

Activité 2 sur le thème 1 : format des données et open data

Consignes :

1. Enregistrer immédiatement le fichier word dans le dossier approprié (snt/theme1).
2. Compléter ce document Word . Il faudra soigner la rédaction (**phrases complètes**). A la fin de l'heure , enregistrer votre travail sur clé usb et déposer le document dans l'espace élève de Pronote. (page d'accueil ou cahier de texte – onglet déposer ma copie)

I-Format de données

- Une donnée structurée est organisée sous forme d'un tableau appelé table de données.
- La première ligne décrit la forme des lignes suivantes ; les éléments de cette ligne donnant la signification portent plusieurs noms possibles : ce sont les noms de champ ou **descripteurs** ou critères.
- Les lignes suivantes sont appelées des **objets** : elles sont la liste des valeurs de chacun des descripteurs.

Voici ci-dessous une table de données :

Descripteurs	Prénom	Nom	Âge
Une valeur du descripteur « Prénom »	Antoine	Ledoux	15
	Pauline	Darcis	16
	Karim	Chanhoun	15
Un objet	Sarah	Goldberg	14

Pour pouvoir stocker, transmettre et exploiter ce type de tableau de valeurs, il peut être représenté sous différentes formes appelées format.

Les principaux formats utilisés pour représenter un ensemble de données sont le CSV , le JSON et le XML.

- L'acronyme CSV signifie "Comma Separated Values" c'est-à-dire "valeurs séparées par une virgule".
- L'acronyme JSON signifie "JavaScript Object Notation" c'est-à-dire "notation d'objets en Javascript".
- L'acronyme XML signifie "eXtensible Markup Language " c'est-à-dire "langage de balisage extensible".

En CSV, les données sont stockées dans un fichier texte :

- où la première ligne donne la liste des descripteurs, ,
- où chaque autre ligne du tableau correspond à un objet,
- où les valeurs (d'un objet) sont séparées par une virgule.

prenom	nom	age	classe
Jean	Céplusse	16	Seconde 3
Leïla	Mina	14	Seconde 2
Arjuna	Mahatman	15	Seconde 3
Mathilda	Vantage	15	Seconde 3

Voici comment sont écrites les données du tableau ci-dessus au format CSV :

```

1 prenom,nom,age,classe
2 Jean,Céplusse,16,Seconde 3
3 Leïla,Mina,14,Seconde 2
4 Arjuna,Mahatman,15,Seconde 3
5 Mathilda,Vantage,15,Seconde 3

```

Au format JSON :

Au format :XML

```

1 <eleve>
2   <prenom> Jean </prenom>
3   <nom> Céplusse </nom>
4   <age> 16 </age>
5   <classe> Seconde 3 <classe>
6 </eleve>
7 <eleve>
8   <prenom> Leïla </prenom>
9   <nom> Mina </nom>
10  <age> 14 </age>
11  <classe> Seconde 2 <classe>
12 </eleve>
13 <eleve>
14   <prenom> Arjuna </prenom>
15   <nom> Mahatman </nom>
16   <age> 15 </age>
17   <classe> Seconde 3 <classe>
18 </eleve>
19 <eleve>
20   <prenom> Mathilda </prenom>
21   <nom> Vantage </nom>
22   <age> 15 </age>
23   <classe> Seconde 3 <classe>
24 </eleve>

```

```

1  {
2      {
3          "prenom": "Jean",
4          "nom": "Céplusse",
5          "age": "16",
6          "classe": "Seconde 3"
7      }
8      {
9          "prenom": "Leïla",
10         "nom": "Mina",
11         "age": "14",
12         "classe": "Seconde 2"
13     }
14     {
15         "prenom": "Arjuna",
16         "nom": "Mahatman",
17         "age": "15",
18         "classe": "Seconde 3"
19     }
20     {
21         "prenom": "Mathilda",
22         "nom": "Vantage",
23         "age": "15",
24         "classe": "Seconde 3"
25     }
26 }

```

Le format XML avec son système de balises est plus long à écrire est plus rapide à traiter par la machine.

En SNT, nous n'utiliserons que le format CSV.

- Attention ! La virgule est le séparateur standard pour les données anglo-saxonnes, mais pas pour les données aux normes françaises puisque, en français, la virgule est le séparateur des chiffres décimaux (= "nombres à virgule"). C'est pourquoi on utilise un autre séparateur en français : le point-virgule (;)
- Dans certains cas, le fait d'avoir comme séparateur une virgule ou un point-virgule peut engendrer quelques problèmes. Vous devrez donc rester vigilants sur le type de séparateur utilisé.
Un tableur tel que "Calc" (de Libre Office) est capable de lire les fichiers au format CSV mais d'autres tableurs (comme Excel) gèrent moins bien le séparateur CSV "point-virgule".

CAPACITÉS : Identifier les principaux formats et représentations de données. Identifier les différents descripteurs d'un objet.
Distinguer la valeur d'une donnée de son descripteur.

On a enregistré les données d'un répertoire au format CSV et au format JSON.

Répertoire au format CSV

```
"nom";"ville";"adresse"  
'Pierre';'Ivry';'5, rue du fort'  
'Pauline';'Toulon';'18, rue de la poste'  
'Kilian';'Vesoul';'2, avenue des oeillets'
```

1. Quels sont les descripteurs du répertoire ?
2. Quelles sont les différentes valeurs du descripteur **nom** ?
3. Écrire les fichiers CSV et JSON représentant les données du tableau suivant :

Aliment	Nutrinote	Calories/100g
Gaufre	C	291
Yaourt	A	65
Muesli	A	175
Pain de mie	B	234

Répertoire au format JSON

```
{  
  {  
    'nom': 'Pierre',  
    'ville': 'Ivry',  
    'adresse': '5, rue du fort'  
  },  
  {  
    'nom': 'Pauline',  
    'ville': 'Toulon',  
    'adresse': '18, rue de la poste'  
  },  
  {  
    'nom': 'Kilian',  
    'ville': 'Vesoul',  
    'adresse': '2, avenue des oeillets'  
  }  
}
```



1. Quels sont les descripteurs du répertoire téléphonique ?

On distingue trois descripteurs : nom , ville , adresse.

2. Quelles sont les différentes valeurs du descripteur nom ?

Les différentes valeurs du descripteur nom sont Pierre, Pauline, Kilian.

3. Écrire les fichiers CSV et JSON représentant les données du tableau :

Fichier CSV :

```
"Aliment";"Nutrinote";"Calories"  
"Gaufre";"C";291  
"Yaourt";"A";65  
"Muesli";"A";175  
"Pain de mie";"B";234
```

Fichier JSON :

```
{
  {
    'Aliment': 'Gaufre',
    'Nutrinode': 'C',
    'Calories': '291'
  },
  {
    'Aliment': 'Yaourt',
    'Nutrinode': 'A',
    'Calories': '65'
  },
  {
```

II-Open data

L'open data correspond à une volonté politique d'ouvrir les données jugées d'intérêt public afin d'encourager leur réutilisation par tout un chacun. Cela permet d'encourager la transparence démocratique, de bénéficier de services au quotidien ou de prendre des décisions plus éclairées.

L'État français rend publique les données qu'il a développé dans le cadre de ses missions et qu'il juge non sensibles à l'adresse suivante : <https://www.data.gouv.fr>.

1. Aller sur le site des données publiques de l'État français : [lien direct](#).
2. Dans le moteur de recherche de ce site tapé le mot *secours* afin d'obtenir la dernière table de données sur les "Interventions réalisées par les services d'incendie et de secours".
3. Télécharger le fichier le plus récent (au format CSV).
4. Ouvrir ce fichier avec Excel
5. Enregistrer ce fichier sur votre clé USB (snt/theme 1)

L'utilisation de données ouvertes ou rendues publiques est une aide au journalisme d'investigation.

L'OSINT (Open Source INTelligence) ou Recherche d'informations à partir de données ouvertes est un ensemble de méthodes qui permettent d'exploiter des sources en libre accès depuis Internet.

Des exemples de telles données sont des textes, des vidéos ou des photos postés sur des réseaux sociaux, des images satellites, des positions d'avions ou de bateaux, des données de géolocalisation, ...

Plusieurs collectifs de personnes travaillant dans l'OSINT existent.

[Bellingcat](#) est un site Web où des journalistes d'investigation citoyens effectue de la vérification de faits et des enquêtes en se servant de renseignements obtenus et de données ouvertes.

Ce site s'est fait connaître au travers de plusieurs enquêtes de qualité. En voici quelques unes :

- Le créateur du site démontre que le régime du dictateur syrien Bachar el-Assad utilise des armes chimiques et des armes à sous-munitions contre sa population.
Cette démonstration repose en partie sur l'analyse de vidéos accessibles sur Internet que le journaliste arrive à localiser et qui lui permet d'analyser les détails des armes utilisés.
- En 2016, le site prouve que lors de la guerre du Donbass la Russie, depuis son territoire, a bombardé une partie du territoire ukrainien à l'été 2014 pour soutenir les rebelles pro-russes des régions de Donetsk et Louhansk.
Cette démonstration repose en partie sur l'analyse d'images satellites publiques disponibles permettant d'étudier la trajectoire des projectiles à partir des impacts au sol.
[Voici un accès direct à l'article \(en anglais\).](#)
- Bellingcat démontre aussi en novembre 2014 que l'avion de la Malaysia Airlines qui a été abattu le 17 juillet 2014 au-dessus de l'est de l'Ukraine, avec à son bord 298 personnes, a été abattu par un missile russe et incrimine directement une brigade l'armée russe.
Cette démonstration repose en partie sur l'analyse des photos et de vidéos obtenus sur les réseaux sociaux puis sur des images satellitaires.
[Voici un accès direct à l'article \(en anglais\).](#)
Le rapport d'enquête internationale sur les raisons du crash confirmera ensuite cette affirmation.
- En décembre 2020, le site Bellingcat publie une enquête dans laquelle sont identifiés les deux agents des services secrets russes qui ont empoisonnés l'opposant russe Alexis Navalny et affirme que deux instituts russes poursuivent des recherches sur les substances chimiques mortelles.
Cette démonstration repose en partie sur l'analyse de données obtenues en Russie contenant des relevés téléphoniques, des bases de données en lien avec des numéros de passeports, ...

Penser à sauvegarder votre travail (snt/theme1)