

Structurer des données

■ Définitions

Commençons par quelques définitions !



Qu'est-ce qu'une donnée ? Qu'est-ce qu'une donnée personnelle ?

Une **donnée** est une valeur ou une information décrivant un objet, une personne, un événement, etc.

Une **donnée personnelle** est toute information se rapportant à une personne physique identifiée ou identifiable.

Il est important de noter qu'une personne physique peut être identifiée :

- de manière directe (exemple : nom et prénom)
- de manière indirecte (exemple : par un numéro de téléphone, une plaque d'immatriculation, etc.)



Avez-vous d'autres exemples de données personnelles qui permettent d'identifier indirectement une personne physique ?

Notez que l'identification d'une personne physique peut être réalisée :

- à partir d'une seule donnée
- à partir du croisement d'un ensemble de données (exemple : une femme vivant à telle adresse et membre dans telle association)

■ Le format CSV

Pour structurer des données

Le format CSV est très courant sur le Web. 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
Johnson,Katherine,26/08/1918
Lovelace,Ada,10/12/1815
Hamilton,Margaret,17/08/1936
```

Je pense qu'il est évident pour vous que nous avons ici 3 femmes :

- Katherine Johnson qui est née le 26/08/1918
- Ada Lovelace qui est née le 10/12/1815
- Margaret Hamilton qui est née le 17/08/1936

Dans ce fichier :

- Les éléments sur la première ligne, à savoir `nom`, `prenom` et `date_naissance`, sont appelés des **descripteurs**. Ces descripteurs permettent de décrire les données en définissant leur contenu.
- `Johnson`, `Lovelace` et `Hamilton` sont les **valeurs** du descripteur `nom`.
- Chacune des lignes de ce fichier (hormis la première qui contient les descripteurs) s'appelle un **objet** ou un **enregistrement**. Ce fichier possède donc 3 enregistrements (ou 3 objets).
- Un regroupement d'objets (ou personnes, etc.) partageant les mêmes descripteurs s'appelle une **collection** de données. Le tableau ci-dessus présente donc une collection sur des femmes scientifiques célèbres.

 **Question 1** : Donnez les différentes valeurs du descripteur `date_naissance` du fichier précédent.

 **Question 2** : Choisissez l'une des trois personnalités de ce fichier puis :

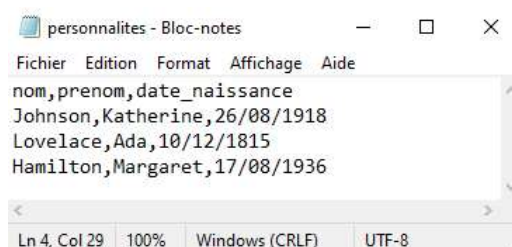
- résumez en une phrase de qui il s'agit ,
- et donnez le lien vers sa page Wikipedia.



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 souvent, en France, 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.

Ouverture d'un fichier CSV

Comme un fichier CSV est un fichier texte, on peut le visualiser avec n'importe quel éditeur de texte (capture d'écran ci-dessous).



Fichier CSV ouvert avec un éditeur de texte

Mais comme un fichier CSV représente des données tabulaires, on peut représenter ces données dans un tableau qui s'appelle une **table de données** :

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

Et à ce titre, on peut l'ouvrir avec n'importe quel tableur (capture d'écran ci-dessous) qui range les données du fichier CSV dans un tableau avec des lignes et des colonnes :

personnalites.csv - LibreOffice Calc

Echier Edition Affichage Insertion Format Feuille

Liberation Sans 10

F10

	A	B	C	D
1	nom	prenom	date_naissance	
2	Johnson	Katherine	26/08/18	
3	Lovelace	Ada	10/12/15	
4	Hamilton	Margaret	17/08/36	
5				
6				

Fichier CSV ouvert avec le tableur Calc de LibreOffice

Question 3 : Téléchargez ce fichier CSV en cliquant sur le lien suivant : personnalites.csv. Puis ouvrez-le de deux manières :

- **avec un éditeur de texte** : clic droit sur le fichier téléchargé puis "Ouvrir avec" et sélectionnez "Bloc-notes"
- **avec un tableur** : clic droit sur le fichier téléchargé puis "Ouvrir avec" et sélectionnez "LibreOffice" puis **sélectionnez la virgule comme séparateur** lors de l'ouverture. Vous devriez obtenir des écrans similaires à ceux donnés au-dessus.

Question 4 : Fermer le tableur puis, dans l'éditeur de texte, ajoutez une ligne (un objet) pour la graphiste [Susan Kare](#) et enregistrez les modifications. Vérifiez en ouvrant le fichier avec le tableur, que la ligne a bien été ajoutée dans le tableau.

■ Exercices

Exercice 1 : Télécharger un fichier CSV de données ouvertes



Source : <https://data.angers.fr/pages/home/>

La ville d'Angers met à disposition des citoyens les données en temps réel sur les disponibilités dans les parkings angevins (il y a plein d'autres jeux de données).

Ces données sont accessibles via cette page <https://data.angers.fr/explore/dataset/parking-angers/information/>.

Question 1 : Télécharger ces données au format CSV en allant dans l'onglet *Export* de cette page.

Question 2 : Ouvrez le fichier téléchargé avec un éditeur de texte et répondez aux questions suivantes :

1. Quel est le *caractère de séparation* utilisé dans ce jeu de données ?
2. Quels sont les *descripteurs* de ce jeu de données ? (revoir la définition dans le paragraphe précédent si besoin).
3. Combien d'*enregistrements* possède ce jeu de données ? (revoir la définition dans le paragraphe précédent si besoin).

N'hésitez pas à ouvrir également ce fichier avec un tableur en veillant à sélectionner le bon caractère de séparation !

Exercice 2 : Les vidéos YouTube

Dans cette activité, on considère un objet particulier : une **vidéo YouTube**.

Les vidéos YouTube et toutes leurs caractéristiques sont enregistrées de manière **structurée** sur les serveurs de YouTube (ce n'est pas dans des fichiers CSV mais dans ce qu'on appelle des bases de données, mais le principe est similaire). C'est ce qui

permet de rechercher des vidéos à partir de mots clés et d'afficher les résultats comme sur la capture d'écran ci-dessous.

chatgpt

X

Q

Filtres

À propos de ces résultats



Ce qui se cache derrière le fonctionnement de ChatGPT

529 k vues • il y a 5 jours

ScienceEtonnante

Je vous explique les bases de ChatGPT, et comment il a été conçu à partir du modèle de langage GPT. Le serveur Discord de ...

Nouveau

Sous-titres



Comment utiliser chatGPT (TUTO Complet)

250 k vues • il y a 2 mois

Labo Des Réseaux

Dans ce tuto, on va voir étape par étape comment utiliser chatGPT. Je te donnerai également les astuces pour affiner tes prompts ...



Chat GPT-4 : la révolution de l'intelligence artificielle

78 k vues • il y a 4 semaines

TF1 INFO

L'outil conversationnel ChatGPT répondait à la plupart de nos questions en quelques instants. Une nouvelle version de cette ...



Pourquoi ChatGPT est une révolution (qui inquiète)

376 k vues • il y a 2 mois

HugoDécrypte

Tout le monde en parle depuis des semaines : dans cette vidéo, je t'explique ce que permet ChatGPT, pourquoi certains ...

4K

Intro | Capacités de ChatGPT | Qui est derrière ChatGPT ? | Comment ça fonctionne ? | Quelles... 8 chapitres

Capture d'écran des premiers résultats YouTube pour le mot clé *chatgpt*.

-  **Question 1** : Vous devez constater que les données affichées en résultat ont la même forme quelle que soit la vidéo. Identifiez tous les descripteurs des vidéos YouTube qui apparaissent à l'écran.
-  **Question 2** : Construisez une table de données avec les descripteurs sur la première ligne et leurs valeurs pour l'une des 4 vidéos (celle de votre choix).

 Exercice 3 :

Dans cet exercice on considère la table de données suivante :

nom	prénom	classe
Dubois	Gabriel	2C
Marchand	Élodie	1G1

Pour mémoriser cette table de manière permanente dans un ordinateur, il faut la stocker dans un fichier. L'activité propose de travailler sur deux formats très utilisés : le format CSV (abordé plus haut) et le format JSON (JavaScript Object Notation). Voici la table écrite dans deux fichiers au format CSV puis au format JSON :

Données au format CSV

nom, prénom, classe
Dubois, Gabriel, 2C
Marchand, Élodie, 1G1

Données au format JSON

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



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

Question 1 : Quels sont les objets de cette table ? Quels sont les descripteurs ?

Question 2 : Dans le format CSV, quel est le caractère permettant de séparer les données de chaque colonne ?

Question 3 : Écrivez le contenu des fichiers CSV et JSON correspondant à la table ci-dessous :

Date	Equipe 1	Equipe 2	Score
10/02/2022	Angers	Paris	7-0
02/03/2020	Nantes	Angers	2-8

Question 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

■ Bilan

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), ce qui sera vu par la suite.
-

Références :

- Cours de David Roche sur les [différents types de données structurées](#)
- CNIL pour la définition d'une donnée personnelle : <https://www.cnil.fr/fr/definition/donnee-personnelle>

Voir en ligne : info-mounier.fr/snt/donnees/donnees_structurees.php