

Séance 2 : Structurer des données

Contenus :

- Définir une donnée personnelle
- Identifier les principaux formats et représentations de données
- Identifier les différents descripteurs d'un objet
- Distinguer la valeur d'une donnée de son descripteur

Activité 1 Définitions, vocabulaire et exemples

Qu'est-ce qu'une donnée ?

Qu'est-ce qu'une donnée personnelle ?

Descripteurs et valeurs (rappels séance 1)

Vous avez travaillé dans la séance 1 sur des femmes scientifiques célèbres.

Les données connues sur ces femmes sont leur prénom, leur nom et leur date de naissance. On appelle cela les

descripteurs du jeu de données : ils décrivent les données dont on dispose en définissant leur contenu. Chaque descripteur possède des **valeurs** : par exemple, le descripteur « nom » a pour valeurs Johnson, Lovelace et Hamilton. Un regroupement d'objets (ou personnes, etc.) partageant les mêmes descripteurs s'appelle une **collection** de données. Le tableau ci-dessus présente la collection sur les femmes scientifiques célèbres.

prénom	nom	date_naissance
Katherine	Johnson	26/08/1918
Ada	Lovelace	10/12/1815
Margaret	Hamilton	17/08/1936

Activité 2 Les vidéos YouTube

Dans cette activité, on considère un **objet** particulier : une vidéo YouTube. Suite à une recherche, on obtient des propositions de vidéo en lien avec la recherche, comme sur la capture d'écran plus bas.

1. Vous devez constater que les données affichées ont la même forme quelle que soit la vidéo. Identifiez tous les **descripteurs** des vidéos YouTube apparaissant à l'écran.
2. Ces descripteurs permettent de structurer les données sur les vidéos YouTube. La façon la plus simple de représenter des données structurées est d'utiliser un tableau que l'on appelle **table de données**.
Proposez un tableau permettant de représenter les données des vidéos YouTube.

angèle flou

FILTRER

Angèle - Flou [CLIP OFFICIEL]
Angèle · 16 M vues · il y a 5 mois
"Brol la suite", inclus 7 titres inédits, écoutez-les maintenant : <https://lnk.to/BrolLaSuiteYD> Brol Shop : <https://lnk.to/AngeleShopYD> ...

Angèle - Flou (LYRICS/paroles)
lyriczzz vidzzz · 1,5 M vues · il y a 1 an
Si ça s'trouve ça va pas durer
Ou peut-être seulement quelques mois
Voilà je commence déjà àangoisser
Crise après crise j'arrive plus à être moi

Angèle - Oui ou Non [CLIP OFFICIEL]
Angèle · 28 M vues · il y a 3 mois
"Brol la suite", inclus 7 titres inédits, écoutez-les maintenant : <https://lnk.to/BrolLaSuiteYD> Instagram ...
Sous-titres

Angèle - Medley / Victoires de la Musique 2020 (Balance ton quoi, Oui ou Non, Flou...)
Guewen · 214 k vues · il y a 2 semaines
Angèle Medley / Victoires 2020.

Activité 3 Les élèves du lycée

1. Proposez des *descripteurs* permettant de décrire les élèves du lycée.
2. Construisez la *table de données* correspondante et la compléter avec les *valeurs* concernant les membres du binôme/trinôme.
3. Parmi les données que vous avez renseignées, entourez toutes les *données personnelles*.
4. Selon vous, quel est l'intérêt (pour le lycée) de structurer ces données ?

Activité 4 Deux formats de données structurées

Pour mémoriser les tables de manière permanente dans un ordinateur, il faut les stocker dans des fichiers. L'activité propose de travailler sur deux formats très utilisés : le format CSV (*Comma-Separated Values*, déjà découvert dans la séance 1) et le format JSON (*JavaScript Object Notation*).

Voici un exemple de table de données et sa mémorisation au format CSV puis au format JSON.

Table de données

nom	prénom	classe
Dubois	Gabriel	2C
Marchand	Elodie	1G1

Données au format CSV

```
nom,prénom,classe
Dubois,Gabriel,2C
Marchand,Elodie,1G1
```

Données au format JSON

```
[
  {
    "nom": "Dubois",
    "prénom": "Gabriel",
    "classe": "2C"
  },
  {
    "nom": "Marchand",
    "prénom": "Elodie",
    "classe": "1G1"
  }
]
```

1. Quels sont les *objets* de cette table ? Quels sont les *descripteurs* ?
2. Dans le format CSV, quel est le caractère permettant de séparer les données de chaque colonne ?
3. Ecrivez le contenu des fichiers CSV et JSON correspondant à la table ci-dessous.

Date	Equipe 1	Equipe 2	Score
10/02/2020	Angers	Paris	7-0
02/03/2020	Nantes	Angers	2-8

4. Construisez la table correspondant au fichier CSV suivant.

```
nom_commune,code_commune,pop2007,pop2012
BRIOLLAY,49048,2565,2751
VILLEVEQUE,49377,2738,2858
ECOURLANT,49129,3747,3775
```

VOCABULAIRE

Format CSV (*Comma Separated Values*, soit « valeurs séparées par des virgules »)

Un fichier CSV est un fichier texte représentant une table de données structurées. Les valeurs sont séparées par des virgules. Certains fichiers CSV utilisent d'autres symboles comme caractère de séparation, notamment le « ; » pour la plupart des données françaises enregistrées au format CSV. On peut ouvrir un fichier CSV avec un tableur pour le visualiser sous la forme d'un tableau.

Format JSON (JavaScript Object Notation)

Un fichier CSV est aussi un format de données textuelles. Ce format permet de représenter des données plus complexes que des tables. C'est un format très utilisé sur le Web pour récupérer et échanger des données

CONCLUSION

Complétez les pointillés par les mots suivants : donnée, structurées, CSV, collection, table, descripteurs, ligne, colonne

- Une est une valeur décrivant un objet, une personne, un événement digne d'intérêt pour celui qui choisit de la conserver.
- Plusieurs peuvent être utiles pour décrire un même objet (par exemple des descripteurs permettant de caractériser un contact : nom, prénom, adresse et numéro de téléphone).
- Une regroupe des objets partageant les mêmes descripteurs (par exemple, la collection des contacts d'un carnet d'adresses).
- La structure de permet de présenter une collection : les objets en, les descripteurs en et les données à l'intersection. Les données sont alors dites
- Pour assurer la persistance des données, ces dernières sont stockées dans des fichiers. Le format (Comma Separated Values, les données avec des séparateurs) est un format de fichier simple permettant d'enregistrer une table. Il en existe d'autres, comme le format JSON.
- Structurer des données permet d'effectuer des opérations très rapidement sur ces données (rechercher, filtrer, trier, calculer) → cf. séances suivantes

EXERCICES SUPPLÉMENTAIRES

Exercice 1

Choisir la (ou les) bonne(s) réponse(s).

- Une donnée personnelle peut être
 - Un numéro de sécurité sociale
 - Une adresse postale
 - Numéro de train
- Un fichier CSV peut ressembler à :
 - Prénom,Âge
Léa,17
Marc,24
 - ```
[
 {
 "prénom": " Claire ",
 "âge": "16",
 },
 {
 "prénom": " Pierre ",
 "âge": "34",
 }
]
```
  - Prénom,Sexe  
Léa:F  
Marc:M

On considère le fichier CSV ci-contre.

- Dans ce fichier, « Poire » est :
  - Un objet
  - Un descripteur
  - Une valeur d'un descripteur
- Dans ce fichier, il y a :
  - 2 descripteurs
  - 4 objets
  - 15 données

|   | Code | Fruit      | Quantité |
|---|------|------------|----------|
| 1 | 1    | Banane     | 38       |
| 2 | 2    | Poire      | 16       |
| 3 | 3    | Clémentine | 52       |
| 4 | 4    | Kiwi       | 31       |

### Exercice 2

On a enregistré les données d'une table concernant des pays au format CSV et au format JSON.

#### Au format CSV

```
Pays;Population;Superficie
Allemagne;81197537;357,3
Danemark;5659715;42,9
France;66415161;632,8
```

#### Au format JSON

```
[
 {
 "Pays": "Allemagne",
 "Population": 81197537,
 "Superficie": 357.3
 },
 {
 "Pays": "Danemark",
 "Population": 5659715,
 "Superficie": 42.9
 },
 {
 "Pays": "France",
 "Population": 66415161,
 "Superficie": 632.8
 }
]
```

- Quel est le caractère de séparation du fichier CSV ? Selon vous, pourquoi la virgule n'a pas été utilisée ?
- Quels sont les descripteurs de la table ?
- Quels sont les différentes valeurs du descripteur Pays ?
- Écrire les fichiers CSV et JSON représentant les données du tableau suivant :

| Article   | Nutri-score |
|-----------|-------------|
| Chips     | D           |
| Yaourt    | A           |
| Soda      | E           |
| Biscottes | C           |