## 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.

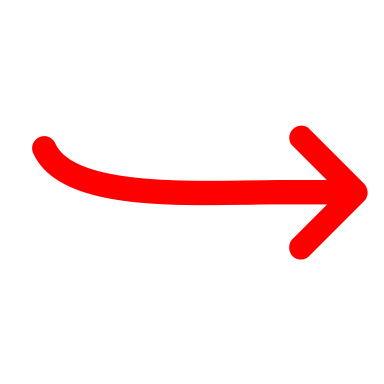
### De quoi va-t-on parler dans ce thème ?

* 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…

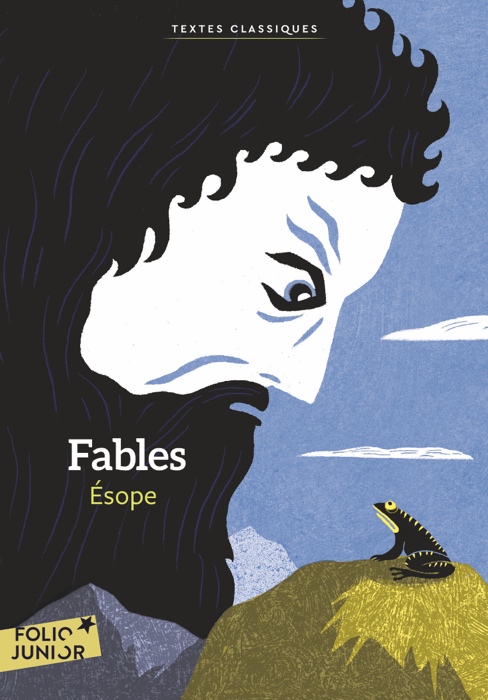
### Structuration de données

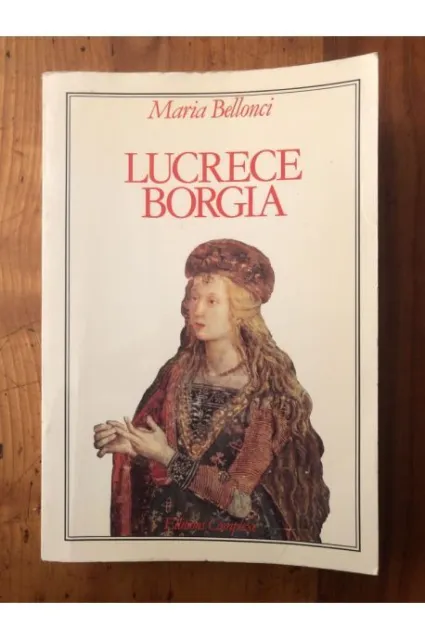
#### Activité – la bibliothèque (*source : Louis Paternault)* :

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 ?

A long shot of a library with Trinity College Library in the background

Description automatically generated

1. Considérez ces notices :

A white paper with black text

Description automatically generated

A white paper with black text

Description automatically generated

🡺 Comment identifier chaque livre de manière unique ?

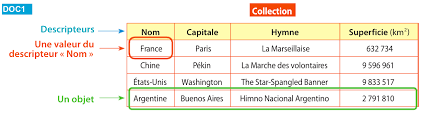
🡺 Quelle(s) donnée(s) vous « parle(nt) » ? Quelle(s) donnée(s) parlerai(en)t plutôt à un ordinateur ?

🡺 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 :

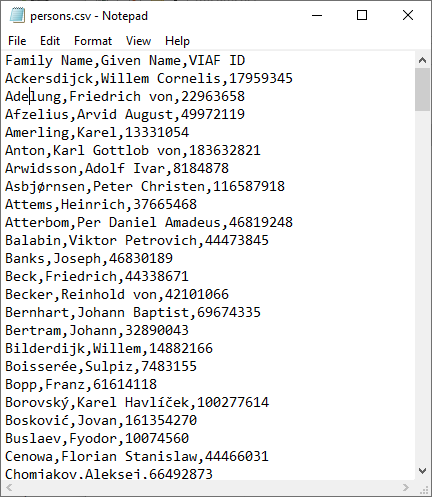


* On organise des données dans des **tables**, ou des **collections**.
* Chaque ligne de ces tables est un **objet** – toutes les données de la ligne concernent la même « chose ».
* Chaque colonne de ces tables est un **attribut** de ces **objets** qui porte un nom – c’est le **descripteur**.
* Chaque 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.



* 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).

A computer code with text

Description automatically generated with medium confidence

A computer screen shot of a computer code

Description automatically generated

**🡺 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.

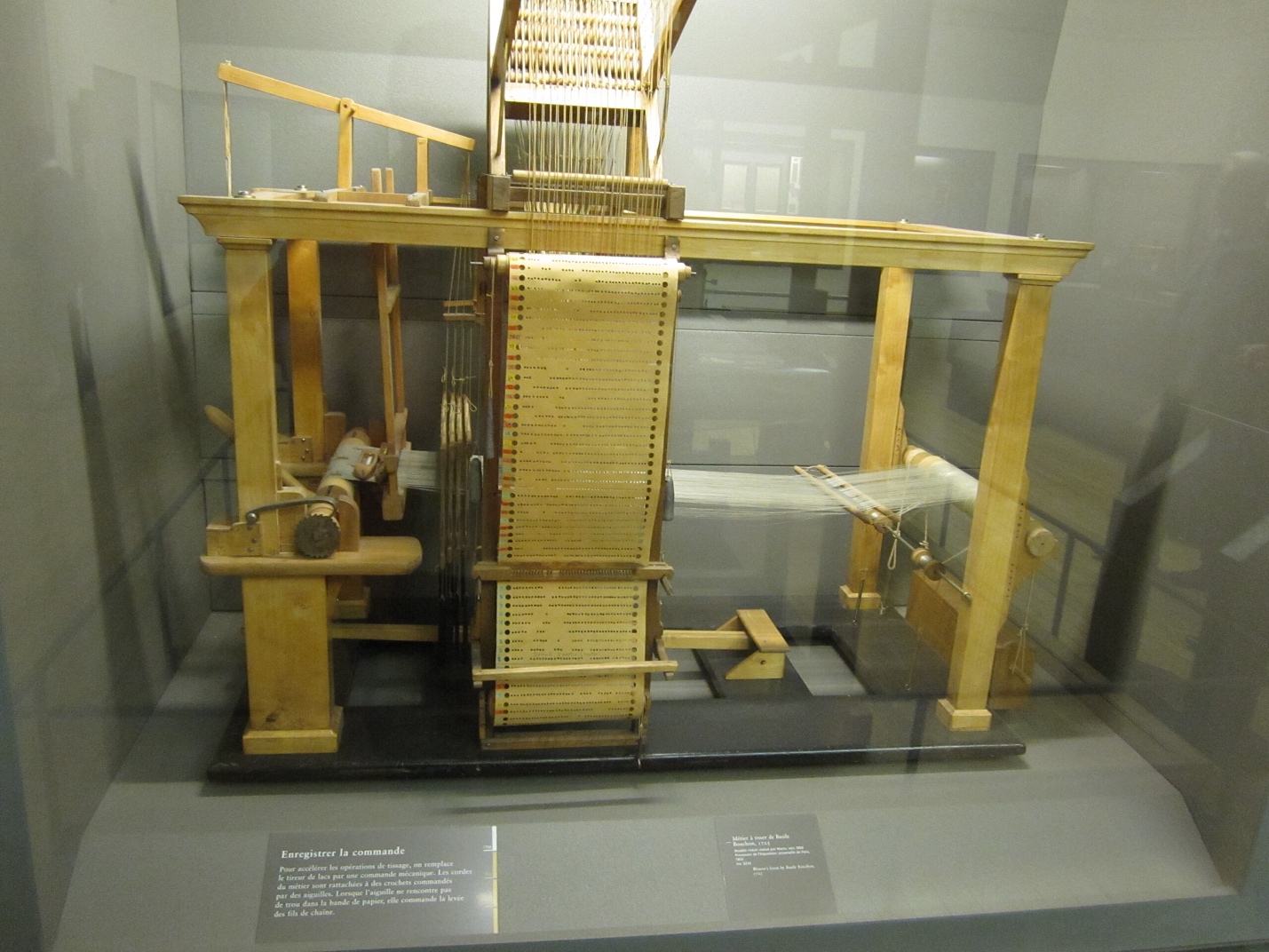
*🡺 Utilisation de la donnée par d’autres humains.*

A stone wall with writing on it

Description automatically generated

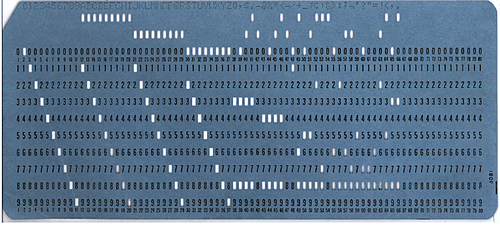
* + Les rubans perforés inventés il y a 300 ans en France.

*🡺 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.*

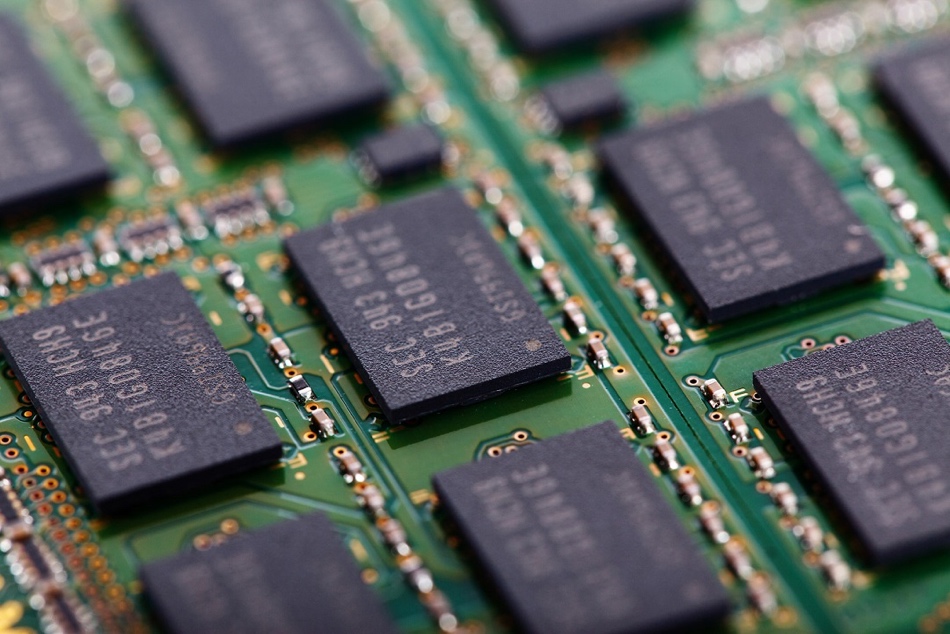


* + Les disques durs inventés dans les années 1950 par IBM – le premier, l’IBM 350 pèse une tonne et peut contenir… 3,75 Mo (soit une chanson MP3 ou 1/35.000ème de téléphone 128 Go).

A person standing next to a large machine

Description automatically generated

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



### Données personnelles & métadonnées

A blue and white cover with white text

Description automatically generated

* RGPD : Règlement Général sur la Protection des Données.
* Une donnée personnelle permet d’identifier une personne soit directement (identifiant) soit par recoupement (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.

A person sitting on a rocky cliff overlooking the ocean

Description automatically generated

* + 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…