



Les données

➤ Qu'est-ce qu'une donnée ?

Une **donnée** est une valeur attribuée à une entité pour la décrire. Il peut s'agir d'un objet, d'une personne, d'un événement ou par exemple d'un numéro de téléphone. L'**information** est l'interprétation que l'on fait d'une donnée.

Les données peuvent être renseignées par un humain, comme lorsque nous nous inscrivons sur un site qui nous incite à renseigner nos données personnelles pour créer un compte. Les données peuvent aussi être capturées et enregistrées par un dispositif matériel, comme les caméras de surveillance à reconnaissance faciale.



➤ Les bases de données

Pour retrouver et traiter des données facilement, on les organise sous forme d'un tableau appelé **table de données**. Une collection regroupe des **objets** partageant les mêmes **descripteurs**. Un **objet** est donné par la liste des **valeurs** de tous ses **descripteurs**. Les données sont alors dites « **structurées** ».



Collection					
Descripteurs	→	Nom	Capitale	Hymne	Superficie (km²)
Une valeur du descripteur « Nom »	→	France	Paris	La Marseillaise	632 734
		Chine	Pékin	La Marche des volontaires	9 596 961
		USA	Washington	The Star-Spangled Banner	9 833 517
Un objet	→	Argentine	Buenos Aires	Himno Nacional Argentino	2 791 810

➤ Métadonnée

Une **métadonnée** est une donnée particulière qui donne des informations sur la donnée principale, comme le titre d'une vidéo. Les données personnelles sont souvent associées à de nombreuses métadonnées qui donnent davantage d'informations sur le fichier principal et sur leur propriétaire. Ces métadonnées varient selon le type de fichier qu'elles décrivent.





➤ Le Big Data

En 2018 IBM a annoncé que 90% des données mondiales ont été créées au cours des deux dernières années. Chaque jour, 2,5 quintillions (10^{30}) d'octets de données sont générés. Ces données constituent le **Big Data** : ensemble de données numériques produites par l'utilisation des nouvelles technologies ou l'échange d'informations. Il répond à trois principes résumés par les cinq « V » de l'analyste Doug Laney :



- le Volume de données de plus en plus conséquent ;
- la Variété de ces données qui peuvent être brutes, non structurées ou semi-structurées ;
- la Vitesse qui désigne le fait que ces données doivent être vite produites, récoltées et analysées en temps réel ;
- La Véracité concerne la fiabilité et la crédibilité des informations collectées ;
- La Valeur correspond au profit qu'on puisse tirer de l'usage du Big Data.

Les données ouvertes : Des données ouvertes ou Open Data, sont des données totalement publiques et libre de droit. De nombreux sites offrent l'accès à ces données. Il ne doit y avoir aucun obstacle technique et aucune restriction à l'utilisation des données, même commerciale.

90%
DE L'ENSEMBLE DES DONNÉES
AUJOURD'HUI DISPONIBLES
ONT ÉTÉ CRÉÉES CES
2
DERNIÈRES
ANNÉES

2 Mo

La quantité de données produites par personne... et par **seconde** dans le monde en 2020, selon IDC

Le potentiel colossal du « big data »

