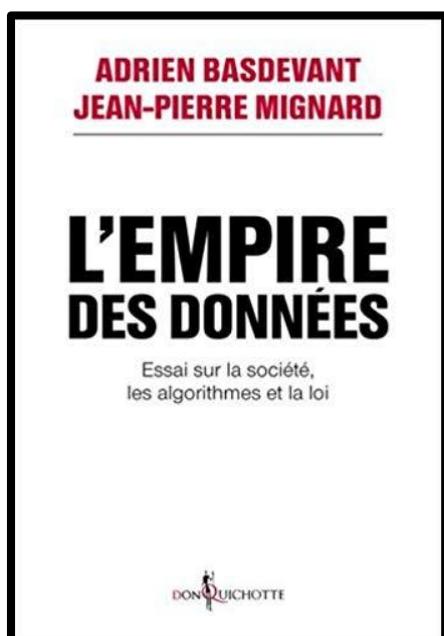

L'EMPIRE DES DONNEES.

ESSAI SUR LA SOCIETE, LES ALGORITHMES ET LA LOI

Adrien Basdevant, Jean-Pierre Mignard. Editions Don Quichotte, 2018.

Clarisse Buchot, Alyona Lapko, Pierre Ledoux, Fabien Reichard, Hugo Sanchez



Adrien Basdevant et Jean-Pierre Mignard sont avocats au Barreau de Paris et associés au sein du cabinet Lysias Partners. Ces deux hommes de droit passionnés par les questions numériques font le constat dans leur livre "L'Empire des données" d'un changement de paradigme avec l'arrivée des données. Les relations entre Etats, entreprises privées et consommateurs-citoyens ont profondément évoluées. Selon eux, on ne peut pas faire marche arrière face à ce coup data. Il s'agit désormais de créer des lois et des pratiques éthiques pour accompagner ces évolutions.

Pour les auteurs, il n'y a qu'une seule question qui vaille : *"comment concilier l'efficacité d'une technologie avec des principes de vie sociale et de liberté individuelle ? Et quel nouveau partage de responsabilité doit s'opérer entre la puissance publique, les entreprises privées et la société civile ?"*

Leur point de départ est simple. Depuis la démocratisation de l'usage des ordinateurs personnels dans les années 1980, le monde bascule dans une nouvelle ère : celle des données. Celles-ci deviennent un outil incroyablement efficace, à des fins de marketing dans un premier temps, mais aussi dans des domaines comme la santé, la fiscalité ou encore la lutte contre le terrorisme. Le "Big Data" arrive au XXIème siècle comme un processus prometteur qui pourrait repousser les limites de l'intelligence, qu'elle soit artificielle ou non. Cependant, une telle avancée ne signifie pas pour autant progrès. Nombreux sont les scientifiques et intellectuels redoutant l'influence des données dans notre vie en société et les intrusions dans la vie privée des individus.

Si comme les auteurs le pensent, le monde est en proie à un "coup data" alors il faut des mettre en place des principes d'éthiques, de régulations juridiques et de gouvernance pour s'en prémunir.

Pour rendre compte de cet ouvrage, nous traiterons dans une première partie de l'importance constante de la donnée au fil de l'Humanité pour préciser en quoi elle devient aujourd'hui capitale. Il s'agit d'explicitier ce qu'est une donnée, comment elle se collecte et comment elle peut être utilisée. Dans une seconde partie, nous aborderons le "coup data" sur la gouvernance mondiale cher à Adrien Basdevant et Jean-Pierre Mignard, soit l'avènement du Big Data et les mesures juridiques prises pour contrer ce phénomène aussi prometteur que dangereux, avant de conclure sur les propositions des auteurs.

L'EMPIRE DES DONNÉES

QUE SONT LES DONNEES ?

Pour définir ce qu'est une donnée, les auteurs empruntent une définition de Serge Abiteboul dans Sciences des données : de la logique du premier ordre à la Toile (2012). La donnée est une "description élémentaire d'une réalité. La donnée est dépourvue de tout raisonnement, supposition, constatation, probabilité. Étant indiscutable ou indiscutée, elle sert de base à une recherche, à un examen quelconque."

En fonction du contexte ou à elle seule, une donnée peut être insignifiante ou capitale, mais son importance s'accroît considérablement lorsqu'elle devient le carburant d'un puissant moteur que sont les algorithmes. Il y a des algorithmes partout : dans le fonctionnement simple de nos appareils électroniques connectés, sur notre fil d'actualité Facebook, Twitter, sur notre navigateur web, etc. Selon les auteurs, les algorithmes ont déjà une place prépondérante dans nos vies et ce de manière inconsciente. Pourtant, tout indique qu'ils seront toujours plus importants : les auteurs précisent en effet que ce seront ces algorithmes qui risqueront de déterminer l'obtention d'un prêt et de son taux à la banque ou d'une nouvelle assurance mais aussi la répartition des élèves dans une filière post-bac.

En ce sens, Adrien Basdevant et Jean-Pierre Mignard affirment que la donnée, en tant que carburant de ces algorithmes déterminants, sera la ressource du XXIème siècle. C'est déjà le cas aujourd'hui où l'on ne parle plus de "données" mais de "Big Data".

Selon les auteurs, le Big Data sera le nouveau modèle d'analyse du monde à travers le nombre croissant de données collectées. Ils le prouvent en s'appuyant sur des données chiffrées. Le Big Data, en 2011, est la création de plus de données que dans toute l'histoire de l'humanité. En 2013, 4,4 zettaoctets ont été créés par nos soins à travers nos messages, e-mails et clics soumis à nos ordinateurs, tablettes, smartphones et autres. Selon les auteurs, les données ne sont pas collectées sans but précis : elles représentent des

informations qui peuvent être précieuses à des fins de marketing. C'est pourquoi certains experts parlent de "profilage" : ces données produites par notre activité sur l'Internet vont être collectées et utilisées par divers algorithmes qui pourront ainsi nous catégoriser pour mieux nous cibler sur notre identité, sur nos pratiques, sur nos préférences vestimentaires, alimentaires, politiques, etc. Les auteurs affirment que, grâce aux données, le marketing est passé d'un marketing de masse peu efficace à un marketing ciblé, segmenté qui permet d'améliorer son efficacité. Le basculement dans l'ère de la Big Data cristallise aujourd'hui les préoccupations des experts qui préconisent une régulation de ces processus de collecte et de traitement.

L'HISTOIRE DES DONNEES

Afin de mieux comprendre tout l'enjeu des données, les auteurs offrent à leur lecteur un historique des nombres et de ses sciences. En effet, selon eux, l'apparition des données est une conséquence de l'évolution de la relation entre l'Homme et les nombres. L'Homme a toujours compté. Au départ, il s'agissait d'une nécessité pour vivre en société. On peut prendre l'exemple des censeurs pendant la Rome Antique qui comptaient les familles et évaluaient leur patrimoine. Puis, l'utilisation des nombres s'est peu à peu transformé en statistique. Le terme "statistique" vient du latin "Status" qui signifie "Inventaire", mais on peut aussi trouver un lien avec "statisticus" qui veut dire "relation avec l'Etat". Ce n'est pas anodin, la statistique a d'abord été utilisée par les Etats pour s'informer sur les individus de leur territoire comme a pu le faire Gottfried Achenwall, le créateur de la statistique, dans l'Allemagne du XVIIIème siècle. Ainsi, le nombre devient un outil de pouvoir, ou comme le disait Alexis de Tocqueville "*le nombre seul fait la loi et le Droit. Toute la politique se réduit à une question d'arithmétique*".

La démocratisation de la statistique est capitale : selon les auteurs, c'est elle qui fait passer l'humanité d'une logique descriptive à une logique prescriptive en calculant les risques et dangers de chaque action. C'est par cette approche désormais prescriptive que seront pensés les Etats-providences et les systèmes de sécurité sociale. Cette approche prescriptive de la société va de pair avec l'avènement d'une théorie critiquable : l'Homme Moyen d'Adolphe Quetelet. Au XIXème siècle, cet astronome, mathématicien et statisticien belge est le premier à étudier des données telles que la naissance, le décès, la taille et tout autre phénomène pour créer un "*Homme Moyen parfaitement déterminé que l'on pourrait [...] considérer comme le type du beau ; et tout ce qui s'éloignerait le plus de ressembler à ses proportions ou à sa manière d'être constituerait les difformités et les maladies*." Ainsi par cette logique, la moyenne devient la norme.

Cependant l'Homme après avoir appris à utiliser la moyenne invente les probabilités. Ce changement radical dans l'utilisation du dénombrement apporte des risques inhérents. Le raisonnement est passé de la déduction à l'induction. Une autre transition de

raisonnement est importante : la causalité et la corrélation. Là où le premier outil de raisonnement met en exergue des rapports de causes à effet indubitables, le second se contente de marquer une relation entre les deux éléments sans lien de causalité avéré. Or l'utilisation de la corrélation est devenu la norme privilégiée de la plupart des algorithmes. Immédiatement, la confiance attribuable à ces algorithmes doit être limitée, en ce que le raisonnement proposé tend à se satisfaire de sophismes.

Comme le précisent les auteurs, les années 1970 marquent un nouveau virage dans l'histoire du dénombrement et de la statistique puisque l'on assiste à la démocratisation des ordinateurs, véritables machines à statistiques et algorithmes. Ainsi, chaque personne munie d'un ordinateur va créer des données par son activité sur les réseaux. Les auteurs insistent sur le lien entre prolifération des ordinateurs personnels et début du fantasme cybernétique, soit l'auto-gouvernance de la société à travers le programme informatique. Selon Robert Wiener, mathématicien et penseur de la cybernétique, ce modèle fonctionnerait grâce aux systèmes munis d'entrées (input) et de sorties (output) influencés par des *feedbacks* (rétroaction). Ces derniers contrôlent l'information et la modifient en fonction de la réaction du système. Ainsi, le système est capable de s'autoréguler jusqu'à trouver la situation d'équilibre qui résoudra une situation.

COMMENT SONT COLLECTEES LES DONNEES ?

L'ère du Big Data dans laquelle nous sommes ne néglige aucune donnée. Les auteurs distinguent deux types de collectes de données : celles effectuées en ligne, et les données collectées hors-ligne.

Il n'est jamais aisé de savoir où, quand et comment sont prélevées les données. Un outil est primordial dans la collecte des données en ligne : le cookie. Il faut voir ce dernier comme un traceur que nous transportons avec nous à chaque voyage sur le web. Ce traceur va enregistrer les connexions, les liens sur lesquels on a cliqué, la navigation, etc. Ainsi, le cookie va pouvoir identifier les sites sur lesquels nous allons pour savoir ce qui nous intéresse et nous le proposer sous forme de publicité. En bref, ce véritable traceur qu'est le cookie va retenir la totalité de notre navigation pour en tirer les conclusions essentielles : qui sommes-nous, et qu'est ce nous que nous voulons ? Les auteurs précisent qu'il existe trois types de cookies : les first party cookie, soit les cookies propres à un site qui permettent de reconnaître un individu qui est déjà venu sur le site, les third-party cookies sont les cookies issus d'accords entre des sites différents. Concrètement, ce sont ces cookies qui permettent les publicités sponsorisées de marques sur le fil d'actualité de nos réseaux sociaux. Enfin, le dernier type de cookie est nommé "flash cookie". Les auteurs le différencient des autres par une résistance plus importante aux suppressions de cookie et permet aux sites web qui en emploient de connaître les usages hors-lignes des individus.

Quant aux collectes hors-ligne, celles-ci sont aussi orchestrées par les "flash-cookies" , mais surtout par ce qu'on appelle "l'internet des objets". Cet internet particulier représente l'interconnexion entre Internet et la multitude d'objets qui peuvent s'y connecter. Parmi ces objets, on retrouve les ordinateurs mais aussi les smartphones, les téléviseurs, les montres connectées, les enceintes et de plus en plus les maisons elle-même. Ainsi, un lien se crée entre le virtuel et le physique. C'est ce lien qui va permettre la collecte de données en piochant dans les différentes connexions entre l'Internet et les objets. Grâce à l'internet des objets, de plus en plus de données sont collectées chaque jour car le nombre d'objets connectés augmente fortement. D'ici 2020, l'Institut Gartner prévoit la connexion d'environ cinquante milliards d'objets sur Internet.

La collecte massive de données sur Internet peut paraître secrète voire illégale. En effet, un utilisateur n'a pas nécessairement conscience qu'en naviguant sur le web, des cookies peuvent collecter des informations sur lui-même. Or, les auteurs nous prouvent que ce n'est pas le cas. Cette collecte massive de données, nous l'approuvons en acceptant les cookies ou bien les Conditions Générales d'Utilisation (CGU) qui nous sont présentées de manière explicite. Cependant, par manque de temps, d'intérêt et même de connaissances sur le sujet, nous acceptons machinalement tout contrat nous séparant de l'utilisation d'un site web ou d'un réseau social.

À QUOI SERVENT LES DONNEES ?

Le Big Data en est à ses prémises et nous sommes loin d'avoir vu toute l'étendue de son potentiel. Les auteurs insistent tout le long du livre sur le fait que l'avenir du Big Data est un enjeu fondamental pour l'avenir de notre société car il a le pouvoir d'être aussi prometteur que destructeur. Pour enrichir ces propos, ils proposent quatre domaines dans lesquels le Big Data est ou peut devenir primordial : la santé, la fiscalité, la lutte contre le terrorisme et la transition écologique.

Le Big Data pourrait bouleverser le domaine de **la santé** en apportant ses qualités d'analyse et de traitement à la recherche scientifique et médicale. La revue *Nature* a notamment confirmé que les professionnels de la santé auront dans un futur proche bien plus de données à leur disposition que dans le monde de la finance. Cette quantité de données pourrait être un facteur non-négligeable dans la production de traitements contre diverses maladies. Les progrès de la médecine grâce au Big Data ne sont pas que futurs : les auteurs donnent notamment l'exemple d'une patiente atteinte d'une leucémie a été soignée grâce au logiciel Watson d'IBM qui a recoupé vingt millions de données d'autres patients pour conclure que la leucémie était atypique et demandait un traitement particulier pour être soignée.

En ce qui concerne le domaine de **la fiscalité**, là aussi le Big Data s'inscrit dans une démarche de recoupement d'informations pour lutter contre la fraude fiscale. Si bien des

choses peuvent être mises en place dans les prochaines années, le Big Data s'imisce déjà dans la lutte contre la fraude fiscale avec le logiciel Connect que les auteurs présentent comme un outil de choix permettant de déterminer les anomalies entre ce qu'un individu déclare aux impôts et la vie qu'il dit mener en analysant les photographies postées sur divers réseaux sociaux. Pour preuve, les auteurs précisent que 83% des enquêtes fiscales ouvertes au Royaume-Uni l'ont été grâce à ce logiciel.

Selon les auteurs, **la lutte contre le terrorisme** doit aussi passer par l'utilisation du Big Data. Aujourd'hui, des logiciels de *data mining* (exploration des données) localisent et travaillent à la détection de la propagation des influences radicales.

Enfin, après avoir évoqué l'urgence que représente la nécessité d'effectuer une **transition écologique**, les auteurs rappellent aux acteurs de ce changement peuvent aussi compter sur le Big Data. Il est déjà utilisé de manière préventive en Asie pour détecter les éventuels séismes mais il s'inscrit aussi dans une transformation de la ville via le fantasme de la Smart City, cette ville totalement connectée et soumise à la cybernétique pour évoluer.

LES DATAS AU 21^{EME} SIÈCLE

LE COUP DATA SUR LA GOUVERNANCE DU MONDE

Le "coup data" est une expression des auteurs qui met en avant les problématiques de gouvernance que pose le nouveau paradigme créé par les algorithmes et le Big Data. Les auteurs précisent que ce "coup data" n'est possible que sur fond de crise politique et de l'enfoncement dans la "spirale du tout-sécuritaire".

Cette expression marque également la rapidité de la modification du paradigme par les données. Il existe différentes définitions de la gouvernance. Nous retiendrons celle de la Banque Mondiale qui aide à particulièrement bien comprendre le propos des auteurs. La gouvernance recouvre les normes, traditions et institutions à travers lesquelles un pays exerce son autorité sur le bien commun, dans une optique de développement. Il s'agit de comprendre tout ce que modifient les données.

Dans un premier temps, les auteurs montrent la modification des normes. Les géants technologiques commencent à prendre la place de régulateurs privés (voir affaire Apple contre FBI). Cela menace le débat démocratique et montre la puissance de l'utilisation habile des données face aux gouvernements. Les auteurs parlent d'un affrontement états-plateforme (les géants technologiques) contre les états-nations (les gouvernements).

Les auteurs alertent également sur la puissance des données face à une potentielle éditorialisation de la culture. Les auteurs proposent d'observer comment Google Book tend à devenir la bibliothèque officielle de l'humanité. De cette manière, ce géant du

numérique collecte et choisi quels peuvent être les ouvrages facilement accessibles en ligne et lesquels ne le seront pas. Qui et qu'est-ce qui guide ces choix ? Le secret industriel protège l'accès à ces réponses.

Cette problématique touche le concept de soft-power. Une culture éditorialisée et mondialement partagée rend cet arme d'influence vaine. Et pis encore, elle marque l'hégémonie d'une seule culture.

La domination des territoires est également menacée par les données. Cela fait écho à la matérialité d'internet. Cela revient à poser ces questions : où sont les serveurs, qui les possède, comment est réglementé l'accès à ces serveurs, où sont les câbles souterrains qui connectent les espaces entre eux, quels espaces sont connectés ? Voici tant de question qui montre la réalité d'internet et met en exergue son rôle géopolitique majeur. Les auteurs mettent en avant le projet Internet.org. Il s'agit d'un projet de développement du continent africain mené par Facebook avec des associés économiques de la télécommunication. Ce modèle de développement, marqué par un choix éditorial très restreint, comparable à une tablette munie d'une sécurité pour les enfants, pose des questions de légitimité, des intentions réelles des acteurs privés qui asservissent d'une certaine manière ce continent.

Pour conclure, on peut noter que le Danemark a créé la fonction de "diplomate pour le numérique", le 26 mai 2017, reconnaissant ainsi sur la scène internationale la puissance et le pouvoir de ces géants dans la sphère géopolitique mondiale. C'est une prise de pouvoir technologique : c'est le coup data.

LES MODES DE REGULATION

L'avancée la plus notable à l'heure actuelle en terme de régulation est le RGPD. Le Règlement Général sur la Protection des Données, mis en vigueur le 25 mai 2018, est un outil de régulation propulsé par l'Union Européenne. Il a pour ambition d'harmoniser la gestion et la protection des données personnelles des personnes résidentes dans l'Union Européenne.

Il s'applique à tous les acteurs économiques et sociaux proposant des biens et des services sur le marché de l'Union Européenne dès lors que leur activité traite des données personnelles des résidents de l'Union Européenne. Il peut s'agir d'entreprises, d'associations, d'organisations de marché public, voire même d'entreprises extérieures à l'Union Européenne qui traitent les données des résidents européens.

L'objectif est de donner aux citoyens européens davantage de contrôle et de visibilité sur leurs données privées. Cela passe par une obligation des collecteurs de données de faire savoir aux usagers quelles sont les données collectées, à quelle fin et quelles sont leurs durées de conservation.

L'enjeu pour les acteurs économiques et sociaux concernés est de savoir à un instant donné où sont les données et pouvoir sur simple demande les collecter et les transmettre à l'individu qui en fait la requête. Il faut noter que des sanctions sont prévues en cas de non-conformité. Cela peut être aller jusqu'à 20 millions d'euros d'amende ou 4% du chiffre d'affaire annuel mondial pour les multinationales à cela s'ajoute également des dommages et intérêts aux personnes pour préjudice subi dû au non-respect du règlement.

On peut lister cinq principes fondateur de ce texte, qui aide à comprendre sa conception.

1. La logique de responsabilisation (**Accountability**) impose aux entreprises de prendre toutes les mesures pour garantir la conformité de leur activité au RGPD. Il s'agit également pour elles d'être capable de prouver la mise en place de toutes les mesures pour tendre vers une protection maximum des données
2. Le principe de **privacy by design** suppose que la protection des données personnelles doit être prise en compte dès la conception du produit ou du service
3. Le principe de sécurité par défaut (**Security by default**) renforce le rôle de la sécurité dans le système d'information (celui-ci doit être sécurisé à ses différents niveaux, du physique jusqu'au logique avec par exemple des contrôles d'accès)
4. La nomination d'un DPO, **Data protection officer** : il s'agit d'un délégué à la protection des données. Il doit être associé aux différentes questions et problématiques de protection des données à caractère personnelle de l'entreprise. Il est également le point de contact avec les autorités de contrôle
5. Réalisation d'études d'impact sur la protection des données personnelles avant la mise en oeuvre de nouveaux traitements de données qui pourraient potentiellement porter atteinte aux données personnelles et aux principes de liberté

"QUELLE SOCIETE VOULONS-NOUS DEMAIN ?"

"Quelle société voulons-nous demain ?" voici l'interrogation centrale. Pour les auteurs cela commence par appréhender le numérique comme un fait social pour mieux l'accompagner. Il s'agit pour cela d'hybrider les connaissances pour une meilleure formation des générations futures. Il faut pour cela développer une culture du numérique au sein de l'Union Européenne. Les auteurs veulent que l'habileté numérique (internet literacy) soit enseignée. La formation aux médias numériques accessible au plus grand nombre est essentielle.

Les auteurs prônent également un apport de l'éthique dans le numérique. Cela peut prendre la forme d'un institut pluridisciplinaire de recherches. Ils ajoutent que les réflexions peuvent être nourries des réflexions éthiques autour du nucléaire, en ce qu'elles ont soulevés des enjeux de productions et de sécurité similaire.

Le nouveau paradigme de l'empire des données peut être très efficace pour les sociétés mais doit être cadré rapidement. Il est urgent d'affirmer l'humanisme, de se recentrer autour du droit des individus. Les auteurs appellent à un débat de la société, pour réfléchir et repenser en profondeur le système de gouvernance des données. Une gouvernance partagée serait, pour eux, la plus efficace. Ils ont écrit : *“les États, de préférences réunis, les plateformes et, enfin, les milliards d'utilisateurs, devront être associés, directement ou indirectement, à un système de droit durable dans le temps et dont le maître mot sera la confiance”*.

UN DEBAT NAISSANT

La lecture de “L'Empire des données” redonne un souffle d'espoir là où la croyance principale décrit un monde submergé et asservi par les algorithmes. Parfois très technique (la partie traitant du RGPD est certainement réservée aux experts) l'ouvrage s'accorde autour d'une note optimiste. Les auteurs ne veulent pas apporter des solutions toutes faites, mais militent pour un réel débat citoyen autour de ces questions aussi cruciales qu'urgentes. Le ton est volontairement naïf et les préconisations relativement générales et peu nombreuses. Cependant cet ouvrage ouvre la réflexion à la réalité des moyens d'actions et permet tout un chacun de prendre part au débat.