doit être doublé.

Exemple : **ab"c** devient **"ab""c"**

Donc : 123,456,"789","ab""c","DEF"





### 4.1.2 Lecture d'un fichier

La lecture d'un fichier devrait se faire en mode « stream », c'est-à-dire qu'il est préférable d'analyser un caractère à la fois, et non une ligne à la fois. La raison est qu'un saut de ligne peut apparaître à l'intérieur d'un champ délimité sans toutefois représenter la fin d'un enregistrement.

- ✓ Si le champ lu commence par un guillemet, alors :
  - Le champ se terminera seulement lorsqu'un guillemet *qui ne fait pas partie d'une paire de 2 guillemets consécutifs* est rencontré, peu importe les autres caractères rencontrés avant d'arriver à ce marqueur de fin de champ.
  - Le guillemet initial et le guillemet final devront être ignorés.
  - Toute paire de 2 guillemets consécutifs dans le champ devront être remplacés par un seul guillemet.
- ✓ La première paire CRLF rencontrée à l'extérieur d'un champ délimité signifie la fin de l'enregistrement.
- ✓ Le dernier enregistrement du fichier peut se terminer par EOF au lieu d'un CRLF; mais si la fin du fichier est atteinte et que le champ délimité n'est pas, cela causera une erreur.
- ✓ Aucun CR ou LF présent seul (hors d'une paire CRLF) ni aucun caractère d'espacement vertical (VT) n'est permis en à l'extérieur d'un champ délimité; leur présence est donc une erreur.

## 4.2 Encodage des fichiers

Les fichiers CSV peuvent, en théorie, utiliser n'importe quel encodage, en autant que l'encodage à utiliser soit convenu d'avance entre les deux parties.

Le système d'imports automatisés de Dialog Insight est en mesure d'utiliser une variété d'encodages, mais les imports manuels sont pour l'instant limités à l'import de fichiers encodés en utf-8 ou iso-8859-1.

### 4.2.1 Détection de l'encodage utf-8 lors des imports

L'application Dialog Insight permet deux types d'imports de fichiers : l'import manuel (procédure lancée à la main par un utilisateur) et l'import automatisé (procédure récurrente qui récupère un fichier et l'importe selon un horaire établi).



Pour des raisons historiques, lors d'un import manuel, le fichier sera réputé être encodé en iso-8859-1 à moins qu'il ne contienne un marqueur qui permette de reconnaître l'encodage utf-8. Ce « Byte Order Mark » ou « Préambule » utf-8 est une chaîne de 3 octets (0xEF 0xBB 0xBF en hexadécimal) injectée au début du fichier et qui sert de signature pour reconnaître l'encodage utf-8. Pour plus de détails, visitez :

[http://en.wikipedia.org/wiki/Byte\\_order\\_mark](http://en.wikipedia.org/wiki/Byte_order_mark).

Les imports automatisés permettent de définir explicitement l'encodage qui sera utilisé pour chaque fichier échangé. Les encodages possibles incluent, en plus d'iso-8859-1 et utf-8, les encodages utf-32 et utf-16 dans les deux ordres de bytes possibles (« Big-endian » et « Little-endian »).

#### 4.2.2 Fichiers produits par Dialog Insight

Puisque les bases de données de Dialog Insight acceptent des données UNICODE, les fichiers produits à l'export utiliseront par défaut l'encodage utf-8 et le « Byte Order Mark » sera inclus au début du fichier. Ceci permet d'assurer que tous les caractères contenus dans la base de données pourront être représentés dans le fichier. Certaines fonctions d'exports manuels dans l'application permettent de choisir un encodage iso-8859-1, mais soyez avisé que l'export échouera si des caractères contenus dans les données à exporter ne peuvent pas être représentés en iso-8859-1.

#### 4.2.3 Choisir le bon encodage

Bien que Dialog Insight supporte l'encodage iso-8859-1, cet encodage est restreint à des caractères très spécifiques de l'alphabet nord-américain et européen; nous ne recommandons donc pas son utilisation pour des bases de données modernes. Un exemple fort simple est que cet encodage ne supporte pas les apostrophes ou guillemets stylisés qui sont utilisés couramment ( ' au lieu de ' ). Le choix de l'encodage iso-8859-1 est cependant acceptable lorsque vous devez transmettre des données à partir d'anciennes bases de données qui ne supportent pas UNICODE.

#### Contact

Canada : 1 866 529-6214


France : 01 84 88 40 66

Russie : +7 (495) 226-04-11

Courriel : [info@dialoginsight.com](mailto:info@dialoginsight.com)

Site Web : [www.dialoginsight.com](http://www.dialoginsight.com)

Blogue : [www.dialoginsight.com/fr/ressources/academie](http://www.dialoginsight.com/fr/ressources/academie)

 @DialogInsight

 Dialog Insight

