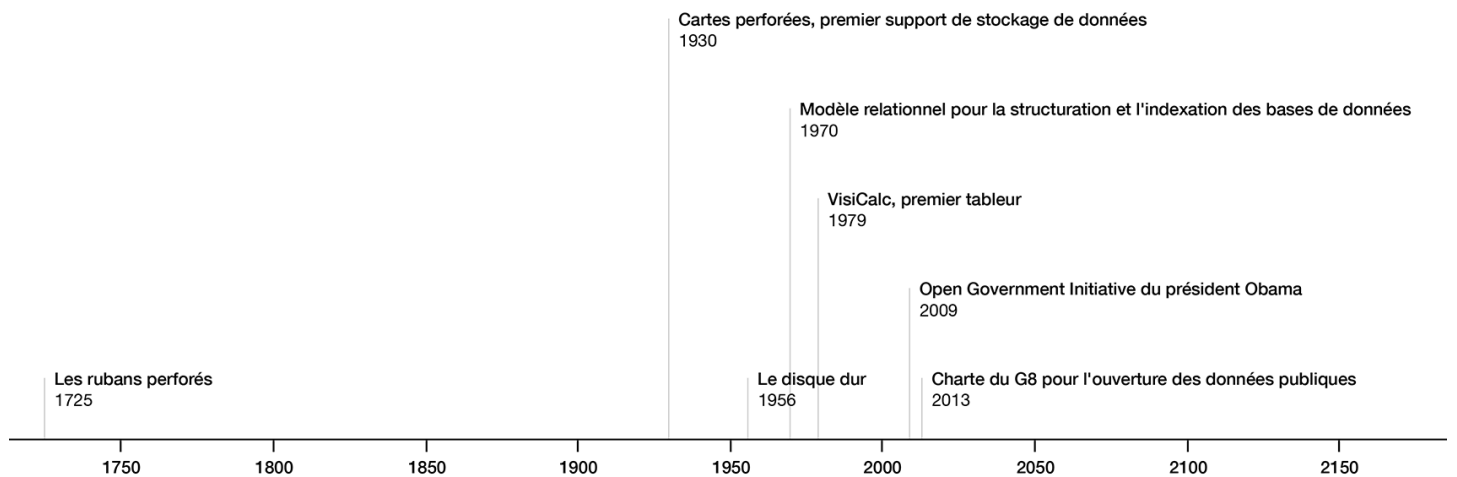


Sciences Numériques et Technologie

Chapitre III

Les données structurées

# I. Histoire et chronologie



Le stockage de données a commencé en 1725 environ avec les rubans perforés, puis les cartes perforées en 1930 et enfin le disque dur en 1956.

La création de bases de données a eu lieu dans les années 1970 avec la création des modèles relationnels.

## II. Les données

### II.1. Généralités

#### Définition

Une **donnée** est une valeur décrivant un objet, une personne, un événement digne d'intérêt pour celui qui choisit de la conserver.

On parle de **donnée personnelle** lorsque la donnée identifie directement ou indirectement une personne physique.

Une **métadonnée** est une donnée qui permet de décrire le contenu d'un fichier.

#### Exemple

- 25/12 est une donnée correspondant à la date de Noël.
- Une adresse postale, un numéro de téléphone sont des données personnelles.
- Les coordonnées GPS, la date et l'heure de la prise d'une photographie numérique sont des métadonnées du fichier contenant la photographie numérique.

### II.2. Données structurées

#### Définition

Pour décrire un objet, on utilise plusieurs **descripteurs**.

Une **collection** regroupe des objets partageant les mêmes descripteurs. La structure de table permet de présenter une collection : les objets en ligne, les descripteurs en colonne et les données à l'intersection. Les données sont alors dites **structurées**.

### Exemple

Si on veut décrire les trousse des élèves, on peut choisir différents descripteurs : la marque, la forme, la contenance, la couleur, etc.

On peut regrouper ces informations dans un tableau.

Numéro de la trousse	Marque	Forme	Contenance	Couleur
1	Eastpak	cylindrique	30 stylos	noire
2	Zipit	plate	10 stylos	bleue
3	Tann's	parallélépipédique	25 stylos	noire
⋮	⋮	⋮	⋮	⋮

### Définition

Une **base de données** regroupe plusieurs collections de données reliées entre elles.

### Exemple

La collection de données suivante peut être reliée à la précédente.

Nom	Trousse
Jean	2
Marie	1
Sophie	2
Claude	3
⋮	⋮

Ainsi on peut dire que Jean possède une trousse de la marque Zipit, qui est plate et bleue et pouvant contenir 10 stylos.

### Persistance ces données

Pour assurer la persistance des données, ces dernières sont stockées dans des fichiers.

Le format CSV (Comma Separated Values) est un **format** de fichier simple permettant d'enregistrer une **table de données**.

### Exemple - utilisation du vocabulaire

La *base de données* d'une bibliothèque regroupe 3 *collections* de données :

- celle sur les livres
- celle sur les abonnés
- celle sur les emprunts effectués

Les objets de la collection des abonnés sont les abonnés eux-mêmes et ont les mêmes *descripteurs* : nom, prénom, numéro de téléphone, adresse.

Des *données* de cette collection sont donc les noms, prénoms, numéros de téléphone et adresses des abonnés. Elles permettent d'identifier une personne physique. Elles sont donc des *données personnelles*. De plus, elles sont *structurées* dans des *tables de données*.

La base de données (et donc les 3 *tables de données*) est enregistrée dans des fichiers au *format* CSV par la bibliothèque d'Arras.

Une *métadonnée* de ce fichier peut être par exemple « utilisateur : bibliothèque d'Arras ».

### III. Le traitement des données

#### Définition

Le **traitement des données** consiste à faire des opérations sur les données appelées **requêtes**. Elles sont exprimées dans un langage informatique tel que le SQL (Structured Query Language).

#### Remarque

Les opérations les plus utilisées sont la recherche, le tri, le calcul, le filtre sur une ou plusieurs tables de données.

### IV. Stockage de données

On stocke en général nos fichiers sur des supports de stockage : internes (disque dur ou SSD) ou externes (disque dur ou SSD ou clé USB), locaux (sur l'ordinateur) ou distants (**cloud**).

#### Définition

Le **cloud** désigne donc l'accès à des données distantes mais aussi à des ressources informatiques telles que le stockage, la puissance de calcul, etc.

#### Remarque

Les ressources du Cloud sont situées dans des serveurs qui doivent être alimentés en électricité constamment et maintenus à des températures suffisamment basses pour fonctionner.

L'impact environnemental est donc très important. En effet, les centres de données (data centers) consomment plus de 10% de l'électricité mondiale. Un centre de données peut consommer autant d'électricité que 30000 habitants européens.

De nos jours, on assiste à un phénomène de surabondance des données. De nouveaux algorithmes se développent afin d'exploiter ces données. L'exploitation de ces données massives (**Big Data**) est en plein essor dans des domaines aussi variés que les sciences, la santé ou encore l'économie. Il y a donc des conséquences sociétales nombreuses : surveillance de masse, exploitation des données personnelles, etc.

Certaines de ces données sont dites ouvertes (OpenData) car leurs producteurs considèrent qu'il s'agit d'un bien commun. Chacun peut donc y accéder et les utiliser comme bon leur semble.

Cependant on assiste aussi au développement d'un marché de la donnée où des entreprises collectent et revendent des données sans transparence pour les usagers.

D'où l'importance d'un cadre juridique afin de les protéger. C'est le Règlement Général sur la Protection des Données (**RGPD**) qui s'en charge.