

L'internet, vecteur de puissance des Etats-Unis ?

4 - Commerce des mégadonnées (Big Data)

samedi 28 avril 2018, par [Laurent BLOCH](#)

Alors que l'Intelligence artificielle passionne, Laurent Bloch aborde dans ce quatrième chapitre le commerce des mégadonnées (Big Data). Il présente successivement avec pédagogie Le commerce de données et ses acteurs ; Concurrence et conflits réglementaires ; Un pays évincé de ce domaine serait condamné au déclin.

***Diploweb.com*, publie cet ouvrage de Laurent Bloch, *L'Internet, vecteur de puissance des Etats-Unis ?* pour proposer à chacun les éléments nécessaires à une juste appréciation de la situation. Ce livre est déjà disponible sur Amazon au [format numérique Kindle](#) et [au format broché imprimé sur papier](#). Il sera publié ici sous forme de feuilleton, chapitre par chapitre, au rythme d'environ un par trimestre.**

L'industrie des mégadonnées est un enjeu crucial dans la mesure où le négoce de ces données en masse est un secteur en plein essor et déjà très lucratif ; la quasi-totalité des autres secteurs de l'économie, de la santé à l'agriculture en passant par les transports, en sont affectés, en bien ou en mal selon leur capacité à s'y adapter. [Les entreprises américaines](#) géantes du cyberspace, [Google](#), [Apple](#), [Facebook](#), [Amazon](#) et leurs collègues, sont à l'avant-garde de leur collecte et de leur exploitation : les Européens pourront-ils rattraper leur retard ?



Le commerce de données et ses acteurs

Les GAFA (Google Apple Facebook Amazon) sont des entreprises de première grandeur, leaders dans le cyberspace, toutes américaines. Apple est la troisième entreprise américaine par le chiffre d'affaires avec 234 milliards de dollars, juste derrière le géant de la grande distribution Walmart et le pétrolier Exxon-Mobil (au classement mondial Apple n'est qu'au 12ème rang derrière quelques pétroliers chinois et européens, Volkswagen, Samsung et d'autres, cf. [le classement ici](#)). Google est à 75 milliards de dollars,

Les revenus de ces entreprises proviennent en partie (pour Apple et Amazon) ou en quasi-totalité (pour Google et Facebook) de la commercialisation de données secondaires laissées sur leurs sites par les internautes. Chaque recherche sur le site de Google, chaque bavardage sur Facebook, chaque achat sur Amazon engrange dans les greniers électroniques de ces entreprises des informations dites secondaires, éventuellement calculées à partir des méta-données (origine et destination des communications, identité des interlocuteurs, adresses réseau, dates et heures...), relatives à vos goûts, votre résidence, vos fréquentations, etc., qui, dûment anonymisées, agrégées et soumises à des traitements statistiques, seront vendues à des agences publicitaires, des cabinets de marketing, des agences matrimoniales, des opérateurs touristiques, etc. Quant aux données primaires, ce sont la teneur de vos bavardages ou vos messages de blog, que vous aurez fournis gratuitement et qui attireront d'autant plus d'internautes que vous aurez raconté des bêtises et publié des photos scabreuses qui vous nuiront auprès des conjoints et employeurs potentiels pour le reste de votre vie en ce bas monde.

Les données, qui deviennent des informations dès lors qu'elles atteignent un esprit humain et qu'elles le modifient, comme nous l'apprend Gilbert Simondon, sont un bien en réseau, c'est-à-dire que plus elles sont « consommées » (consultées), plus leur valeur s'accroît (des données auxquelles personne n'a accès ont une valeur nulle, elles ne transmettent aucune information). Une donnée, sauf si elle est protégée par un dispositif de contrôle d'accès tel que chiffrement ou tatouage électronique, est un bien non-rival (la consommation du bien par un agent n'a aucun effet sur la quantité disponible de ce bien pour les autres individus) et non-exclusif (une fois que le bien est produit, tout le monde peut en bénéficier), en d'autres termes un bien public (cf. Wikipédia).

Il ne faudrait surtout pas croire qu'il ne s'agirait là que de l'activité habile de quelques jeunes gens décontractés autour de deux ou trois ordinateurs dans le garage de leurs parents. Ainsi on a pu lire sous la plume de Charles de Laubier (dans La Tribune du 20 février 2015) que le prélèvement de 20 % du prix de la course effectué par Uber sur la recette du chauffeur de VTC affilié relevait « du servage », mais le même supplément hebdomadaire que La Tribune a consacré à cette plate-forme d'intermédiation nous apprend quelques pages plus loin qu'Uber employait à l'époque 450 développeurs à San Francisco et à Amsterdam, qu'il faut bien rémunérer (sans oublier les autres personnels, par exemple à la hot-line). D'ailleurs ces 20 % sont à comparer aux 33 % des compagnies de taxis traditionnelles.

Un livre édité par Diploweb.com, format Kindle et broché

Uber est une plate-forme d'intermédiation parce qu'au lieu de simplement percevoir un fermage elle conserve, organise et exploite toutes les données qu'elle reçoit sur ses chauffeurs et sur ses clients. Mais c'est encore une petite entreprise comparée aux GAFAs. Il faut savoir que Google ou Amazon exploitent chacun des millions de serveurs à la surface de la terre, ils achètent des centrales hydroélectriques pour en assurer le fonctionnement, ils exploitent leurs propres réseaux de fibres optiques transocéaniques. Bref, derrière le commerce éthéré et en apparence immatériel des données il y a une industrie lourde.

Concurrence et conflits réglementaires

Ce commerce de données produites gratuitement par des centaines de millions d'internautes et collectées par les entreprises qui ont créé les plates-formes géantes que sont les sites des GAFAs est d'ores et déjà un axe de l'économie mondiale, mais ce n'est qu'un début parce que ce phénomène qui a commencé par concerner la publicité et le marketing étend progressivement son influence à tous les secteurs.

On sait que sous le régime de l'économie numérique, que l'on propose de baptiser économie, et qui est la concurrence monopolistique, le gagnant ramasse tout : Google ou Facebook disposent d'une telle avance, et partant d'une telle avance financière, qu'il n'est pas possible de créer un concurrent sur le même terrain. Mais si la locution « concurrence monopolistique » évoque la notion de monopole, elle suggère aussi celle de concurrence, parce qu'il est possible de créer une activité dans une niche de marché voisine mais séparée. C'est ainsi qu'Apple, que beaucoup croyaient condamné il y a vingt ans face à Microsoft, a su se réinventer et est aujourd'hui trois fois plus gros que son concurrent.

Un pays évincé de ce domaine serait condamné au déclin

La négociation du projet de traité de libre échange transatlantique (TAFTA/TTIP) place ces entreprises géantes de l'industrie des données dans une position à bien des égards analogue à celle des marchands britanniques d'opium qui voulaient imposer au marché chinois la commercialisation de leurs stupéfiants, d'où résultèrent au XIX^e siècle les deux guerres de l'opium qui asservirent la Chine, jusqu'alors première puissance économique mondiale. De la même façon, les grands acteurs de l'Internet souhaitent accéder aux divers marchés nationaux sans en accepter les systèmes fiscaux ni en respecter les législations sur la protection des données personnelles, pour y vendre ce qui semble bien être une marchandise addictive.

[La puissance publique française](#) recueille et crée une masse considérable de données dont elle ne tire pas tout le bénéfice potentiel, ni ne les met suffisamment à la disposition du public. Il est paradoxal que pour obtenir des données démographiques ou économiques sur la France, plutôt que le site de l'Insee, il faille consulter les sites d'organismes internationaux ou privés qui font le travail de repartir des données de base de l'Insee, de les croiser avec d'autres sources et de les agréger de façon utilisable par quelqu'un qui ne soit pas un statisticien expert : Eurostat, l'OCDE, Wikipédia, Xerfi par exemple. Signalons d'ailleurs que cette difficulté d'interprétation des données brutes publiées par l'Insee et d'autres organismes publics crée un marché pour des entreprises qui les mettent en forme et procurent des outils de navigation, comme par exemple la société Spallian.

De même, il a fallu attendre plus de trente ans pour que l'IGN mette à disposition gratuite du public ses fonds de cartes au niveau communal, et encore sous une forme difficile à utiliser. C'est un contraste saisissant avec la doctrine des agences gouvernementales américaines selon laquelle le contribuable n'a pas à payer une seconde fois pour accéder à des données pour lesquelles il a déjà payé une fois par ses impôts.

Ces politiques différentes ne sont pas sans conséquences économiques : la restriction française d'accès aux données publiques est un handicap pour nos entreprises. Il y a des décennies que les entreprises américaines peuvent accéder aux bases de données d'American Demographics qui mettent en forme de façon inventive les données fournies par le Census Bureau, éventuellement au niveau de l'îlot, ce qui représente un atout considérable pour leur politique commerciale. Sans que ce soit rigoureusement impossible, obtenir la même qualité d'information en France demande un parcours du combattant hors de portée d'une entreprise de taille moyenne ou petite. La situation est la même pour les données cadastrales, elles aussi d'un intérêt social et économique considérable.

L'examen des possibilités d'accès aux données d'EdF ou de la SNCF conduit à des conclusions analogues.

Plus, tout de suite

Ce livre de Laurent Bloch, *L'Internet, vecteur de puissance des Etats-Unis ?*, est déjà disponible sur Amazon au [format numérique Kindle](#) et [au format broché imprimé sur papier](#).

P.-S.

Précédemment responsable de l'informatique scientifique de l'Institut Pasteur, Directeur du Système d'Information de l'Université Paris-Dauphine. Il est auteur de plusieurs ouvrages sur les systèmes d'information et leur sécurité. Il se consacre à la recherche en cyberstratégie. Auteur de « Internet, vecteur de puissance des Etats-Unis », éd. Diploweb 2017.