

Les données structurées

I Où récupérer des données ?

On trouve énormément de données sur internet et une partie de ces données est publique.

Par exemple sur le site data.gouv.fr. Ces données sont librement réutilisables, c'est de l'**open data**.

À la page suivante: <https://doc.data.gouv.fr/>, lire le paragraphe **Découvrir l'open data**.

II Le format CSV

Sur le site data.gouv.fr rechercher, à l'aide du moteur de recherche proposé par le site, les **Données hospitalières relatives à l'épidémie de COVID-19**

Observer le format des fichiers mis à disposition. Indiquer leur extension :

Voici ce qu'en dit [Wikipédia](https://fr.wikipedia.org/wiki/Comma-separated_values) :

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, (...) 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
```

Nous avons ici 3 personnes : Jean-Pierre Durand qui est né le 23/05/1985
 Christophe Dupont qui est né le 15/12/1967
 Henry Terta qui est né le 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".

Remarque : La virgule est un standard anglo-saxon, mais pas français ! En effet, en français, la virgule est le séparateur des chiffres décimaux (alors que c'est un point en notation anglo-saxonne). 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 capables de lire les fichiers au format CSV.

Exercice 1 :

- 1) Télécharger le fichier **liste.csv** puis l'ouvrir à l'aide du tableur LibreOffice Calc.

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			

- 2) Fermer le tableur puis ouvrir de nouveau le fichier **liste.csv** à l'aide d'un **éditeur de texte** (Clic droit/Ouvrir avec/Wordpad ou Notepad ou Notepad++).
- 3) Ajouter votre nom, prénom et date de naissance sur la ligne suivante en respectant le séparateur puis enregistrer le fichier obtenu sous le nom **listenouvelle.csv**.
- 4) Ouvrir maintenant le fichier **listenouvelle.csv** avec le tableur et constater les changements.

Il est possible de trouver sur le web des données beaucoup plus intéressantes à traiter que celles contenues dans le fichier précédent. Le site sql.sh, propose un fichier csv contenant des informations sur l'ensemble des communes françaises.

Télécharger le fichier [Liste des villes françaises au format CSV \(8.1Mo\)](#)

Exercice 2 :

- 1) Ouvrir ce fichier à l'aide d'un tableur. Il comporte 36 700 lignes et 27 colonnes.
Il ne comporte pas de descripteurs mais le site explique les données.
- 2) Rechercher pour votre commune de résidence puis pour Cherbourg et Font-Romeu les douze informations suivantes :
 - dep : numéro de département
 - nom : nom de la commune
 - cp : code postal
 - nb_hab_2010 : nombre d'habitants en 1999
 - nb_hab_1999 : nombre d'habitants en 2010
 - 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 en degrés, minutes et secondes
 - lat : latitude en degrés, minutes et secondes
 - alt_min : altitude minimale de la commune
 - alt_max : altitude maximale de la commune
- 3) À l'aide de l'éditeur de texte créer un fichier csv avec les réponses. N'oubliez pas les descripteurs ce qui vous fera donc 4 lignes dans le fichier.
- 4) Déposer le fichier dans **Devoir 13 Données structurées**.