

Thème 3 : Les Données Structurées

Activité 1 - Les données ouvertes, personnelles, structurées.

À partir du livre p. 62-63.

Une correction possible en italique.

Question 1. Les données ouvertes (Open Data) et les données personnelles.

a) Consulter l'**aide 1**. En quoi consiste l'Open Data ?

C'est une démarche qui vise à rendre les données numériques accessibles par tous, exemples : données publiques de mairies, de l'état, comptes publics, services publics.

Quelles sont les informations qui ne peuvent pas être rendues publiques ?

Ce sont les informations personnelles.

b) Qu'est-ce que le RGPD ? Qu'est-ce qu'une donnée personnelle ? Visionner l'**aide 2** si besoin.

C'est le règlement général pour la protection des données mise en place par l'union européenne. Ça sert à encadrer la manière dont les entreprises, les administrations qui gèrent les données personnelles vont utiliser ces données personnelles.

Une donnée personnelle c'est toute information qui permet d'identifier directement ou indirectement une personne, comme son nom, son adresse, son numéro de téléphone, son adresse email, son activité sur internet, les contenus postés sur internet.

Reprend le droit français.

Consentement des mineurs au dessus 15 ans, en dessous le consentement est demandé au parents.

Droit à la portabilité des données : je peux prendre toutes mes données d'un service (email par exemple) et les donner à un autre service.

Conditions d'utilisation par les entreprises : des règles encadrant l'utilisation des données personnelles claires et concises ! Pas encore respectées !

c) Avec l'**aide 3**, identifier les dispositions qui permettent de :

- donner son accord ou refuser la collecte de données privées ;

Le consentement "explicite" et "positif".

Le profilage (article 22).

- contrôler ce qui est fait avec les données collectées ;

le consentement .

les missions du délégué à la protection des données.

responsabiliser les entreprises.

- se faire oublier.

Le droit à l'effacement (version allégée du droit à l'oubli) (article 17)

Question 2. Les métadonnées.

a) Visionner l'**aide 4**. Que sont les métadonnées d'un fichier ?

Métadonnées : éléments qui servent à décrire une ressource (un livre ou une bibliothèque) tels que le titre, l'auteur, la date de sortie, la maison d'édition, etc.

À quoi peuvent-elles servir ?

Avec l'exemple précédent si on fait une recherche sur le web, elles facilitent la recherche de la ressource, ici le livre.

b) Donner un exemple de format de fichier de données structurées.

Format csv : Coma Separated Values (valeurs séparées par des virgules).

c) À partir du site www.data.gouv.fr, télécharger les données ouvertes sur la playlist de la radio NRJ pour la période du 13 juillet 2012 au 13 août 2012 (vous pouvez vous aider de la barre de recherche en haut de la page du site).

d) Un clic droit sur le fichier permet de consulter ses propriétés, également appelées métadonnées. Quel est le format de ce fichier de données ?

Format csv.

Quelles sont les autres informations disponibles ?

Le nom, la taille, la création, la modification du fichier, le programme d'ouverture.

Quelles sont les informations modifiables ?

Le programme d'ouverture, le nom du fichier.

Question 3. Manipuler les données.

a) Ouvrir le fichier précédent avec un éditeur de texte type Notepad ou Notepad++ : les données s'affichent au format texte où chaque ligne contient les données d'un objet délimitées par ce qu'on appelle un "séparateur".

Combien de données servent à décrire chaque objet de cette collection de données ? 7
Quel est le séparateur utilisé pour délimiter les données ? *Le point-virgule, ";"*

b) Quelles opérations de traitement (filtre, tri, calcul...) sont possibles avec cette représentation de données ?

La recherche (date, heure, nom), le comptage de cette recherche.

c) Ouvrir un tableur et importer les données. **Aide 5** si nécessaire.

d) Les descripteurs des objets de cette collection sont renseignés par le site www.data.gouv.fr : Date;Heure;Artiste;titre;Année de la chanson;Sexe;Pays d'origine. Ajouter une première ligne au tableau et nommer chaque colonne par le descripteur associé.

e) Avec l'aide de l'outil "filtrer" appliqué à la première ligne et de l'outil "trier" appliqué au bon descripteur, déterminer l'année la plus récente et l'année la plus ancienne des titres de cette playlist.

*L'année la plus ancienne : 1979,
l'année la plus récente : 2012*

f) Un autre format de données courant sur le Web est le format json. Faire quelques recherches sur ce format pour identifier les différences principales avec le format csv.

Format json : JavaScript Object Notation.

Différences avec le format csv :

csv, une ligne d'entête ; json, paires clé-valeur

csv : un des formats les plus simples à utiliser

Vous voudrez utiliser ce format CSV quand :

Vos données sont simples et facile à représentées sous forme de tableau.

Vos données sont statiques (pas de vérification, pas de calculs). Bref, on les lis tel quel.

Vous voulez que ce soit lisible dans un tableur.

exemple de forme de fichier au format json :

```
[
  {
    'nom': 'Sam',
    'numero': '555-555-555',
    'adresse': '5, rue des lilas'
  },
  {
    'nom': 'Max',
    'numero': '1234567890',
    'adresse': 'Patong, thailande'
  },
  {
    'nom': 'Bob',
    'numero': '666',
    'adresse': '7eme cercle'
  }
]
```

le même fichier au format csv :

```
"nom";"numero";"adresse"
"Sam";"555-555-555";"5, rue des lilas"
"Max";"1234567890";"Patong, thailande"
"Bob";"666";"7eme cercle"
```