


On trouve énormément de **données** sur internet. Une partie de ces **données** sont publiques, par exemple le site `data.gouv.fr`, recense un grand nombre de données publiques. Ces données sont librement réutilisables.

1^{re} partie : logiciel libre, Open Data

1. Faire une recherche concernant les logiciels libres. Vous répondrez aux questions suivantes : définition, utilisation, intérêt.
2. Voici la licence associée à nos ressources : . Quelle est la signification de cette licence ? Expliquez en quelques lignes.
3. Donnez une définition simple de l'Open Data.
4. Afin de découvrir ce qu'est « l'Open Data », allez sur le site `data.gouv.fr`. En haut et à droite de la page d'accueil, cliquez sur l'onglet « documentation » puis sur résumez en quelques lignes ce que vous aurez appris en lisant cette page.

2^e partie : le format CSV

1. Sur le site `data.gouv.fr`, recherchez les données « Opérations coordonnées par les CROSS » à l'aide du moteur de recherche proposé par le site puis donnez la signification de l'abréviation CROSS. Vous pouvez constater que ces données sont au format csv. Le format csv est très courant sur internet, nous allons l'étudier en premier. 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 :

```
1 nom, prenom, date-naissance
2 Ainhua,Alan,23/05/2005
3 Otega,Loïc,15/12/2007
4 Barsacq,Lorena,12/06/2008
```

Il semble évident que nous avons ici 3 personnes :

- Alan Ainhua qui est né le 23/05/2005
- Loïc Otega qui est né le 15/12/2007
- Lorena Barsacq qui est née le 12/06/2008

« nom », « prenom » et « date_naissance » sont appelés des **descripteurs** alors que, par exemple, "Ainhua", "Otega" et "Barsacq" sont les valeurs du descripteur « nom ».

- Préciser les différentes valeurs du descripteur "date_naissance".

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. "Normalement" car certains tableurs gèrent mal le séparateur CSV « point-virgule » et le séparateur des chiffres décimaux « virgule ».

Après avoir téléchargé le fichier `ident_virgule.csv`, ouvrez ce dernier à l'aide d'un tableur. Si par hasard votre tableur ne gère pas correctement le fichier avec le séparateur « point-virgule », voici une version « séparateur virgule » du fichier : `ident_virgule.csv` Dans la suite, gardez toujours cet éventuel problème à l'esprit (surtout avec des données « made in France »)

Vous devriez obtenir ceci :

| | A | B | C | D |
|---|---------|--------|----------------|---|
| 1 | nom | prenom | date_naissance | |
| 2 | Ainhoa | Alan | 23/05/2005 | |
| 3 | Otega | Loïc | 15/12/2007 | |
| 4 | Barsacq | Lorena | 12/06/2008 | |
| 5 | | | | |

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). Il est possible de trouver sur le web des données beaucoup plus intéressantes à traiter que celles contenues dans le fichier « `ident_virgule.csv` ». Par exemple, le site sql.sh, propose un fichier csv contenant des informations sur l'ensemble des communes françaises.

3^{re} partie : le format JSON

Autre format de données très courant sur le « web », le JSON (JavaScript Object Notation). Le JSON fonctionne avec un système de paire clé/valeur. Un « objet » est encadré par des accolades :

```
1 {cle_1 : val_1, cle_2 : val_2, cle_3 : val_3}
```

souvent, pour une question de lisibilité, on écrira :

```
1 {
2   cle_1 : val_1,
3   cle_2 : val_2,
4   cle_3 : val_3
5 }
```

Un fichier au format JSON peut regrouper un grand nombre d'objets :

```
1 [{
2   "nom" : "Ainhoa",
3   "prenom" : "Alan",
4   "date_naissance" : "23/05/2005"
5 },
6 {
7   "nom" : "Otega",
8   "prenom" : "Loïc",
9   "date_naissance" : "15/12/2007"
10 },
11 {
12   "nom" : "Barsacq",
13   "prenom" : "Lorena",
14   "date_naissance" : "12/06/2008"
15 }
16 ]
```