


LES DONNÉES STRUCTURÉES ET LEUR TRAITEMENT

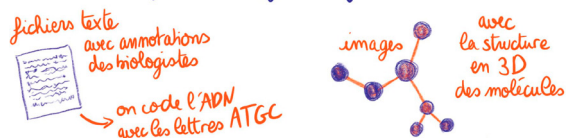
STRUCTURER LES DONNÉES

Pour décrire un lieu, on peut utiliser plusieurs **données** comme des coordonnées GPS, un nom et une brève description. Ceci constitue un enregistrement. Une collection regroupe plusieurs **enregistrements** d'un même type. On parle de données **structurées** quand celles-ci sont organisées et classées en vue de faciliter leur traitement.

Une **collection** regroupe des enregistrements d'un même type (par exemple des lieux). Informatiquement, ils sont regroupés dans un fichier dont le format permet de facilement effectuer des opérations de tri ou de recherche. Une base de données regroupe des collections ayant des liens entre elles.

Pour aller plus loin:  [Leçon inaugurale de Serge Abiteboul au Collège de France en 2012](#)

Les données issues de la biologie moléculaire sont les génomes, les séquences (ADN, protéines, etc.) et peuvent être partagées sous différentes formes :



Le génome humain appartenant à tout le monde, il n'existe pas une seule base de données biologiques mais **PLUS DE 1500!!!**



Toutes ces données sont publiées sur le web, elles sont très nombreuses et volumineuses, on parle de **BIG DATA**.

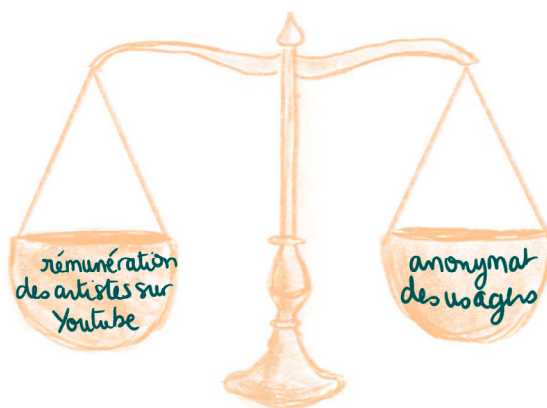
ANGLICISMES DU QUOTIDIEN

Les **big data** désignent l'exploitation de données massives souvent issues d'autres domaines comme les sciences, la santé ou l'économie. Leur traitement soulève des questions sociétales nombreuses tant en termes de démocratie, de surveillance de masse, d'impact énergétique ou encore d'exploitation des données personnelles.

Pour aller plus loin:  [Covid-19: l'indispensable apport de la science des données](#)



On parle d'**open data** pour désigner des données dites ouvertes, c'est-à-dire à la disposition de toutes et tous. Le règlement général sur la protection des données (RGPD) encadre le traitement des données personnelles sur le territoire de l'Union Européenne et le gendarme français pour le RGPD est la CNIL.

Pour aller plus loin:  [L'Open Data, l'ouverture des données pour de nouveaux usages](#)
 [Les données des réseaux sociaux mobilisées contre le Covid-19](#)



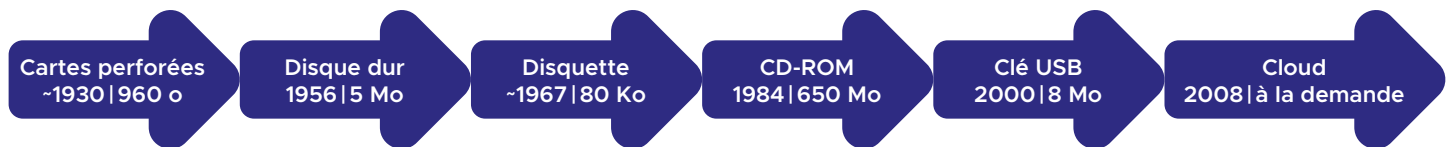
Un **data center** est un entrepôt de stockage de serveurs de calcul et d'instruments de conservation de données. Les data centers soulèvent de nombreux enjeux écologiques: leur fonctionnement est très gourmand en énergie (alimentation électrique mais aussi refroidissement) et la fabrication des serveurs nécessite des métaux rares et génère des déchets.

Le **cloud** désigne un lieu de stockage (dans le "nuage") hébergé sur des serveurs accessibles sur Internet, le **cloud computing** désigne le calcul sur des machines à distance.

Pour aller plus loin:  [Portrait d'Anne-Cécile Orgerie](#)
 [Numérique: le grand gâchis énergétique](#)



LE STOCKAGE DES DONNÉES AU COURS DU TEMPS



Même si l'informatique n'était pas au programme quand j'étais lycéenne dans les années 70, la prof de maths nous a fait découvrir les algorithmes avec des cartes perforées.

La formule mathématique que vous venez de trouver va être appliquée à cette carte, que par la machine et les instructions vont être exécutées.

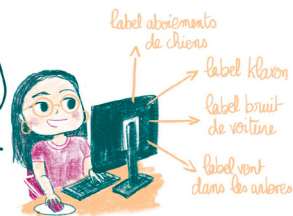
Et oui, ça vous en bouche un coin, je sais !



Pour aller plus loin: [Stockage de données: les promesses de l'ADN synthétique](#)

AU-DELÀ DES DONNÉES STRUCTURÉES?

Avec l'outil de comparaison, je peux classer 100 000 sons automatiquement, sans avoir besoin de les écouter.



De nombreuses données issues de mesures physiques (sons, photos, vidéos...) ne donnent pas accès à l'information utile de façon aussi directe que des données structurées (noms, adresses...). Le **traitement du signal et des images** permet d'extraire de ces mesures un petit nombre de caractéristiques significatives. On dispose alors de données structurées qui permettent l'établissement d'un catalogue rendant possibles des recherches et analyses efficaces. On peut par exemple établir automatiquement une classification de sons distincts, comme dans l'application Shazam.

Pour aller plus loin: [De Fourier à la reconnaissance musicale](#)

QUIZZ

Quel système de stockage Marie-Christine Rousset a-t-elle utilisé au lycée ?

- A. Les cartes perforées
- B. La clé USB
- C. L'ADN des géraniums
- D. Les bandes magnétiques

Qu'est-ce qui prend le plus de place à stocker ?

- A. Tous les livres écrits depuis l'origine de l'humanité
- B. Les données produites par le collisionneur de particules du CERN en une minute
- C. L'intégralité du catalogue Netflix

Quelle doit être la durée moyenne d'un extrait musical pour que Shazam identifie le morceau dont il est issu ?

- A. Il faut écouter tout le morceau
- B. Une minute est nécessaire
- C. Quelques secondes suffisent