



COMMENT ET POURQUOI ACCUMULE-T-ON AUTANT DE DONNÉES ?

▼ Auteurs et date

- Date de production de la fiche : 12/04/2021
- Laurent Devernay ;Réfèrent Formateurs Occitanie ; simplon.co

RÉSUMÉ

Par définition, une donnée est ce qui est connu, ce qui peut servir de base pour un raisonnement. Au cours de l'évolution de l'être humain, il est vite apparu capital de gérer les données pour pouvoir les réunir, les comparer et les transmettre. C'est aussi pour cela que les supports utilisés ont évolué.

DES DONNÉES PARTOUT

L'une des promesses initiales du web était de permettre à chacun d'accéder plus facilement et surtout gratuitement à l'information. Cependant, la croissance des données nécessite traitement, mise en forme, centralisation et partage, ce qui est à la fois chronophage et gourmand en terme de ressources. Ces dernières se sont multipliées et, notamment d'après IBM, 90% des données produites par l'humanité l'auraient été les 2 dernières années¹. Et cette affirmation se vérifie depuis plus de 30 ans. Sauf que, aujourd'hui, 90% des données stockées ne serviraient à rien².

Imaginez par exemple que chaque voiture autonome collecte 4To de données par jour pour 1h30 de circulation quotidienne environ (même si une partie seulement transite ensuite sur le réseau)³ ! Alors qu'on compte des dizaines de milliards d'objets connectés dans le monde qui captent et transfèrent eux aussi des données, le problème devient conséquent. L'impact environnemental du transfert et du stockage de toutes ces informations n'est pas anodin⁴.

L'EXEMPLE DE LA PHOTOGRAPHIE

La photographie illustre bien l'évolution de nos usages liés au numérique. Le passage de l'argentique au numérique a été une révolution. Diminution des coûts (plus besoin de pellicule ni de développement), nouvelles possibilités et nouveaux besoins (imprimer soi-même ses photos mais avec du papier spécifique et parfois des imprimantes conçues pour cela, facilité de stockage). Et avec les smartphones, tout ceci est devenu encore plus facile en apparence. Ce n'est pas grave de s'y reprendre à plusieurs fois pour réussir une photo et des filtres peuvent facilement être ajoutés. Il n'a jamais été aussi facile de partager ses photos par messagerie ou via les réseaux sociaux. Peu de risque de les perdre, elles sont automatiquement envoyées par défaut vers le cloud. Tout ceci est gratuit, mais seulement si

l'on omet le prix des équipements utilisés ainsi que le coût réel de stockage. Jusqu'à récemment, le stockage sur le cloud était le plus souvent gratuit. Jusqu'à ce que Google réalise que ce sont 4 milliards de photos et vidéos qui sont envoyées sur ses serveurs chaque jour⁵. Le service, au-delà d'une certaine limite, est devenu payant. Plus généralement, la dématérialisation et la gratuité ont impacté notre vision du numérique et en particulier du cloud. Il est nécessaire de garder en tête que le cloud est avant tout un ensemble d'équipements physiques et que tout ce que nous stockons a un coût. Pas forcément un coût financier mais pour la planète (des équipements à fabriquer, à alimenter, à refroidir et qui seront régulièrement remplacés). Sauf que l'on tend toujours plus vers le numérique au quotidien. Pour les démarches administratives mais pas seulement. La dématérialisation des tickets de caisse est proposée (voire imposée) comme une bonne pratique pour l'environnement, ignorant totalement la pollution numérique générée (et la question des données personnelles)⁶. Et la tendance se propage aux factures et autres relevés bancaires. De son côté, suite à des analyses poussées, La Poste a réalisé que ses communications promotionnelles étaient plus impactantes via le numérique qu'en papier⁷.

LE CAS DES DONNÉES PERSONNELLES

Avec le web, un type de donnée en particulier a pris de la valeur : les données personnelles. Une donnée personnelle c'est une information concernant une personne physique. Cela concerne votre prénom, votre nom de famille mais aussi ce qui a trait à votre religion, à vos opinions politiques, aux programmes que vous regardez à la télé, etc. Sur le web⁸, ces informations sont récupérées principalement de trois façons :

1. Les informations que vous donnez volontairement, par exemple lorsque vous remplissez un formulaire pour faire un achat en ligne.
2. Les cookies, ces fichiers texte utilisés par les sites web.
3. Les trackers, des bouts de code conçus pour capter des informations sur l'internaute, notamment les pages qu'il visite, ce qu'il recherche sur le web, etc.

Grâce à tout cela, il est possible de constituer pour chaque internaute un profil. Certaines grandes entreprises font reposer leur modèle économique sur la collecte et la revente de ces données ainsi que sur des algorithmes qui permettent, à partir de données personnelles, de déduire de nouvelles informations sur quelqu'un (si vous avez aimé ce livre, vous aimerez probablement tel autre livre). Parmi ces entreprises, on parlera entre autres de data brokers pour qui les données personnelles sont un business à part entière⁹.

Sauf que ces collectes de données personnelles sont problématiques. C'est pour protéger les citoyens qu'est apparu en 2018 le RGPD¹⁰. Ainsi la collecte des données d'internautes européens ou par des structures européennes ne peut se faire qu'avec :

- le consentement explicite des internautes
- une totale transparence (sur la nature des données collectées et l'utilisation qui en est faite, en particulier si elles sont revendues à des tiers)

- un droit à l'oubli (afin de pouvoir demander à une entreprise de supprimer les données collectées sur soi)

En complément, des outils pour se protéger des trackers voient le jour (Ghostery, Blacklight, etc). Il est dans tous les cas important aujourd'hui de se demander ce que deviennent nos données.

SOURCES

1. Bernard Marr. How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read [en ligne], 2018. Disponible sur [Forbes](#) [04/06/21] [↩](#)
2. Karlin Lillington. Digital pollution: 'We have a pandemic of data and 90% of it is crap' [en ligne], 2020. Disponible sur [irishtimes.com](#) [04/06/2021] [↩](#)
3. Brian Krzanich. Data is the New Oil in the Future of Automated Driving [en ligne], 11/2016. Disponible sur [newsroom.intel.com](#) [04/06/2021] [↩](#)
4. Gerry McGovern. The data pollution problem [en ligne], 12/2020. Disponible sur [gerrymcgovern.com](#) [04/06/2021] [↩](#)
5. Shimrit Ben-Yair. Updating Google Photos' storage policy to build for the future [en ligne], 11/2020. Disponible sur [Google](#) [04/06/2021] [↩](#)
6. Les tickets de caisse et leurs impacts environnementaux, 05/02/2020. Disponible sur le site [Zéro déchet Strasbourg](#) [04/06/2021] [↩](#)
7. Pour une communication responsable. L'impact environnemental des supports de la communication clients. La Poste [en ligne], 2020. Disponible sur [le site de La Poste](#) [04/06/2021] [↩](#)
8. Do Not Track (série documentaire consacrée à la vie privée et à l'économie du Web). Réalisée par Brett Gaylor, 2015. Disponible sur [donottrack-doc.com](#) [04/06/2021] [↩](#)
9. Wolfie Christl. Corporate Surveillance in Everyday Life, [en ligne], 2017. Disponible sur [crackedlabs.org](#) [04/06/2021] [↩](#)
10. RGPD, de quoi parle-t-on? [↩](#)