

2nde SNT – Thème #1

Données structurées

Marc Biver, Septembre 2023

Lycée Fustel de Coulanges, Massy

Quelques questions pour commencer...

- Avant de parler de données « structurées » - c'est quoi, une donnée ?
- Un fichier, ça peut contenir quoi ?
- La capacité d'un smartphone pour stocker des fichiers se mesure en... ?
- Vrai ou faux ?
 - L'adresse d'une personne n'est plus une donnée personnelle si elle a été donnée volontairement (sur un site web par exemple) ?
 - Les informations de votre carte Vitale sont stockées uniquement sur la carte elle-même.
 - Le « cloud » est une série de millions de disques durs localisée dans la Silicon Valley, en Californie.

De quoi va-t-on parler dans ce thème ?

- Comment on structure / on organise des données ?
- Un peu d'histoire – comment l'humain enregistre des données, et depuis quand ?
- Comment on traite / on analyse des données ? (tableurs, bases de données)
- Des données un peu particulières – données personnelles, métadonnées.
- Où sont toutes ces données – big data, cloud...

Structuration de données

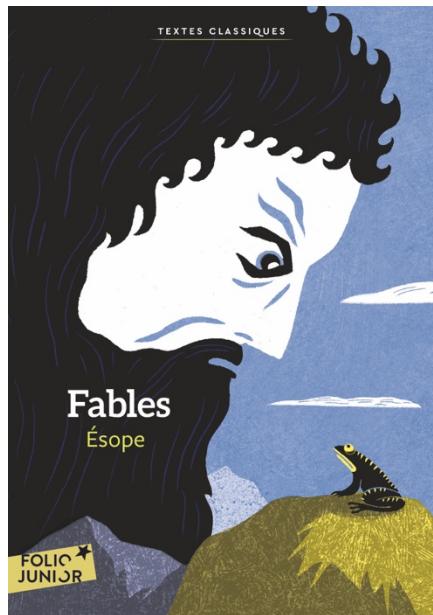
Activité – la bibliothèque (source : Louis Paternault) :

1. Par binôme, choisissez un livre que vous aimez. Décrivez-le en une ligne sur une feuille de papier et passez cette feuille à votre camarade. Est-ce que vous pensez que, dans une bibliothèque infinie, il ou elle pourrait le retrouver ?

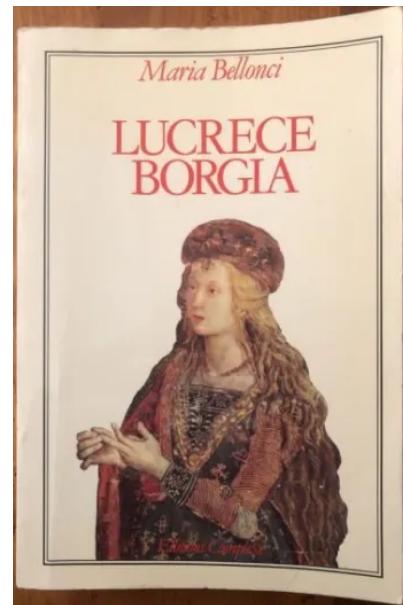


2. Considérez ces notices :

- ▶ **Auteur(s)** : Ésope (0620 ?-0560 ? av. J.-C.).
- ▶ **Titre conventionnel** : Fables (français). Extrait]
- ▶ **Titre(s)** : Fables
- ▶ **Publication** : Paris : Gallimard jeunesse, DL 2021
- ▶ **Impression** : impr. en Espagne
- ▶ **Description matérielle** : 1 vol. (100 p.) : ill. ; 18 cm
- ▶ **Collection** : Folio junior. Textes classiques
- ▶ **Autre(s) auteur(s)** : Saillard, Rémi (1960-....). Illustrateur ; Chambry, Émile (1864-1951). Traducteur ; Cossou, Bruno. Éditeur scientifique.
- ▶ **Identifiants** : ISBN 978-2-07-512050-0 (br.)
- ▶ **Prix** : 4,50 EUR



- ▶ **Auteur(s)** : Bellonci, Maria (1902-1986)
- ▶ **Titre(s)** : Lucrèce Borgia
- ▶ **Traduction de** : Lucrezia Borgia
- ▶ **Publication** : Bruxelles : Éditions Complexe ; [Paris] : [diffusion Presses universitaires de France], 1983
- ▶ **Impression** : impr. en Belgique
- ▶ **Description matérielle** : IV-469 p. : couv. ill. en coul. ; 23 cm
- ▶ **Collection** : Le Temps et les hommes ; 14
- ▶ **Note(s)** : Bibliogr. p. 463-469
- ▶ **Identifiants, prix et caractéristiques** : ISBN 2-87027-106-9 (Br.) : 125 F



- ➔ Comment identifier chaque livre de manière unique ?
- ➔ Quelle(s) donnée(s) vous « parle(nt) » ? Quelle(s) donnée(s) parlerai(en)t plutôt à un ordinateur ?
- ➔ Y a-t-il une donnée qui vous semble « de trop » ?
- ➔ Comment organiseriez-vous ces données si vous étiez en charge de la bibliothèque infinie ?
- Sur du papier ?
 - Sur un ordinateur ?

Éléments de cours :

DOC1

Collection

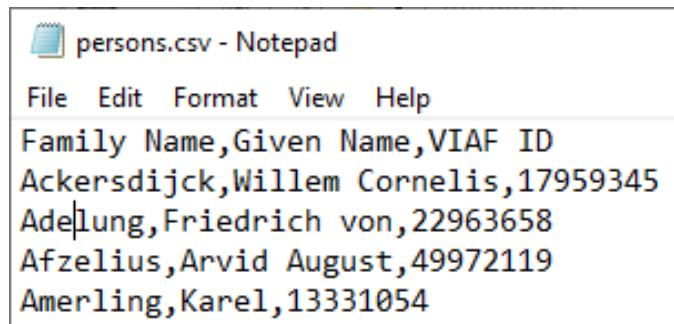
Descripteurs	Nom	Capitale	Hymne	Superficie (km ²)
Une valeur du descripteur « Nom »	France	Paris	La Marseillaise	632 734
	Chine	Pékin	La Marche des volontaires	9 596 961
	États-Unis	Washington	The Star-Spangled Banner	9 833 517
Un objet	Argentine	Buenos Aires	Himno Nacional Argentino	2 791 810

- On organise des données dans des tables, ou des collections.
- Chaque ligne de ces tables est un objet – toutes les données de la ligne concernent la même « chose ».
- Chaque colonne de ces tables est un attribut de ces objets qui porte un nom – c'est le descripteur.
- Chaque case, ou cellule, de ces tables comporte une valeur – c'est un élément unitaire d'information.
- Une donnée, c'est une valeur, un code – en soi elle ne veut rien dire. Sa valeur c'est qu'associée à une multitude d'autres données elle fournit de l'information :
 - Une lettre de l'alphabet – la donnée.
 - Un mot – un objet.
 - Une phrase – de l'information.

Les formats des données :

Des données structurées et organisées peuvent être représentées sous de nombreuses formes, comme par exemple :

- CSV – des tables décrites sous formes de lignes de texte, très facile d'utilisation et la base du fonctionnement des tableurs qu'utilisent directement les humains.



A screenshot of a Windows Notepad window titled "persons.csv - Notepad". The menu bar includes File, Edit, Format, View, and Help. The content of the file is a CSV table:

Family Name	Given Name	VIAF ID
Ackersdijck	Willem Cornelis	17959345
Adelung	Friedrich von	22963658
Afzelius	Arvid August	49972119
Amerling	Karel	13331054

- JSON ou XML – ce sont des formats d'échanges de données destinés à transmettre des informations de machine à machine pour exploitation par des programmes. Les données sont toutes associées à des descripteurs (sous forme d'étiquettes ou de balises).

```
{  
  "menu": {  
    "id": "file",  
    "value": "File",  
    "popup": {  
      "menuitem": [  
        { "value": "New", "onclick": "CreateNewDoc()" },  
        { "value": "Open", "onclick": "OpenDoc()" },  
        { "value": "Close", "onclick": "CloseDoc()" }  
      ]  
    }  
  }  
}
```

```
<menu id="file" value="File">  
  <popup>  
    <menuitem value="New" onclick="CreateNewDoc()"/>  
    <menuitem value="Open" onclick="OpenDoc()"/>  
    <menuitem value="Close" onclick="CloseDoc()"/>  
  </popup>  
</menu>
```

➔ Ces différents formats offrent différentes possibilités de consultation, de traitement et d'analyse des données – et on en verra quelques-unes dans la troisième section de ce chapitre.

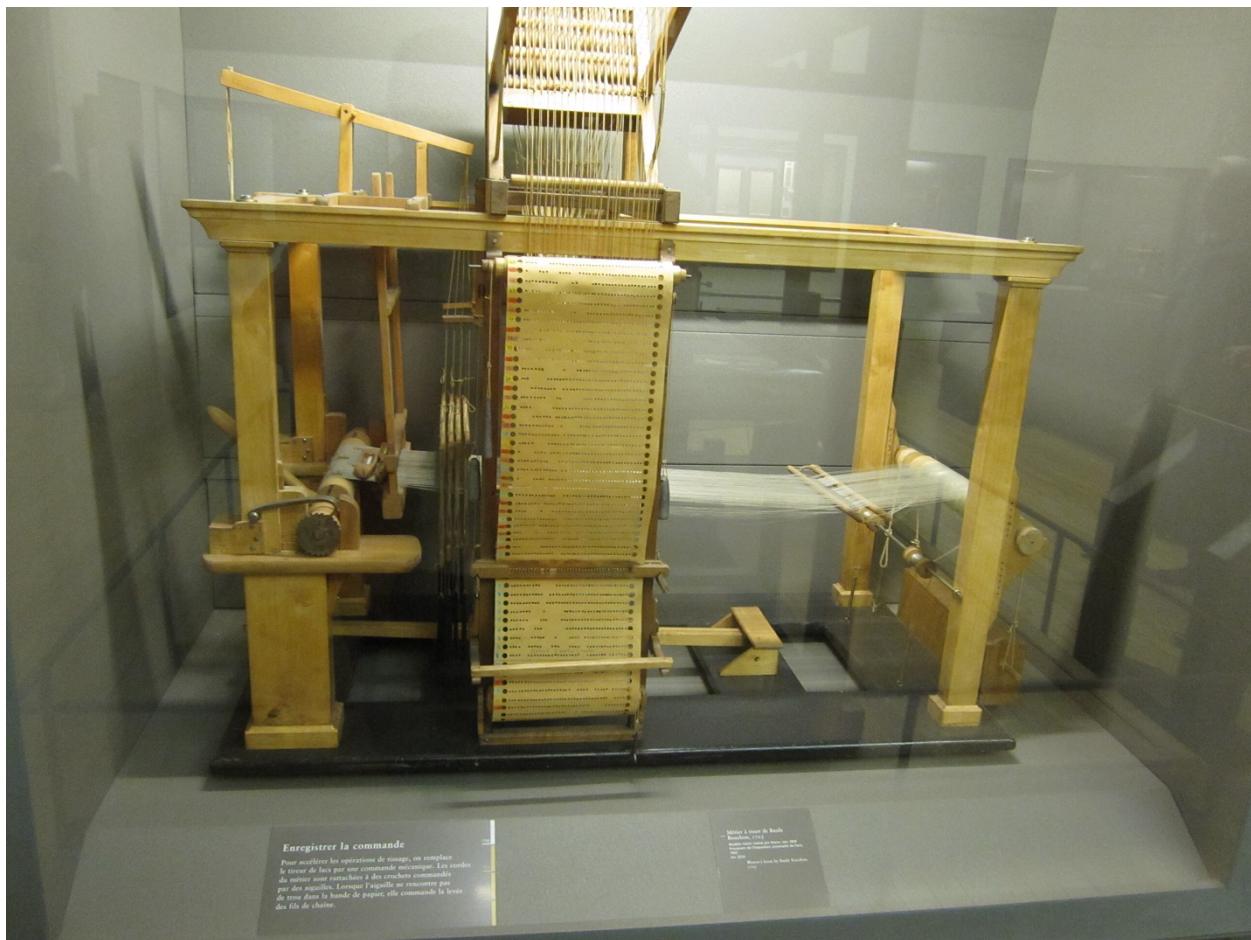
Un peu d'histoire...

- L'humain enregistre sur des *supports externes* des données depuis très longtemps :
 - L'écriture – sans doute apparue pour la première fois en Mésopotamie il y a environ 5.500 ans.
- ➔ *Utilisation de la donnée par d'autres humains.*



- Les rubans perforés inventés il y a 300 ans en France.

➔ *Utilisation de la donnée par un système mécanique.*



- Les cartes perforées inventées à la fin des années 1930 par IBM.

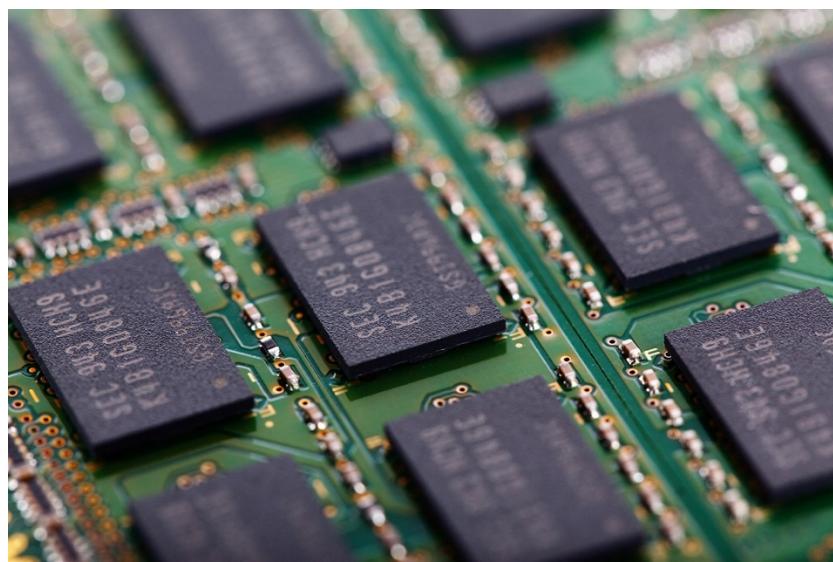
➔ *Première utilisation de la donnée par un ordinateur.*



- Les disques durs inventés dans les années 1950 par IBM – le premier, l'IBM 350 pèse une tonne et peut contenir... 3,75 Mo (soit une chanson MP3 ou 1/35.000^{ème} de téléphone 128 Go).



- La mémoire flash ou SSD inventée par Toshiba en 1984 – plus rapide, sans pièces mécaniques, elle remplace progressivement les disques durs.





© DIRECTION DE L'INFORMATION
LÉGALE ET ADMINISTRATIVE | PARIS 2017

Qu'est-ce qu'une donnée personnelle ?

Toute information relative à un particulier identifié ou identifiable, directement ou indirectement, grâce à un identifiant ou à un ou plusieurs éléments propres à son identité

Par exemple :



Source : Règlement général sur la protection des données (RGPD) du 27 avril 2016

- RGPD : Règlement Général sur la Protection des Données.
- Une donnée personnelle permet d'identifier une personne soit directement (identifiant) soit par recouplement (vous vous souvenez de la bibliothèque infinie ?).

- Origine : un humain, une caméra, un transfert...

Activité – métadonnées :

- ➔ Prise en main des ordinateurs de la salle / vérification des connexions.
- Dans un navigateur (Firefox ou Chrome par exemple) :
 - Rendez-vous à l'adresse suivante :

<https://tinyurl.com/2p9a67yz>

 - Cliquez sur la croix pour ne pas vous connecter ou créer de compte.
 - Téléchargez la photo que vous trouvez.



- Qui peut me dire ce que j'utilise comme téléphone et où j'ai passé des vacances récemment ? Et qu'est-ce que vous pouvez me dire sur mon honnêteté dans ce que je poste... ?
- Quelle leçon tirer de cela ?
- Comment se protéger ? ➔ Sortez vos téléphones...