

Thème 4 – Les données structurées et leur traitement

Je découvre, p. 50

- Une donnée est une valeur décrivant un élément que l'on veut conserver. Les données constituent la matière première de toute activité numérique. Afin de permettre leur réutilisation, il est nécessaire de les conserver de manière constante. Les structurer correctement garantit que l'on puisse les exploiter facilement pour produire de l'information et les rechercher. Cependant, les données non structurées peuvent aussi être exploitées, par exemple par les moteurs de recherche.
- Aujourd'hui le volume mondial de production de données numériques a explosé. Ceci est dû au fait que les individus sont désormais interconnectés. Ils interagissent avec les objets, générant ainsi une quantité exponentielle de données et de traces numériques au travers de leurs smartphones, ordinateurs, objets connectés (caméras, jouets, compteurs électriques...). Ce volume faramineux de données est ce qu'on appelle communément le « **Big Data** ».
- Les enjeux actuels sont donc la capacité à les stocker, les traiter et les sécuriser. Pour cela, de nouveaux outils de gestion et de nouvelles compétences métiers voient le jour. On cherche aussi à contrer les nombreuses conséquences sociétales tant en termes de démocratie, de surveillance de masse que d'exploitation des données personnelles.

Info : 90 % des données disponibles en ligne en 2019, ont été créées depuis 2017.

J'illustre, p. 50

- Les données ne sont pas permanentes et ont un cycle de vie. Il existe plusieurs étapes de gestion, de leur création à leur effacement quand elles deviennent obsolètes.
- **Réordonner les étapes suivantes pour représenter leur cycle de vie :**
exploitation et partage des données – destruction des données –
création/collecte de données – archivage des données – analyse des données –
sauvegarde des données.

Activité 1 – Structurer les données d'une page Web, p. 51

L'expression « **données structurées** » signifie qu'elles sont définies par des référentiels et présentées dans les champs d'une base de données (type tableur Excel).

Plusieurs **descripteurs** peuvent être utiles pour décrire un même objet (par exemple pour caractériser un contact : nom, prénom, adresse et numéro de téléphone). Il faut aussi les regrouper dans des collections qui partagent les mêmes descripteurs, afin de pouvoir croiser les recherches par la suite.

Activité 2 – Les métadonnées, p. 52

Les **métadonnées** sont des données qui permettent d'en décrire d'autres. Elles résument les informations de base sur les données, facilitant leur recherche et leur gestion. Les métadonnées peuvent être créées manuellement pour être plus précises ou automatiquement pour contenir plus d'informations. Si l'on souhaite identifier une donnée dans une table, les possibilités de descriptions semblent

infinies. On peut alors imposer un filtre pour les distinguer davantage, par la date de sauvegarde, la taille, l'auteur, le type, le nom du fichier. Ainsi, les schémas de métadonnées peuvent être simples ou complexes, mais ils ont tous des choses en commun.

Activité 3 – Les principaux formats de données, p. 53

Les formats de données sont conçus pour faciliter la transmission d'informations. Chaque format possède sa propre écriture et est voué à être utilisé par des programmes spécifiques.

Certains formats sont créés sur la base du même langage, ce qui facilite leur collaboration et d'autres sont ouverts, c'est-à-dire qu'ils peuvent être utilisés et partagés plus facilement. Ces formats sont le support de l'**Open Data** qui permet l'**interopérabilité** des fichiers.

Activité 4 – Traitement des données structurées, p. 53

Traiter des données, c'est savoir donner du sens à des données brutes. Pour être en mesure d'extraire une donnée, il faut savoir réaliser des opérations de recherche, filtre, tri ou calcul sur une ou plusieurs tables de données.

L'intérêt du traitement de données est généralement d'offrir une analyse qui provient de la fusion de celles-ci, de l'extraction d'informations ou de la transformation de sa représentation.

Activité 5 – Analyse des données et visualisation, p. 54

Nos écrans sont remplis de données en tous genres : chiffres, sondages, pourcentages, etc. L'important est de pouvoir visualiser l'information pour

comprendre l'ensemble des liens entre les données les plus importantes.

La **datavisualisation** ou « dataviz » est la capacité de communiquer des données brutes de façon graphique. C'est une des nouvelles formes de traitement de l'information.

Activité 6 – Big Data, p. 55

Le **Big Data** est un terme simple utilisé pour décrire l'abondance des données numériques et l'émergence des moyens développés pour y accéder et l'analyser. Son rôle est de traiter des informations pour acquérir de nouvelles connaissances. Pour en extraire du sens il faut trier d'énormes volumes de données, trouver des corrélations inattendues et des connexions étonnantes. Aujourd'hui le Big Data est déjà utilisé pour apprendre et résoudre des problèmes dans de nombreuses disciplines, notamment grâce aux technologies d'**intelligence artificielle** qui reposent dessus.

Info : Chaque minute, plus de 48 heures de vidéos sont mises en ligne sur YouTube.

Activité 7 – Données dans le nuage, p. 56

Le « nuage » ou « **cloud** » est le réseau de stockage des données dans des serveurs. Ces données étaient conservées sur l'ordinateur de l'utilisateur ou sur du matériel de stockage. Il existe différents types de cloud, avec des fonctionnalités plus ou moins étendues. Cela va du simple espace de stockage pour sauvegarder et partager des fichiers, à une application en ligne qui n'a pas besoin d'être installée sur l'ordinateur.

Ainsi, les **datacenters**, lieux de rassemblement des serveurs, se multiplient pour pouvoir répondre au besoin de plus en plus croissant de stockage. Le maintien du

bon fonctionnement de ces infrastructures pose d'ailleurs de nouveaux problèmes de sécurité et d'écologie.

Activité 8 – La consommation énergétique des centres de données,

p. 57

Avec la numérisation de la société et la population ultra-connectée, les datacenters, cloud et autres outils qui centralisent les données sur des serveurs, sont de véritables gouffres énergétiques et une catastrophe écologique pour la planète.

À son niveau, chacun peut diminuer ces effets en éteignant les appareils reliés à internet lorsqu'ils ne sont pas utilisés (la nuit notamment) et éviter de saturer les serveurs, ou en supprimant par exemple ses vieux emails.

Info : En 2017, l'écosystème numérique représentait environ 7 % de la consommation mondiale d'électricité. D'ici 2020, cette consommation pourrait atteindre 12 % et augmenter d'environ 7 % par an d'ici 2030, d'après le Visual Networking Index de Cisco.