

# L'open data et les formats de fichiers

## Table des matières

Qu'est-ce que l'Open Data ?.....	2
Quels sont les formats d'échange ?.....	4
Les formats ouverts.....	4
Qu'est-ce que le format csv ?.....	5
Les données brutes.....	5
Le jeu de caractères.....	5
Les séparateurs.....	6

Université du Littoral Côte d'Opale



Licence Creative Commons

Attribution - Partage dans les Mêmes Conditions

## Qu'est-ce que l'Open Data ?

Les services publics, les collectivités, les associations ou les entreprises collectent et traitent de nombreuses données numériques. Ces *données non personnelles* peuvent être mises à disposition de tous et réutilisées librement.

Par exemple, il peut s'agir :

- du recensement des installations sportives en France sur [data.gouv.fr](https://data.gouv.fr) ;

The screenshot shows the data.gouv.fr interface for the dataset 'Recensement des équipements sportifs, espaces et sites de pratiques'. The page title is 'RECENSEMENT DES EQUIPEMENTS SPORTIFS > FICHES INSTALLATIONS - T4 2017'. It includes a 'TÉLÉCHARGER' button and a 'DÉTAILS' button. The 'Informations' section lists 'Openfield 16' and 'Licence Ouverte / Open Licence'. The 'Producteur' section shows the logo of the 'MINISTÈRE DES SPORTS'.

- des livres les plus empruntés dans les bibliothèques parisiennes sur l'[Open Data de Paris](https://data.paris.fr) ;

The screenshot shows the PARISDATA website for the dataset 'Les titres les plus prêtés dans les bibliothèques de prêt'. The page title is 'Les titres les plus prêtés dans les bibliothèques de prêt'. It includes a 'Licence Open Database License (ODbL)' and a 'Producteur' as 'Mairie de Paris / Direction des Affaires Culturelles'.

- des disponibilités en temps réel des stations V-Lille sur l'[Open Data](#) de la métropole lilloise ;

**MEL** LES **DONNÉES OUVERTES**  
DE LA **MÉTROPOLÉ EUROPÉENNE DE LILLE**

ACCUEIL DONNÉES DÉMARCHE CARTOGRAPHE

### V'Lille - Disponibilité en temps réel

Informations Tableau Carte Analyse Export API

Disponibilité des vélos et des places de vélos en libre service en temps réel  
Fréquence de mise à jour : 1 minute

**Identifiant du jeu de données** ville-realtime

**Thèmes** Transports, Déplacements

**Mots clés** Vélo, V'Lille, VLS, Transpole, Temps réel

**Licence** Licence Ouverte (Etalab)

**Producteur** MEL

Ce jeu de données pourrait être utilisé pour alimenter une application pour smartphone

Ces données sont associées à des *licences ouvertes* : elles peuvent être réutilisées par tous sans aucune restriction. Il s'agit de **Données Ouvertes** (ou **Open Data**).

Pour faciliter la réutilisation des données, il est préférable qu'elles soient diffusées dans un format exploitable par tous.

## Quels sont les formats d'échange ?

En général, ces données ouvertes sont proposées sous différents formats.



### Les formats ouverts

Certains de ces formats sont des **formats ouverts**<sup>1</sup> favorisant l'*interopérabilité* : les données pourront être réutilisées indépendamment du logiciel qui a été utilisé pour les créer.

Les principaux formats ouverts sont :

- Texte brut : ASCII (.txt) / sans extension
- Texte formaté : TeX (.tex), OpenDocument Text (.odt), Hypertext Markup Language (.htm ou .html), XHTML (.xhtml), Feuilles de style en cascade (.css)
- Tableur : OpenDocument Spreadsheet (**.ods**)
- Document imprimable : Document PDF (.pdf)
- Livre numérique : EPUB (.epub)
- Données brutes : CSV (**.csv**) / sans extension, JSON (**.json**), XML (**.xml**)
- Données géographiques : KML (.kml), SHP (.shp)

Pour en savoir plus, consultez l'article [Format Ouvert sur Wikipédia](#).

<sup>1</sup> son mode de représentation a été rendu public par son auteur et aucune entrave légale ne s'oppose à son utilisation

# Qu'est-ce que le format csv ?

## Les données brutes

Le **format csv** (*comma-separated values*) est souvent utilisé pour échanger des données brutes. Il s'agit d'un *fichier texte* composé uniquement de caractères (dont l'espace et le retour à la ligne).

Ces données textuelles peuvent être présentées sous forme de tableau : chaque ligne du texte correspond à une ligne du tableau et on utilise des séparateurs (virgule, point-virgule, tabulation ou autre) pour les colonnes.

Exemple d'un fichier csv ouvert avec un éditeur de texte :

```
ID;nom;adresse;commune
219;GARE JEAN LEBAS;PLACE DE LA GARE;ROUBAIX
227;BVRD ARMENTIERES;1, BOULEVARD D'ARMENTIÈRES;ROUBAIX
247;ROUGET DE L ISLE;SECTEUR BVRD DE L'EGALITÉ;TOURCOING
```

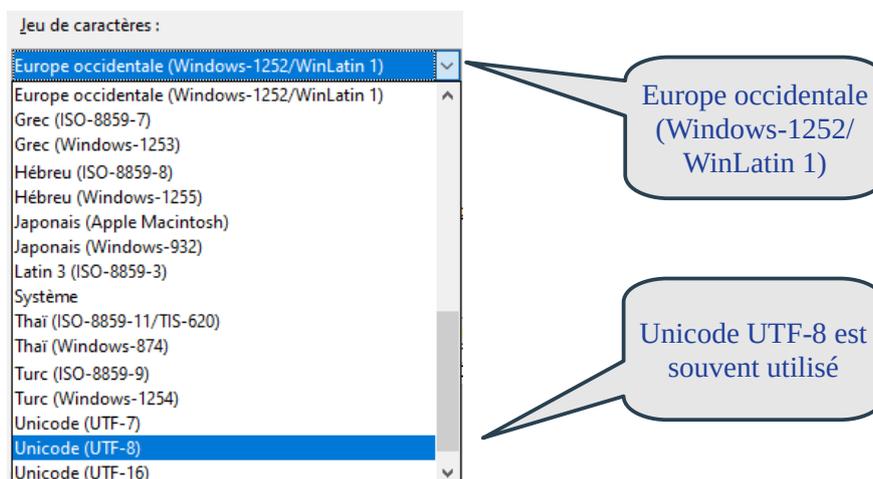
Exemple du même fichier csv ouvert avec un tableau :

	A	B	C	D
1	ID	nom	adresse	commune
2	219	GARE JEAN LEBAS	PLACE DE LA GARE	ROUBAIX
3	227	BVRD ARMENTIERES	1, BOULEVARD D'ARMENTIÈRES	ROUBAIX
4	247	ROUGET DE L ISLE	SECTEUR BVRD DE L'EGALITÉ	TOURCOING

L'ouverture d'un fichier au format csv peut nécessiter quelques paramètres.

## Le jeu de caractères

- Le jeu de caractères utilisé pour l'encodage doit être précisé.



Import de texte - [ville-realtime.csv]

**Importer**

Jeu de caractères : Europe occidentale (Windows-1252/WinLatin 1)

Langue : Par défaut - Français (France)

À partir de la ligne : 1

**Champs**

Type de colonne :

	Standard	Standard	Standard	Standard
1	ID	nom	adresse	commune
2	219	GARE JEAN LEBAS	PLACE DE LA GARE	ROUBAIX
3	227	BVRD ARMENTIERES	1, BOULEVARD D'ARMENTIERES	ROUBAIX
4	247	ROUGET DE L ISLE	SECTEUR BVRD DE L'EGALITE	TOURCOING

Le jeu de caractères doit correspondre à celui de l'encodage

Anomalie constatée

## Les séparateurs

- Le type de séparateur doit être indiqué : on peut en sélectionner un, plusieurs ou en choisir un spécifique.

Import de texte - [ville-realtime.csv]

**Importer**

Jeu de caractères : Unicode (UTF-8)

Langue : Par défaut - Français (France)

À partir de la ligne : 1

**Options de séparateur**

Largeur fixe  Séparé par

Tabulation  Virgule  Point-virgule  Espace  Autre

Fusionner les séparateurs Séparateur de texte : "

**Champs**

Type de colonne :

	Standard	Standard	Standard	Standard
1	ID	nom	adresse	commune
2	219	GARE JEAN LEBAS	PLACE DE LA GARE	ROUBAIX
3	227	BVRD ARMENTIERES	1 BOULEVARD D'	ROUBAIX
4	247	ROUGET DE L ISLE	SECTEUR BVRD DE L'EGALITE	TOURCOING

Dans ce cas, il faut éviter de choisir la virgule comme séparateur car elle peut être présente dans l'adresse

	A	B	C	D	E
1	ID	nom	adresse	commune	
2	219	GARE JEAN LEBAS	PLACE DE LA GARE	ROUBAIX	
3	227	BVRD ARMENTIERES	1 BOULEVARD D'ARMENTIERES	ROUBAIX	
4	247	ROUGET DE L ISLE	SECTEUR BVRD DE L'EGALITE	TOURCOING	

Attention, le format csv est un **format très basique**. Une fois les données importées dans une feuille de calcul, il est conseillé d'enregistrer le fichier dans un format de type classeur (.ods par exemple) pour conserver les modifications éventuelles : la mise en forme, les formules ou l'organisation sur plusieurs feuilles.