

# Les données structurées au format CSV

---

On trouve énormément de données sur internet. Une partie de ces données sont publiques, par exemple le site [data.gouv.fr](https://data.gouv.fr) recense un grand nombre de données publiques. Ces données sont librement réutilisables.

Afin de découvrir ce qu'est "l'open data", nous allons utiliser le site [data.gouv.fr](https://data.gouv.fr). Résumez en quelques lignes ce que vous aurez appris en lisant la page suivante : <https://doc.data.gouv.fr/>.

Sur le site [data.gouv.fr](https://data.gouv.fr). Recherchez les données "Opérations coordonnées par les CROSS" à l'aide du moteur de recherche proposé par le site  
Vous pouvez constater que ces données sont au format CSV

Le format CSV est très courant. Voici ce que nous dit Wikipédia sur le format CSV :

***Comma-separated values, connu sous le sigle CSV, est un format informatique ouvert représentant des données tabulaires sous forme de valeurs séparées par des virgules.***

***Un fichier CSV est un fichier texte, par opposition aux formats dits « binaires ». Chaque ligne du texte correspond à une ligne du tableau et les virgules correspondent aux séparations entre les colonnes. Les portions de texte séparées par une virgule correspondent ainsi aux contenus des cellules du tableau.***

Voici un exemple du contenu d'un fichier CSV :

```
nom,prenom,date_naissance
Durand,Jean-Pierre,23/05/1985
Dupont,Christophe,15/12/1967
Terta,Henry,12/06/1978
```

"nom", "prenom" et "date\_naissance" sont appelés des descripteurs alors que, par exemple, "Durand", "Dupont" et "Terta" sont les valeurs du descripteur "nom".

ATTENTION :

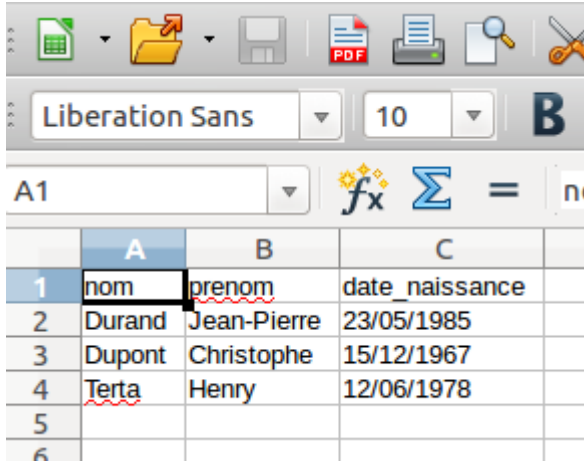
La virgule est un standard pour les données anglo-saxonnes, mais pas pour les données aux normes françaises. En effet, en français, la virgule est le séparateur des chiffres décimaux. Il serait impossible de différencier les virgules des décimaux et les virgules de séparation des informations. C'est pourquoi on utilise un autre séparateur : le point-virgule (;). Dans certains cas cela peut engendrer quelques problèmes, vous devrez donc rester vigilants sur le type de séparateur utilisé.

Les tableurs, tels que "Calc" (Libre Office), sont normalement capables de lire les fichiers au format CSV. J'ai précisé "normalement" car certains tableurs gèrent mal le séparateur CSV "point-virgule" et le séparateur des chiffres décimaux "virgule".

## Expérimentation :

Téléchargez l'archive csv.zip qui se trouve ici : <http://ressource.elec.free.fr/docs/softs/NSI/csv.zip>  
Après avoir décompressé l'archive, ouvrez le fichier **ident\_pointVirgule.csv** à l'aide d'un tableur (Excel ou Calc Open office).

Vous devriez obtenir ceci :



	A	B	C
1	nom	prenom	date_naissance
2	Durand	Jean-Pierre	23/05/1985
3	Dupont	Christophe	15/12/1967
4	Terta	Henry	12/06/1978
5			
6			

Vous pouvez constater que les données sont bien "rangées" dans un tableau avec des lignes et des colonnes (voilà pourquoi on parle de données tabulaires).

Ouvrez aussi le fichier avec le bloc note pour voir comment sont écrites les données. Vous voyez qu'elles sont séparées par des points-virgules.

Il est possible de trouver sur le web des données beaucoup plus intéressantes à traiter que celles contenues dans le fichier "ident\_pointVirgule.csv". Par exemple, le site [sql.sh](http://sql.sh), propose un fichier csv contenant des informations sur l'ensemble des communes françaises.

## Expérimentation :

Ouvrez le fichier **ville\_point\_virgule.csv** à l'aide d'un tableur (c'est une version légèrement modifiée de celle disponible sur le site [sql.sh](http://sql.sh), on y a notamment ajouté des entêtes).

Comme vous pouvez le constater, nous avons 12 colonnes (et 36700 lignes si on ne compte pas l'entête !), voici la signification de ces colonnes :

*dep : numéro de département*

*nom : nom de la commune*

*cp : code postal*

*nb\_hab\_2010 : nombre d'habitants en 2010*

*nb\_hab\_1999 : nombre d'habitants en 1999*

*nb\_hab\_2012 : nombre d'habitants en 2012 (approximatif)*

*dens : densité de la population (habitants par kilomètre carré)*

*surf : superficie de la commune en kilomètre carré*

*long : longitude*

*lat : latitude*

*alt\_min : altitude minimale de la commune (il manque des données pour certains territoires d'outre-mer)*

*alt\_max : altitude maximale de la commune (il manque des données pour certains territoires d'outre-mer)*

## expérimentation :

déterminez l'altitude maximale et l'altitude minimale de votre commune.