



*Ce que je dois  
retenir*

Définir une donnée personnelle. Identifier les principaux formats et représentations de données.  
Identifier les différents descripteurs d'un objet. Distinguer la valeur d'une donnée de son descripteur.  
Utiliser un site de données ouvertes, pour sélectionner et récupérer des données.  
Réaliser des opérations de recherche, filtre, tri ou calcul sur une ou plusieurs tables.  
Retrouver les métadonnées d'un fichier personnel.  
Utiliser un support de stockage dans le nuage. Identifier les principales causes de la consommation énergétique des centres de données ainsi que leur ordre de grandeur.

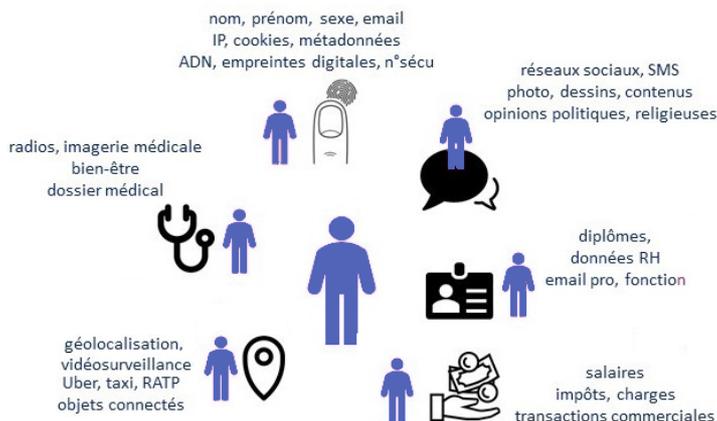
### Données personnelles

Une **donnée personnelle** est toute information relative à une personne physique susceptible d'être identifiée, directement ou indirectement. Peu importe que ces informations soient confidentielles ou publiques.



*Données personnelles :  
vous avez des droits !*

Exemples de données personnelles :



Toute entité manipulant des données à caractère personnel doit se conformer au nouveau règlement mis en vigueur à partir du 25 mai 2018, le **RGPD** et doit faire également l'objet d'une déclaration auprès de la **CNIL** (Commission Nationale de l'Informatique et des Libertés).

### Données ouvertes ou Open Data

Les **Open Data**, ou **données ouvertes**, sont des données auxquelles l'accès est totalement public et libre de droit, au même titre que l'exploitation et la réutilisation. L'**interopérabilité**, c'est à dire la mise la possibilité de mélanger librement ces données, est donc possible. Cela offre de nombreuses opportunités pour étendre le savoir humain et créer de nouveaux produits et services de qualité par exemple.

Source : <https://www.lebigdata.fr/open-data-definition>

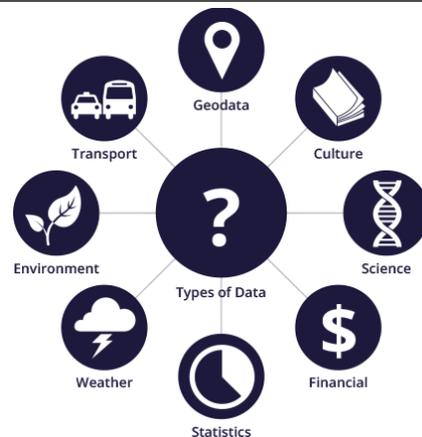
Exemple : Le gouvernement français met à disposition des jeux de données publiques ouvertes via les sites :



[data.gouv.fr](http://data.gouv.fr)



[data.education.gouv.fr](http://data.education.gouv.fr)



Pour aller plus loin : [youtu.be/aHxv\\_2BMJfw](https://youtu.be/aHxv_2BMJfw)

Source : libertic

### Des données structurées : Formats de fichiers et Descripteurs

Pour être utilisables et partageables les données doivent être **structurées** :

Une **donnée** est une **valeur** de type texte, nombre, ... décrivant un **objet**. Plusieurs **descripteurs** peuvent décrire un même objet.

L'interopérabilité des données est possible si des **formats** de fichiers ouverts sont utilisés pour stocker ces données. Les plus utilisés sont :



#### Descripteurs

Index	n_client	nom	prenom	ville
0	1212	Lacasse	Aubrey	Anancy
1	1343	Primeau	Angelette	Tours
2	2454	Gabriaux	Julie	Bordeaux
3	895	Gaulin	Élodie	Lyon

Un objet

Données

## Traitement des données : recherche, filtre , tri, calcul, graphiques, ...

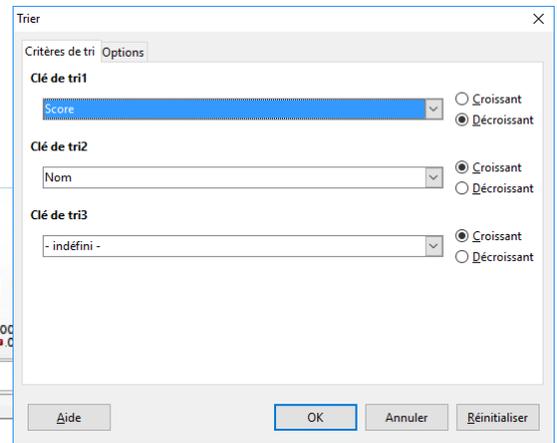
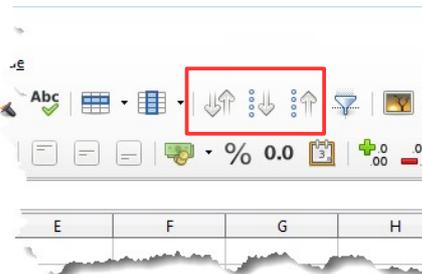
L'intérêt d'obtenir et de structurer des données est de pouvoir les **traiter** de manière très rapide afin d'obtenir les informations souhaitées.

Les **tableurs**, les **bases de données** ou les **algorithmes** sont des outils qui permettent différents **traitement des données** : Filtrer, Trier, Calculer, Rechercher, Exploiter, ...



Pour effectuer un **tri** dans un ordre croissant, décroissant ou alphabétique, il faut sélectionner le bloc à trier et lancer la commande « Tri croissant ou décroissant » ou « Trier » dans le menu « Données » ou cliquer sur l'icône approprié. Il est possible de configurer plusieurs clés de tri et d'utiliser les descripteurs pour faciliter le traitement.

Nom	Sexe	Âge	Score
Jules	G	14	9923
Hugo	G	13	7562
Clara	F	11	6345
Julia	F	10	6345
Louis	F	13	1208
Théo	G	12	890



Pour effectuer ces différents traitements comme le tri de données structurées, on peut également utiliser des **algorithmes** et des **langages de programmation** comme Python. De nombreuses **bibliothèques** permettent de réaliser cela.

*Exemple : la librairie Pandas permet de manipuler les données grâce à ce qu'on appelle un « dataframe ».*

```
In [55]: import pandas as pd
df4 = pd.read_csv("data_test.csv")
```

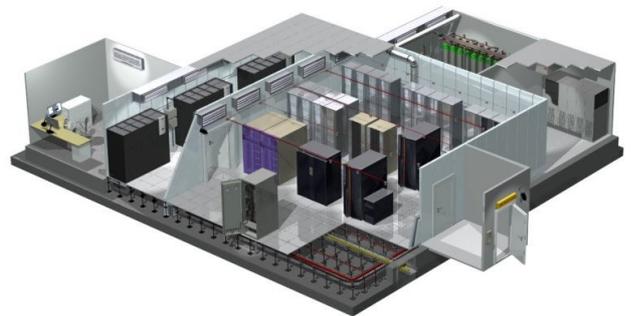
```
In [9]: df4.sort_values(by="date")
```

	date	taille	prix (€)	acheteur
9	2015-12-11	XS	40.72	Gloria Mason
7	2016-02-01	L	16.23	Kathleen Brown
5	2016-04-23	L	15.74	Juan Wallace
4	2016-04-25	3XL	30.79	Julia Bennett
8	2016-05-18	L	13.52	Timothy Carroll
2	2016-05-24	L	34.60	Carl Hunter
6	2016-06-26	M	37.68	Jessica Fuller
3	2016-07-13	XL	44.63	Elizabeth Martin
1	2016-11-08	XL	16.91	Debra Taylor
0	2016-11-10	L	38.36	Christopher Kelley

## Le stockage des données

Les fichiers de données sont stockés sur des supports de stockage internes ou externes, locaux ou distants.

Le stockage distant se fait généralement dans des **centre de données** (en anglais **data center**) qui sont des lieux regroupant des différents équipements : ordinateurs centraux, serveurs, baies de stockage, équipements réseaux et de télécommunications, etc...



Le **cloud** est un service qui permet de stocker, traiter et partager ses données.

Chaque utilisateur peut automatiser le stockage par une **synchronisation** des fichiers depuis son ordinateur ou son smartphone.

Les flux de données et leur stockage pose des problèmes de **sécurité** et de **consommation énergétique**.