

Thème 3 : Les Données Structurées

Bilans

Les données

Une donnée est une information qui se rapporte à une chose, une personne ou un événement : un nom, une date, un lieu, un titre, un son, une image, etc.

Des descripteurs sont utilisés pour décrire les données (exemple, pour une personne il faut le nom, le prénom, la date de naissance, l'adresse, le numéro de téléphone, etc.).

Quand des données partagent des mêmes descripteurs, on peut les organiser en collection (un annuaire téléphonique, une playlist), chaque ligne de cette collection est un objet (un nom, un prénom, une date de naissance, une adresse, etc...).

Exemple de table de données

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

Une base de données regroupe plusieurs collections de données qui ont des liens entre elles. Les données peuvent être personnelles, elles concernent des personnes, ou elles peuvent être publiques, on parle alors d'open data. Le site data.gouv.fr est une mine d'informations sur les données publiques ou ouvertes que l'on peut consulter et utiliser.

Le règlement général sur la protection des données (RGPD), règles européennes, permet de protéger les individus contre les risques liés aux usages abusifs des données collectées par des organismes.

Les métadonnées servent à décrire un objet : pour une photo prise par un smartphone, ça peut être la date de prise de vue, le lieu (obtenu par le GPS), la taille de la photo, l'appareil, la définition, etc.

Les données structurées

Les données, dans une collection qui partagent les mêmes descripteurs, peuvent être structurées en table (ce sont des tableaux) : les objets sont en lignes, les descripteurs sont en colonnes et les données sont à l'intersection. Ces données sont dites structurées.

Les fichiers avec l'extension csv sont des fichiers de données structurées.

Cette structuration facilite la lecture et le traitement de ces données.

Le traitement des données

Les traitements peuvent être des recherches d'informations, des tris, des calculs, des représentations graphiques. Pour ces traitements, les données peuvent être ouvertes dans un tableur ou bien on peut utiliser un langage de programmation, comme Python.

Le cloud

Les données très nombreuses que nous produisons sans cesse doivent être stockées sinon elles disparaissent. Divers supports de stockage existent ;
des stockages locaux comme :

- des stockages internes avec les disques durs,
- des stockages externes avec les CD DVD ou clés USB,
et des stockages distants comme :
- les nuages ou cloud.

Un cloud est en fait un immense bâtiment renfermant des milliers de serveurs informatiques, on les appelle les data centers.

La consommation des data centers

Les data centers consomment énormément d'énergie et produisent de la chaleur qu'il faut dissiper en les climatisant.

En 2018, les industries numériques consomment 10 % de l'électricité mondiale, dont 3 % pour les data centers.

La consommation énergétique des data centers double en moyenne tous les quatre ans.
Des actions doivent être menées pour limiter les transformations écologiques engendrées par les usages numériques.