

4. Big Data & données personnelles

Fabrice LEQUEUX

Maître de conférences à l'Université Paris Sud XI

Université Paris I Panthéon-Sorbonne

ENSTA IP Paris



I. Données personnelles



Dans notre monde numérique, la collecte, le traitement et l'utilisation des **données personnelles** revêt une importance stratégique pour la stratégie des firmes.

Les données personnelles sont les informations qui permettent d'identifier directement ou indirectement une personne physique.

Elles correspondent aux noms, prénoms, adresses (physique et électronique), numéro de téléphone, lieu et date de naissance, numéro de sécurité sociale, état matrimonial, nombre d'enfants, numéro de carte de paiement, agenda, contacts, plaque d'immatriculation, photos et vidéos, empreinte digitale, ADN, etc.



I. Données personnelles



Fabrice Rochelandet, dans son article *Economie des données personnelles et de la vie privée*, distingue les données objectives des données subjectives.

Les **données objectives** sont toutes celles qui concernent l'identification et le contact, le signalement, mais aussi les données sociodémographiques et les données juridiques et financières.

Les **données subjectives** regroupent quant à elles les préférences et les centres d'intérêts, les opinions religieuses et politiques, les données comportementales, les données géographiques et les données relationnelles.

Toutes ces données individuelles deviennent **personnelles** lorsqu'elles permettent une identification de la personne ou de sa singularité dans un groupe social donné.

I. Données personnelles



La question de la collecte des données personnelles n'est pas nouvelle : depuis longtemps les entreprises de **vente par correspondance** collectent des données personnelles sur leurs clients et revendent ces fichiers à d'autres entreprises commerciales.

Mais avec le développement des performances informatiques, et la multiplication des terminaux et applications numériques, de plus en plus de données personnelles sont dévoilées et collectées auprès des utilisateurs, pour être traitées et utilisées à des fins plus ou moins morales, plus ou moins légales...



I. Données personnelles



Ces empreintes des informations laissées sur le Net ou sur les smartphones sont captées par diverses technologies qui les stockent, les transforment et les transmettent.

La production exponentielle de données alimente aujourd'hui de nombreuses activités économiques et **le marché de l'exploitation des données personnelles** a explosé et fait rayonner des secteurs d'innovations comme le big data ou l'intelligence artificielle.

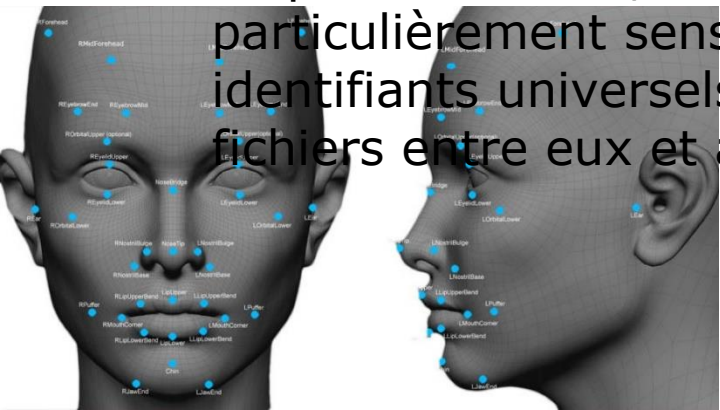
Cette nouvelle **matière première brute** semble s'être peu à peu imposée comme le fer de lance d'une révolution industrielle 3.0 qui tend à s'étendre et se perfectionner.

I. Données personnelles



Aujourd'hui, avec le recoupement des informations, les bases de données des *Data brokers* peuvent identifier et retrouver avec précision n'importe quel individu, connaître ses habitudes, son entourage social, ses transactions commerciales, ses déplacements, ses données médicales, biométriques, etc.

Certaines de ces données, comme le numéro de sécurité sociale ou les données biométriques (empreinte digitale, empreinte faciale, échantillon ADN, etc.), sont particulièrement sensibles, car elles fonctionnent en tant qu'« identifiants universels », qui permet de raccorder différents fichiers entre eux et ainsi d'opérer leur interconnexion.



I. Données personnelles



Le plus grave encore, toutes ces données sont concentrées dans les mains de quelques entreprises seulement. Ainsi, on estime que Google, Amazon, FaceBook et Apple détiennent **80% des données personnelles**.



I. Données personnelles



Par exemple, sur Facebook, un utilisateur *like* une marque de yaourts nature. **Facebook**, plateforme d'intermédiation, va capter ces données et pouvoir lui proposer des publicités adaptées. Idem pour une recherche sur **Google**.

Mais elle va aussi pouvoir les vendre aux sociétés de data-brokers. La société américaine **Datalogix**, par exemple, traque les datas issues des transactions bancaires transitant par les sites de commerce en ligne ou chez les commerçants en magasins physiques.

Acxiom est l'un des data brokers les plus influents. Il fournit des données et des statistiques pour des sociétés de marketing et de détection de la fraude. Elle posséderait d'après la Federal Trade Commission environ 700 millions de données sur les consommateurs à travers le monde, lui permettant de dégager un revenu de près de 850 millions d'euros en 2016

I. Données personnelles



Parallèlement à ces géants (et en partenariat avec eux), fleurissent donc une multitude de **data brokers** qui collectent ou achètent un maximum de données personnelles, les croisent, et les analysent pour les revendre à prix d'or, et réaliser ainsi des bénéfices records.

Selon la dernière étude d'IDC, le chiffre d'affaires mondial issu des activités de Big Data est estimé à plus de 187 milliards de dollars en 2019, soit une augmentation de plus de 50% par rapport à 2015 (122 milliards de dollars). Plus encore, Wikibon prédit un taux de croissance annuel de près de 14.4% pour le secteur.



II. 2-sided markets



En sciences économiques, un **marché biface** est un marché spécifique qui entretient l'existence de deux clientèles tout à fait différentes mais finalement interdépendantes l'une de l'autre.

Par exemple, les médias proposent à la fois un support publicitaire à des annonceurs et un contenu « journalistique » à des lecteurs ou téléspectateurs. On peut alors considérer qu'ils s'inscrivent sur un marché biface. Ils prennent alors le rôle d'intermédiaire entre les clients et les annonceurs.



II. 2-sided markets



Par exemple, une entreprise qui facilite les paiements comme VISA, MASTERCARD ou même PAYPAL peut aussi être considérée comme agissant sur un marché à double face.

D'un côté les consommateurs paient une redevance pour l'utilisation du service (ensuite gratuite) et d'un autre côté, les commerçants reversent un pourcentage à cette entreprise pour chaque utilisation de la carte dans leur magasin.

The PayPal logo, featuring the word "PayPal" in a stylized blue font with a white outline.



II. 2-sided markets



La littérature économique identifie 3 caractéristiques essentielles des marchés à double face :

- Ils utilisent une plate-forme, qui propose **des services différents** à deux côtés du marché, proposant à chaque côté un prix qui lui est propre
- Le marché présente des **externalités de réseau croisées** : l'utilité qu'un agent retire de sa participation au service offert par la plate-forme dépend du nombre de participants de l'autre côté du marché, nombre qui dépend lui-même de la stratégie de prix de la plate-forme.
- La plate-forme profite d'un **pouvoir de marché bilatéral** : elle décide du prix demandé à chaque côté du marché et fixe le plus souvent des prix uniformes (le prix est le même pour tous les agents d'un côté donné du marché).



III. Le tracking



Sur les marchés numériques à double face, la question des données personnelles est tout à fait décisive, car il est alors possible d'accroître l'efficacité d'un **service qui devient personnalisé** en fonction des données personnelles de l'utilisateur (marketing ciblé),

Les modes de collecte se sont très diversifiés avec les technologies numériques. Elles peuvent aller de la collecte d'informations contenues via **un formulaire** rempli volontairement par les individus jusqu'à **l'enregistrement de traces** (habitudes de navigation, localisation géographique de l'adresse de connexion, sites consultés, relations établies avec des individus ou des réseaux, géolocalisation mobile ...).

III. Le tracking



Le modèle économique des acteurs majeurs de l'économie numérique, du web (Google, Amazon) et des réseaux sociaux (Facebook, Twitter) repose en grande partie sur l'exploitation des données personnelles des utilisateurs.

Ainsi pour Google, le rachat de la plateforme publicitaire DoubleClick (pour 3,1 milliards de dollars en avril 2007 soit 2x plus que pour YouTube en octobre 2006) constitue la pierre angulaire de sa stratégie de données personnelles.



III. Le tracking



Pour comprendre de quoi il s'agit, il suffit de se rendre dans les cookies de son navigateur, ces petits fichiers espions qui marquent le passage de l'internaute sur les sites Web. Il y a fort à parier qu'on y trouvera quelques fichiers du type 'hpu@doubleclick.txt', traces du parcours de l'utilisateur sur la Toile.

A tel point que Doubleclick peut définir des données socio-démographiques (âge, sexe ou informations comportementales) sur l'utilisateur puis elle rend compte à ses clients annonceurs de l'impact de leurs campagnes et les conseille sur les changements éventuels à y apporter.

III. Le tracking



Si c'est
Gratuit
c'est toi le
produit

III. Le tracking

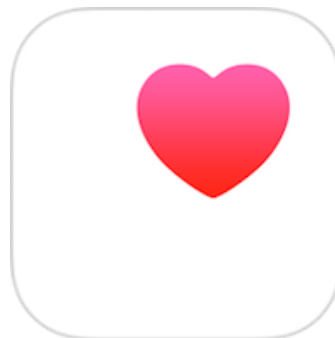


Le principal vecteur de valeur pour la donnée réside dans sa capacité à **influer sur les décisions** ou à donner une information précieuse.

C'est dans cette analyse approfondie de la data, le croisement des renseignements qu'elle fournit, et dans les perspectives stratégiques qu'elle laisse entrevoir que réside la nouvelle valeur de la donnée. On observe alors une mutation substantielle du secteur du **Big data** vers le **Smart Data**.



GPS



Apple santé



Objets connectés

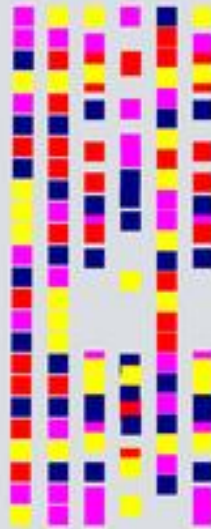
Conclusion



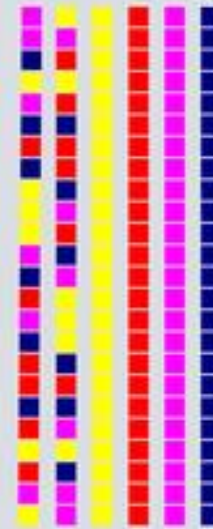
BIG DATA



ANALYTICS



DECISIONS



Conclusion



Applications :

- Publicité, marketing digital
- Médecine, santé : l'ère du microscope temporel
- Sport, politique : prévisions plus fiables
- Sécurité, terrorisme : optimisation des effectifs de police
- Etc etc

Conclusion

