

## II. Les données structurées

Pour être exploitables, les données brutes que nous produisons sont stockées dans des fichiers de manière structurée. Plusieurs formats de fichiers peuvent être utilisés en fonction des besoins.

### 1) Structurer les données

Pour retrouver et traiter des données facilement, on les organise sous la forme d'une table

1. Qu'est-ce qu'une collection ?

*C'est un ensemble d'objets partageant les mêmes descripteurs.*

On donne une collection ci-dessous.

Nom	Capitale	Hymne	Superficie (km <sup>2</sup> )
France	Paris	La Marseillaise	632 734
Chine	Pékin	La Marche des volontaires	9 596 961
Etats-Unis	Washington	<i>The Star-Spangled Banner</i>	9 833 517
Argentine	Buenos Aires	<i>Himno Nacional Argentino</i>	2 791 810

2. Citer un descripteur de cette collection.

Par exemple, le descripteur « Capitale ».

3. Donner une valeur de ce descripteur.

La valeur Paris.

4. Entourer un objet de cette collection.

### 2) Formats des données structurées

Pour assurer leur persistance et leurs échanges, les données sont stockées sous différents formats sous forme de fichier. Un format de données est un mode d'organisation des données qui les rend lisibles, faciles à mettre à jour, échangeables et pérennes.

Pour représenter des données brutes, les formats CSV, JSON et XML sont très utilisés. Le format CSV (*Comma-separated values*) se présente sous forme de table. Les valeurs des descripteurs sont séparées par des virgules. Le format JSON (*JavaScript Object Notation*) associe les données avec une étiquette (descripteur) sous forme d'une liste. Le format XML (*Extensible Markup Language*) utilise des balises pour organiser les informations en sous-éléments.

#### Données au format CSV

```
1 prenom,nom,classe
2 Antoine,Ledoux,seconde
3 Pauline,Darcis,premiere
```

#### Données au format JSON

```
1 {
2   {
3     "prenom": "Antoine",
4     "nom": "Ledoux",
5     "classe": "seconde"
6   }
7   {
8     "prenom": "Pauline",
9     "nom": "Darcis",
10    "classe": "premiere"
11  }
12 }
```

#### Données au format XML

```
1 <eleve>
2   <prenom>Antoine</prenom>
3   <nom>Ledoux</nom>
4   <classe>seconde</classe>
5 </eleve>
6 <eleve>
7   <prenom>Pauline</prenom>
8   <nom>Darcis</nom>
9   <classe>premiere</classe>
10 </eleve>
```

1. Quels sont les avantages des fichiers au format CSV par rapport au JSON et XML ?

*Ils sont plus simples à écrire et sont plus légers.*

2. Donner un inconvénient du format CSV par rapport aux formats JSON et XML.

*Le format CSV est moins souple car tous les descripteurs doivent être renseignés. De plus, les formats XML et JSON sont plus lisibles et permettent d'avoir une hiérarchie et une structure des données beaucoup plus complexes.*

3. Combien y a-t-il d'objets dans la collection représentée sous trois formes différentes par les fichiers au format CSV, JSON ou XML ?

*Il y a deux objets.*

4. Quels en sont les descripteurs ?

*Les descripteurs sont le nom, le prénom et la classe.*

### 3) Un exemple d'utilisation de données ouvertes

Les données ouvertes, ou *Open Data*, sont des données entièrement publiques et libres de droits.

1. Ouvrir un navigateur et se connecter au site <https://data.education.gouv.fr>.

2. Cliquer sur l'onglet « Données ». Dans la partie « Filtres », saisissez « Réussite au baccalauréat selon l'âge » puis cliquer sur le résultat obtenu pour accéder à la page.

3. A quoi servent les différents onglets situés sous le titre de la base de données « Réussite au baccalauréat selon l'âge » ?

Informations : Donne des informations sur le fichier de données

Tableau : Affiche les données sous forme d'un tableau dont chaque colonne correspond à un descripteur

Analyse : Permet d'afficher des graphiques permettant de visualiser les données

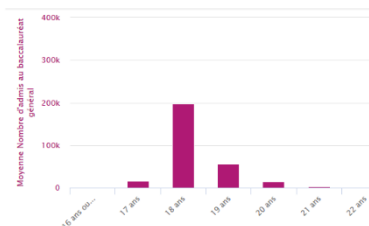
Export : Permet de télécharger les données

API : Permet la recherche et le téléchargement d'enregistrements par plusieurs critères

4. Cliquer sur l'onglet « Analyse ». Dans le menu situé à côté de « Axe Y », sélectionner « Moyenne » et dans le menu encore à droite, sélectionner « Nombre d'admis au baccalauréat général ».

5. Que peut-on conclure sur l'âge des bacheliers à partir de la représentation ?

*On constate que la plupart des bacheliers ont 18 ans.*



6. Sur la gauche de la fenêtre vous avez la possibilité de filtrer les données. Sélectionner les filtres « 1997 » et « 21 ans ». Combien de personnes de 21 ans ont-elles eu leur baccalauréat en 1997 ?

6991

7. Combien de personnes de plus de 30 ans ont eu leur baccalauréat en l'an 2000 ?

45

L'interface du site reste assez limitée lorsqu'il s'agit d'étudier ces données. Pour pouvoir les étudier plus en détail à l'aide de logiciel, on peut télécharger ces données afin de les étudier avec des logiciels plus adaptés.

8. Cliquer sur l'onglet « Export » puis sélectionner « jeu de données entier » dans la ligne CSV.