

Big Data et plateformes

La nouvelle économie des données

+ EN LIGNE

10 études de cas
complémentaires



20 études
de cas :
Uber, Tinder,
TikTok,
Netflix...

Big Data et plateformes

La nouvelle économie des données

Ressources numériques : 10 études de cas

1. LinkedIn,	p. 140	6. Adyen,	p. 189
2. Nexar,	p. 153	7. Square,	p. 189
3. Tesla,	p. 153	8. Datadog,	p. 217
4. Inato,	p. 175	9. Palantir,	p. 217
5. inovie,	p. 175	10. The Trade Desk,	p. 229

Repérez les ressources numériques dans votre livre



lienmini.fr/ressourcesnum-dbs

Accédez directement à votre ressource :

Flashez le code avec votre
téléphone ou votre tablette



OU

Tapez l'URL
dans votre navigateur



Pour toute information sur notre fonds et les nouveautés dans votre domaine de spécialisation, consultez notre site web : www.deboecksuperieur.com

© De Boeck Supérieur s.a., 2022
Rue du Bosquet 7, B-1348 Louvain-la-Neuve

Tous droits réservés pour tous pays.

Il est interdit, sauf accord préalable et écrit de l'éditeur, de reproduire (notamment par photocopie) partiellement ou totalement le présent ouvrage, de le stocker dans une banque de données ou de le communiquer au public, sous quelque forme et de quelque manière que ce soit.

Dépôt légal :

Bibliothèque nationale, Paris : octobre 2022

Bibliothèque royale de Belgique, Bruxelles : 2022/13647/158

ISSN 1782-8147

ISBN 978-2-8073-4815-8

Pour Aurélie

PRÉFACE

Au cours de mes différentes aventures entrepreneuriales, associatives ou dans l'administration d'entreprises cotées, de Montpellier à San Francisco, au sein de Bime Analytics, France Digitale et Sweep, deux évidences se sont imposées à moi : la première est que la meilleure façon de changer le monde c'est d'entreprendre et la deuxième c'est que les données constituent aujourd'hui l'un des actifs les plus précieux d'une entreprise.

L'impact – bon ou mauvais ! – qu'une entreprise peut avoir est proportionnel à sa prospérité, qui est elle-même intimement liée à sa capacité à analyser en quasi temps réel les différents flux de données qui l'irriguent. Ces flux sont une inépuisable source d'informations pour offrir plus aux clients, mais aussi aux partenaires et aux collaborateurs.

Malheureusement, ces données sont souvent mal interprétées par les entreprises qui s'exposent alors au risque de prendre des décisions mal informées les éloignant de leur objectif voire de leur mission première. Les consommateurs exigent, plus que jamais, une expérience client connectée et optimale. Les entreprises agiles doivent utiliser les données en provenance de ces mêmes clients, quels que soient les supports numériques qu'ils utilisent, pour identifier les problèmes et surtout adapter leurs offres en conséquence. Le challenge que les entreprises doivent relever est dorénavant de disposer d'une vue holistique de leur clientèle et de leurs prospects pour se développer, opérer dans le cadre de cette mission et maximiser leur impact sur le monde. En somme, durer pour toujours.

Ce but ultime est partagé par toute nouvelle entreprise et il est difficilement atteignable dans un monde devenu numérique et ultra-concurrentiel sans l'analyse des données et sans la technologie pour éclairer les processus décisionnels.

Les volumes colossaux de données qui sont traitées par les entreprises modernes doivent être considérés comme des actifs et non comme des sources de problèmes. Les analyser peut être une tâche particulièrement complexe, il est par conséquent important de ne pas se laisser dépasser par le volume, mais d'essayer d'identifier les données susceptibles de générer de la valeur. Les laisser en sommeil revient à limiter le champ des opportunités et le développement de l'entreprise.

Certes, il n'existe pas de manuel universel pour expliquer comment les analyser, mais avec les bons experts et les bons outils, ces données peuvent devenir l'avantage concurrentiel le plus précieux d'une entreprise.

Les mégadonnées produites par les nouvelles technologies de l'information irriguent d'ores et déjà nos économies et nos sociétés. Les gains et les enjeux commerciaux sont gigantesques ; les algorithmes d'Amazon et de Facebook anticipent et façonnent nos actions dans une mesure qui était impensable il y a dix ans de cela.

Les données sont également un levier d'innovation et d'orchestration de l'action collective comme l'a prouvé par exemple l'ouverture des fichiers de certains réseaux de transports urbains pour nous aider à mieux nous déplacer et à trouver des alternatives plus écologiques. Autre exemple, l'ouverture de la base des valeurs foncières (DVF) en 2019 a permis la création de nombreuses entreprises et de nombreux services innovants autour de l'analyse des prix de l'immobilier.

Les données sont aussi un moyen efficace pour restaurer la confiance dans l'action publique. À cet égard, la crise sanitaire que nous avons traversée a provoqué dans le pays une prise de conscience vis-à-vis de l'information. Les difficultés pour établir des statistiques épidémiques en temps réel, les conflits d'interprétation des courbes, la fragilité des hypothèses des modèles épidémiologiques, tous ces éléments auront eu finalement une vertu pédagogique : celle de révéler que toute donnée est une construction et qu'elle doit être expliquée. Ouvrir l'accès aux données, c'est enrichir le débat public, cette transparence est le meilleur remède à la défiance et au complotisme. La crise a aussi montré que gouverner par les données nécessite de plus en plus souvent de puiser dans celles détenues par des acteurs privés, sans lesquels prendre le pouls du pays ne serait pas possible. Par exemple, CovidTracker propose des visualisations et des analyses de la situation épidémique ville par ville, avec Google pour la fréquentation des lieux ou Orange pour les déplacements à travers le territoire.

Cependant, les données ouvertes doivent passer à la vitesse supérieure et devenir bien plus fiables et plus qualitatives. Pour ce faire, il faut notamment les documenter (ce qui n'est pas toujours le cas), définir des standards interopérables et disposer de métadonnées plus homogènes. L'enjeu de qualité est crucial pour le développement de l'intelligence artificielle, qui peut aussi bien se nourrir de données publiques que privées, mais qui doivent dans tous les cas être abondantes et intègres si l'on souhaite de meilleures prises de décisions et créer des systèmes plus vertueux.

Le système vertueux le plus menacé aujourd'hui est notre stabilité climatique. Les scientifiques du GIEC ont depuis longtemps compris la trajectoire alarmante du réchauffement climatique grâce à l'analyse de données. Celles-ci sont cruciales pour orchestrer les indispensables efforts de réduction de notre empreinte carbone. Il nous faut comprendre les interactions entre les écosystèmes émetteurs de gaz à effet de serre à l'échelle mondiale. Il faut espérer que l'analyse de ces données permette la coordination de l'action collective en commençant par l'identification des plus grands émetteurs qui sont les grands agents du changement ! Le temps presse pour éviter davantage de phénomènes extrêmes (vagues de chaleur, incendies, inondations, tempêtes tropicales, etc.). Sans technologie numérique pour relier entre eux les écosystèmes de données porteurs d'informations pertinentes pour comprendre notre hypercroissance insoutenable et faire « mieux » ou « moins » aux bons endroits, nous nous éloignerons de la solution.

Comment faire ?

Il faut d'abord mesurer nos émissions, et sur ce plan le marché de la gestion de l'empreinte carbone évolue et mûrit très rapidement. Nous assistons à un engouement qui va de pair avec l'obligation pour les entreprises de publier des rapports sur leurs émissions carbone. L'utilisation d'outils dédiés leur permet d'aller au-delà des simples intentions et de faire état des progrès en termes de réduction des émissions.

Les risques juridiques et réputationnels qui pèsent sur les entreprises font qu'il est devenu nécessaire de rendre compte en continu. L'exercice qui consiste à piocher dans les systèmes d'information internes et externes est cependant très complexe à l'échelle des organismes vivants interconnectés qu'elles représentent, elles et leurs chaînes de valeurs (qui émettent du CO₂ en symbiose !).

La société Sweep que j'ai cocréée développe précisément une offre numérique qui permet l'orchestration et le pilotage de la transition pour une réduction à grande échelle des émissions de carbone. L'ampleur de leurs émissions ne laisse pas le choix aux entreprises : elles se doivent de mettre en place une stratégie technologique pour embarquer leurs chaînes de valeur dans la préservation du bien commun qu'est la stabilité climatique et se préparer à exercer dans une économie bas carbone.

Sweep agrège plus de 30 000 indicateurs pour calculer l'empreinte carbone en tenant compte des particularités géographiques de chaque implantation, du type de mix énergétique et du type de production concerné. La déclaration des émissions directes n'est que la partie émergée de l'iceberg : Sweep relie les grandes entreprises à leur réseau de partenaires commerciaux, à leurs équipes ou à leurs fournisseurs. Cette entreprise est la suite logique de mon parcours dans ce milieu des mégadonnées, c'est une manière pour moi de combattre avec mes quelques armes d'entrepreneure dans ce grand défi climatique.

Dans ce milieu d'innovation technologique, les initiatives américaines et la Silicon Valley sont toujours mises en avant. Cependant, je pense que la France a absolument une carte à jouer de par son engagement historique dans cette lutte depuis les accords de Paris de 2015, et même avant cela : on l'ignore souvent, mais c'est le français Jean-Marc Jancovici, figure emblématique de la vulgarisation de l'action climat, qui inventa le bilan carbone. Autre exemple : aucun autre pays ne dispose d'une institution publique dédiée aussi ancienne que l'Ademe ni d'un pool de talents d'ingénieurs formés sur le sujet du carbone ou du cycle de l'eau. La France inspire sur ces sujets et est observée par les autres pays. Tout ceci, couplé à l'écosystème français du numérique qui a beaucoup progressé depuis dix ans, à l'excellence des formations, à la qualité de la communauté des investisseurs et à l'internationalisation, fait que la France pourrait bien produire les grands acteurs économiques de demain.

Alors que le monde s'engage dans des projets ambitieux pour atteindre les Objectifs de développement durable (ODD) de l'ONU, il est urgent de mobiliser la révolution des données pour tous et pour la planète entière afin de poursuivre les progrès réalisés, de responsabiliser les gouvernements et de favoriser le développement durable. Une information plus diversifiée, intégrée, opportune et digne de confiance, peut mener à

de meilleures prises de décision et à une implication en temps réel des citoyens. Cela permet aux individus, aux institutions publiques et privées et aux entreprises de faire des choix qui sont bons pour eux et pour le monde dans lequel ils vivent.

L'excellent livre de Lionel Melka que vous tenez entre les mains présente de manière claire et illustrée de nombreux cas pratiques et nous explique cette nouvelle ère de la mise en données de nos vies et de nos sociétés, fruit et facteur de la transition numérique. Comme lui, je pense que les défis soulevés par la mise en données du monde ne doivent pas être considérés comme des questions uniquement technoscientifiques. Ils sont – à l'instar de toutes les innovations majeures – politiques et éthiques et même salvateurs au regard du défi climatique actuel.

La question du consentement, du contrôle, mais aussi de la confiance et donc des modalités d'établissement des pratiques éthiques ou socialement acceptables, est au cœur du futur des mégadonnées, et c'est la condition de leur survie. Cependant, des règles trop restrictives entraîneraient la fermeture de nombreux axes de recherche qui permettent de mieux comprendre le fonctionnement des systèmes complexes que sont les sociétés humaines. La réglementation doit être mise en place par des débats démocratiques et informés, dans lesquels chacun peut exposer ses idées et ses contraintes.

Le développement à grande échelle de la culture des données donne aux citoyens les outils nécessaires au plein exercice de leur rôle et à la poursuite de leurs objectifs dans un monde qu'ils comprennent et dont ils appréhendent les opportunités et les menaces.

Remettre en question le fonctionnement des systèmes et des zones de pouvoir vertical actuels (les grandes entreprises, les États-nations) serait presque à portée de clic.

Pour le meilleur selon moi.

Rachel Delacour

INTRODUCTION

Les données sont devenues centrales et essentielles pour un nombre croissant de secteurs économiques contemporains. Les nouvelles technologies, les infrastructures, la finance, l'assurance et l'énergie traitent désormais les données comme une forme d'actif. Jusqu'à récemment, les entreprises supprimaient les données ou choisissaient de ne pas les collecter parce que payer pour le stockage ne semblait pas être un bon investissement. Aujourd'hui, elles en recueillent autant qu'elles le peuvent.

En effet, dans « l'économie numérique », supprimer des données en raison des coûts de stockage reviendrait à brûler des sacs de billets ou à jeter des barils de pétrole parce que louer un entrepôt serait trop cher. Cet élan poussant les entreprises vers la collecte de données s'explique autant par les perspectives florissantes de l'accumulation de ces dernières que par la baisse des coûts de stockage. En d'autres termes, non seulement les barils de pétrole prennent de la valeur, mais la location des entrepôts baisse également. Si les données ne sont pas directement synonymes de profit, elles relèvent d'une logique similaire. Les entreprises ayant toujours été naturellement motivées par le profit, on doit s'attendre maintenant à ce qu'elles soient motivées par les données dont l'accumulation est à l'origine de nouvelles manières de réaliser et de gérer des affaires.

En quelques années à peine, les géants de l'Internet et les réseaux sociaux ont transformé de fond en comble la manière dont l'information est produite, circule, est partagée. Le capitalisme s'est restructuré autour du numérique qui est un réservoir de nouveaux concepts, d'innovations réelles ou potentielles, et qui repose sur l'extraction des données. Elles sont devenues le nouvel or noir du capitalisme, dont la source est l'activité des utilisateurs de services numériques.

Les données sont désormais considérées comme une nouvelle matière première dans l'économie numérique. Dans la société post-industrielle, la richesse ne vient plus de la machine, comme c'était le cas dans la société industrielle, mais de l'information et des données. Elles sont valorisées dans un processus en quatre étapes : génération, qualification, traitement et médiation du traitement. Depuis quelques années, des modèles économiques entiers se sont construits sur l'exploitation des données.

L'essor des mégadonnées (*big data* en anglais), au cours de la dernière décennie et l'émergence plus récente de l'intelligence artificielle (IA) ont suscité de nombreux espoirs et, de plus en plus de craintes quant au sort de l'humanité dans la « quatrième révolution industrielle ». Nous dirigeons-nous vers des temps plus clairs ou plus

sombres ? Les mégadonnées et l'intelligence artificielle représentent-elles des menaces existentielles pour la démocratie ou offrent-elles la possibilité de construire un avenir où les décisions seront plus rationnelles, les politiques plus efficaces, les processus plus justes, les politiciens plus responsables ?

Dans l'ensemble, la révolution des données, les mégadonnées et l'intelligence artificielle actuelles fonctionnent sur la base des données personnelles émises par les personnes qui utilisent des appareils et des services numériques dans le cadre de leurs actions et interactions quotidiennes. Il en résulte de vastes ensembles de données, qui peuvent ensuite être analysés par des algorithmes pour dévoiler des modèles et des corrélations et établir des estimations, des projections, des prédictions et des prescriptions, entre autres.

La plupart d'entre nous s'appuient déjà sur ces outils pour décider des routes à emprunter, des articles à lire, des vêtements à acheter, des contenus à aimer, des vols à réserver ou des personnes à contacter. Nos médecins utiliseront bientôt les mêmes types d'outils pour diagnostiquer le cancer et suggérer des plans de traitement. Il y a, et il y aura, de véritables « robots » qui feront tourner nos usines, effectueront les tâches ménagères et amuseront nos enfants, mais, d'une manière générale, les intelligences artificielles actuelles et futures sont ce que l'on appelait il y a quelques années les mégadonnées : des modèles d'analyse informatique alimentés et entraînés sur de grandes quantités de données traitées par des machines (ordinateurs) pour atteindre un objectif et, dans certains cas, alimenter des machines sophistiquées (robots).

Les applications et les implications de ce phénomène sont déjà considérables. Cette personne va-t-elle aimer ce livre parce que quelqu'un comme elle (y compris le mois dernier) l'a aimé ? Cet adolescent est-il sur le point d'abandonner l'école ? Doit-il ou doit-elle obtenir un prêt ? La voiture sans conducteur doit-elle tuer une femme enceinte ou cinq personnes âgées si elle n'a pas d'autre choix que de les écraser ? Plusieurs questions connexes difficiles viennent à l'esprit et alimentent les débats en cours. Si les algorithmes semblent racistes, est-ce parce que leurs développeurs intègrent leurs préjugés ou plutôt parce que les prédictions répètent les préjugés passés ? Que se passe-t-il lorsque l'algorithme rencontre des cas qu'il n'a jamais vus auparavant (un chien au visage plat ou un humain à la peau plus foncée que dans l'ensemble de données sur lequel il a été entraîné) ?

Fondamentalement, comment ces estimations, prédictions et prescriptions doivent-elles être utilisées, et par qui, quand et si elles le sont ? Ces risques sont réels. Ils doivent être connus et traités pour limiter les pires effets secondaires typiques de l'évolution technologique, au moins à court terme, notamment le creusement des inégalités. Mais les mégadonnées et les intelligences artificielles ne sont pas de la magie noire, pas plus que les algorithmes qui les font fonctionner ne sont de véritables boîtes noires. Compte tenu de leur omniprésence et de leur puissance, il est important de comprendre leur fonctionnement et les enseignements que nous pouvons en tirer pour promouvoir un changement social positif. D'un point de vue critique, il ne s'agit pas (seulement) d'utiliser l'intelligence artificielle pour optimiser les chaînes d'approvisionnement (notamment), ce qui continuera d'avoir un impact majeur sur les sociétés

et les économies, mais d'être inspiré et soutenu par celle-ci pour améliorer les systèmes humains.

Le présent ouvrage traite de cette révolution en marche et est structuré autour de trois parties :

- la première partie explique la montée en puissance des données dans le monde, comment ce mouvement a été mis en œuvre, quels sont les enjeux des politiques publiques dans la collecte et l'utilisation des données et quelles sont les questions de souveraineté numérique et de contrôle des concentrations ;
- la seconde partie passe en revue les modes de création de valeur à partir des données : la pyramide DIKW, les différents modèles économiques, la monétisation et l'essor des plateformes ;
- la troisième partie s'intéresse à une dizaine de secteurs qui ont été bouleversés par l'exploitation massive des données et présente une trentaine de cas pratiques.

CHAPITRE 1

L'ÉMERGENCE D'UNE ÉCONOMIE DES DONNÉES

La plupart des entreprises ne sont pas à l'origine conçues pour utiliser des données, mais pour produire des marchandises. Leur organisation opérationnelle entraîne par exemple une perte considérable d'informations sur l'acheteur final ou sur l'utilisation du produit. Pour mieux extraire les données, un nouveau type d'entreprise s'est développé : la plateforme. Ce nouveau modèle cherche à produire des données exploitables *via* des outils ou une infrastructure dédiée. Les plateformes sont en grande partie des intermédiaires ou des places de marché qui permettent aux individus et aux entreprises de partager des informations, d'acheter, de vendre, ou d'accéder à divers biens et services, en rassemblant le tout en un seul endroit où les données peuvent interagir les unes avec les autres. Elles reposent sur des effets de réseau qui favorisent le volume et créent de la valeur économique bien au-delà de ce qui était observable il y a quelques années dans les entreprises conventionnelles : plus il y a de participants, de fonctions, de contenus ou de services numériques disponibles, plus la plateforme devient utile. Les plateformes produisent très rapidement des monopoles. Contrairement aux entreprises qui cherchent à ne conserver que les compétences essentielles au cœur de leur projet, les plateformes démultiplient les activités rentables et non rentables pour diversifier les sources de données. Google, par exemple, fournit un ensemble de services gratuits pour capter un maximum de données. Enfin, elles se présentent souvent comme des espaces ouverts, alors que les services sont entièrement décidés par leurs propriétaires, ou comme des intermédiaires, mais ce sont elles qui dictent les règles du jeu !

Grâce à des technologies comme l'Internet des objets, les plateformes en ligne et l'analyse des données, la liste des produits et des services numériques s'allonge à un rythme rapide et transforme le monde des affaires. En effet, les systèmes génèrent des volumes de données sans précédent lesquelles ont de plus en plus de valeur. De la « maison intelligente » à la « ville intelligente », en passant par la finance, la gouvernance, la production et la distribution, l'électronique grand public et les systèmes d'information en entreprise, etc., les données sont un capital sans lequel nombre de technologies et d'entreprises ne pourraient pas fonctionner.

La littérature universitaire sur les dimensions sociales, politiques et économiques des données traite ces dernières comme une marchandise. Que ce soit de manière implicite ou explicite, les analyses, tant dans le milieu universitaire que dans les médias, considèrent généralement ce cadre analytique comme acquis. Pourtant, la distinction

entre capital et marchandise est importante. En considérant les données comme une forme de capital, il est possible de mieux analyser la nature et la dynamique de l'économie numérique. Plutôt que de voir la collecte de données comme un simple moyen de produire et d'obtenir des marchandises qui sont d'une manière ou d'une autre converties en valeur monétaire, les données deviennent un régime économique politique dirigé par une logique d'accumulation et de circulation perpétuelles du capital (en l'occurrence, les données).

Aux États-Unis, en Europe et en Chine, la question des données est au cœur des enjeux des politiques publiques. Aux États-Unis, Google a récemment annoncé que d'ici deux ans ses téléphones et ses tablettes Android allaient limiter le partage des données personnelles récoltées par les applications. En d'autres termes, Google va réduire la transmission de ces données à des tiers, notamment aux annonceurs publicitaires, ou entre les applications elles-mêmes, le tout en faveur d'un nouvel équilibre entre confidentialité, publicité et gratuité.

Installé sur 70 % des téléphones, Android est le système d'exploitation le plus utilisé au monde. Il se développe chez Google grâce à la collecte de données personnelles qui est le carburant de l'économie numérique. Mais alors pourquoi Google promet-il de limiter leur collecte ?

En réalité, l'entreprise ne fait qu'emboîter le pas d'Apple qui avait l'an dernier obligé les éditeurs d'application à demander aux utilisateurs la permission de suivre leur navigation sur Internet et de récolter leurs données. Six mois plus tard, 80 % ont refusé, provoquant un manque à gagner de près de dix milliards de dollars pour Snapchat, Facebook, Twitter et YouTube, d'après les estimations du *Financial Times*.

Facebook est l'acteur qui perd le plus dans cette guerre des recettes publicitaires, car il s'agit du réseau social qui en dépend le plus. Début 2022 déjà, alors que le groupe s'effondrait en Bourse, Mark Zuckerberg estimait qu'il allait perdre dix milliards de dollars en 2022 à cause d'Apple et de son parc mondial de 25 % de téléphones portables. Si l'on considère que 70 % de téléphones dans le monde tournent sous Android, la décision de Google pourrait faire grimper les pertes de Facebook à 30 milliards de dollars.

1. LES DONNÉES ET LE NOUVEAU CAPITALISME

Il existe aujourd'hui toute une série de termes qui font référence aux relations politico-économiques entre les données et le capitalisme, notamment « capitalisme de surveillance » (Zuboff, 2015), « capitalisme informationnel » (Fuchs, 2010), « capitalisme communicationnel » (Dean, 2005), « capitalisme de plateforme » (Srniczek, 2016) et « iCapitalisme » (Duff, 2016). Ces différents termes ne sont pas interchangeables, mais ils partagent des thèmes et des conclusions communs.

Trois grandes idées se dégagent dans la littérature académique sur le sujet :

1. les données sont précieuses et créatrices de valeur (Arvidsson, 2016 ; Roderick, 2014 ; Srniczek, 2016) ;

2. la collecte de données a une forte influence sur le comportement des entreprises et des gouvernements (Fourcade et Healy, 2017 ; Zuboff, 2016) ;
3. les systèmes de traitement de données sont truffés de relations d'iniquité, d'extraction et d'exploitation (Aitken, 2017 ; Andrejevic, 2014 ; Fourcade et Healy, 2013 ; Poon, 2016 ; Thatcher *et al.*, 2016).

Fourcade et Healy (2017) indiquent que les « entreprises modernes » sont désormais guidées par un « impératif de données » qui exige l'extraction de toutes les données, de toutes les sources, par tous les moyens possibles. Le stockage et l'étude des activités quotidiennes des individus, même celles qui semblent banales, sont devenus la norme plutôt que l'exception. Pour satisfaire cet impératif, on ne peut pas se contenter de collecter les données, il faut les créer (IBM, 2014).

Cela implique une surveillance des personnes, des lieux, des processus, des choses et des relations entre ces derniers (van Dijck, 2014) et une collecte exhaustive des données. Cisco, l'une des entreprises qui construisent ce système global, appelle ça « l'Internet de tout ». De même, IBM affirme que « tout est fait de données de nos jours ».

Quelles sont les implications d'une vision du monde qui considère que tout est fait de données ? Il ne s'agit pas d'une simple observation neutre sur la nature ou la substance du monde. De telles affirmations ne se contentent pas de révéler ou de refléter le monde. Elles l'ordonnent et le construisent. En opérant de manière rhétorique, elles changent la manière dont nous comprenons le monde et interagissons avec lui et placent ceux qui disposent d'un capital de données en position d'autorité.

Cette vision soutient que l'accumulation et l'utilisation des données constituent non seulement un fait, mais qu'elles deviennent une logique motrice qui influence le comportement. Il s'agit d'une relation pouvoir/connaissance : connaître le monde, c'est exercer un pouvoir sur lui et exercer un pouvoir, c'est le connaître dans toutes ses caractéristiques, c'est le classer en catégories et en normes, c'est le rendre lisible et observable, c'est exclure les autres mesures et méthodes de connaissance.

L'analyse de données est un nom trompeur. Un terme plus approprié serait fabrication de données. Les données ne sont pas là à attendre d'être découvertes comme le pétrole et les minerais. Les données sont une abstraction du monde, enregistrée, créée et valorisée par des individus aidés par la technologie. La vision des données comme ressource naturelle omniprésente et gratuite renforce les régimes d'accumulation.

Une vidéo de 2014 de Siemens, un grand fabricant industriel, illustre cette logique : « Nous vivons dans un univers de données qui gagne non seulement en volume, mais aussi en importance, chaque jour. La question de savoir comment en tirer de la valeur commerciale devient de plus en plus essentielle pour nous. Nous devons comprendre que les données sont partout et qu'elles sont créées chaque seconde. Nous devons voir les données comme un atout et les transformer en valeur. »

Ce n'est pas une coïncidence si les données sont traitées comme une substance universelle au moment où il y a tant à gagner pour celui qui peut les revendiquer et les extraire de toutes les sources. En effet, c'est une boucle qui s'autoalimente : de nombreux

systèmes de contrôle reposent sur la collecte et le traitement constants de données et ces systèmes permettent à leur tour d'en générer encore plus. Les flux de données correspondent aux flux de pouvoir et de profit et c'est une alchimie qui promet de produire des réserves infinies de l'un comme de l'autre. Mathias Herberts, pionnier de l'implémentation des tout premiers clusters Hadoop en France, affirme que l'accumulation et la capacité à exploiter les données font naître une nouvelle forme de concurrence capitaliste plus juste et plus saine que jamais, car elle se fonde avant tout sur des compétences techniques et des choix stratégiques et non sur le favoritisme d'État ou le clientélisme.

L'objectif de tout transformer en données et la recherche de nouvelles sources de données font écho aux modes d'accumulation impérialiste : lorsque le capitalisme est confronté à des crises d'accumulation, il est nécessaire de trouver de nouvelles sources de valeur et de nouveaux lieux pour écouler les marchandises. Les anciennes stratégies d'accumulation sont réutilisées dans les nouveaux espaces et de nouvelles stratégies sont élaborées par essais et erreurs, dans une quête sans fin pour dépasser ou déplacer les contradictions internes qui mènent à la crise. Cela peut signifier soumettre à la logique du capitalisme des parties de la vie qui n'étaient auparavant ni commercialisées ni monétisées, ou coloniser de nouveaux territoires afin d'en faire de nouveaux sites d'extraction. On observe cette dynamique de « colonialisme des données » lorsque Facebook et Google, par exemple, s'installent dans des territoires comme l'Inde et l'Afrique.

Elles le font sous le prétexte d'apporter des services gratuits qui permettent aux personnes marginalisées d'accéder à l'Internet, mais il existe des contreparties évidentes à cette démarche à première vue louable. En réalité, ces entreprises en tirent un avantage significatif en ouvrant des marchés, en attirant de nouveaux publics sur leurs plateformes, et en exploitant leurs données : « si c'est gratuit, c'est que vous êtes le produit » selon l'adage. Ces nouveaux lieux et ces nouvelles populations constituent de nouvelles opportunités d'accumulation. Les mêmes stratégies impérialistes sont rejouées aujourd'hui, mais actualisées à l'ère numérique. Un corpus croissant de recherche sur les études critiques des données a montré comment leur production, leur distribution et leur utilisation se situent dans une économie politique émergente qui a de très fortes implications dans toute la société : de la restructuration des villes et de l'État, au (re) développement des infrastructures électriques et informatiques.

Quelle est la forme économique des données ? Comment la valeur peut-elle en être dérivée ? Pourquoi leur collecte est-elle importante ? Selon moi, le fait de poser ces questions permet de recadrer de manière productive la façon dont nous comprenons la forme et la dynamique des données.

Le « stratège des mégadonnées » d'Oracle, l'un des plus grands éditeurs de logiciels au monde, déclare : « Les données constituent en fait un nouveau type de capital, au même titre que le capital financier, pour la création de nouveaux produits et services. » Et ce n'est pas qu'une métaphore ; les données répondent à la définition littérale du capital telle qu'on la trouve dans les manuels scolaires. Cette déclaration montre un changement politique et économique dans lequel les données sont créées, collectées et circulent en tant que capital.

Comme la finance, elles sont désormais considérées comme un moteur de croissance. Si les entreprises de la finance sont libres de faire la navette entre les capitaux d'un pays à l'autre, les entreprises de la technologie doivent être libres de stocker et de vendre des données là où elles le souhaitent. Cela signifie, par exemple, qu'une entreprise peut collecter des informations personnelles concernant des Américains, les stocker à Taïwan et les vendre en Europe.

L'accent mis sur les données ne doit pas être considéré comme obérant la financiarisation, cela crée de nouvelles sources de valeur et de nouveaux outils d'accumulation. Il y a dans l'histoire une longue liste de croisements entre les innovations dans les technologies de l'information et celles de la finance. Loin d'être en concurrence, Wall Street et la Silicon Valley convergent autour du capital des données qu'ils considèrent comme une nouvelle source d'accumulation et d'échange.

Il est important de savoir quels types de données sont collectés et comment elles sont utilisées. Bien entendu, à l'instar du capital en général, elles ne sont pas toutes identiques ni utilisées de la même manière pour générer de la valeur. Souvent, ce sont des applications mobiles qui sont utilisées pour collecter des données sur leurs utilisateurs, mais il est également possible de valoriser de nombreuses autres choses.

1.1. Les données sont utilisées pour profiler et cibler les individus

De nombreux modèles économiques du capitalisme des données reposent sur l'idée que plus l'on dispose d'informations sur les individus, plus grand sera le profit et/ou le pouvoir que l'on pourra en extraire. Les entreprises numériques tirent souvent leurs revenus des publicités personnalisées. Les courtiers en données rassemblent les informations pour créer des dossiers sur les individus et les classer en segments de marché. Les agences d'évaluation de crédit s'en servent pour noter le risque financier et la fiabilité des emprunteurs. Les commerçants peuvent pratiquer des prix différents en fonction des caractéristiques du client. Les consultants politiques analysent les données pour déterminer qui est susceptible d'être influencé par certains types de messages. Les données sont également utilisées pour optimiser les systèmes et augmenter les profits.

Cette façon de dériver la valeur est également à la base du taylorisme, qui a commencé par des études sur le temps et le mouvement des ouvriers d'usine au début des années 1900. Aujourd'hui, le taylorisme numérique, ce sont les bracelets chez Amazon qui sont portés par les travailleurs des entrepôts pour connaître la position de leurs mains et qui vibrent s'ils travaillent de manière inefficace. Finalement, la surveillance pour récupérer un maximum de données valorisables concerne toutes les sphères de l'entreprise : l'environnement interne pour optimiser les systèmes de production ou les organigrammes et les relations au sein de l'entreprise, l'environnement externe pour comprendre comment les clients utilisent le produit ou le service, les déterminants de l'acte d'achat, ou encore les audiences à cibler pour les dépenses marketing. En savoir plus sur les caractéristiques de l'offre comme de la demande, en monitorant les individus, est une voie de rationalisation avantageuse dans une optique capitaliste de réduction des coûts et de hausse des profits.

1.2. *Les données sont utilisées pour gérer et contrôler les choses*

Il s'agit d'une relation pouvoir/connaissance dans laquelle les données sont une forme numérique, mobile et traitable, de la connaissance. L'idée est qu'en amassant des données sur une chose, il est possible d'exercer un pouvoir sur cette chose et d'en extraire encore plus de données. Il peut s'agir de quelque chose d'aussi banal qu'une personne qui suit son régime alimentaire et ses exercices physiques pour gérer sa santé. Mais cela peut également être la police qui utilise des caméras portées et des drones équipés de logiciels de reconnaissance faciale et de lecture de plaques d'immatriculation. Ou encore un ingénieur qui supervise les schémas de circulation d'une ville afin de gérer la façon dont des millions de personnes se déplacent dans l'espace. Ces données peuvent être utilisées pour éclairer la prise de décision humaine ou alimenter des systèmes automatisés qui réagissent en temps réel.

1.3. *Les données sont utilisées pour modéliser les probabilités*

Avec suffisamment de données couvrant un large éventail de variables sur une période donnée, alimentées par les bons algorithmes, de nombreuses entreprises promettent qu'elles peuvent prédire l'avenir. Bien que ces « prédictions » soient en fait des calculs de probabilités, il existe un marché croissant pour ces services. Par exemple, les services de police utilisent des systèmes « prédictifs » pour créer des « points chauds » qui indiquent qui a une forte probabilité de commettre un crime et où. HunchLab, un outil de police prédictive, utilise des données sur « des dizaines de facteurs comme la densité de population, le recensement, l'emplacement des bars, des églises, des écoles et des centres de transport, les horaires des matchs à domicile et même les phases de la lune ».

De même, les salles de supervision urbaine traitent un flux constant de données pour mettre au point des simulations d'événements et répondre aux catastrophes et pour modéliser l'état possible de la ville à certaines heures et certains jours dans le futur.

Le pouvoir conféré aux modélisations probabilistes par les données justifie des précautions vis-à-vis des risques : abuser de prédictions probabilistes pour attribuer une responsabilité individuelle pourrait signer la fin de la volonté humaine et du libre arbitre. Le défi des autorités de régulation comme des exploitants des mégadonnées est de donner du sens à ces innovations pour se protéger d'une dictature des données.

1.4. *Les données sont utilisées pour fabriquer des objets*

Les systèmes et les services numériques sont souvent construits à partir de données. Ils en ont besoin pour fonctionner, ils utilisent des données existantes et en collectent d'autres. À mesure que les services se transforment en plateformes et que les appareils deviennent « intelligents », ils s'axent sur les données et se connectent à Internet. Par exemple, Uber ne fonctionnerait pas sans données en temps réel.

De nombreux biens, par exemple les maisons intelligentes, et de nombreux environnements urbains comme les villes intelligentes, reposent sur l'extraction et sur

l'exploitation des données. Dans ces *Smart Cities*, les données informant sur les habitants de la ville et leur comportement ainsi que sur les infrastructures permettent de développer de nouvelles solutions en matière d'optimisation de la production d'énergie ou des réseaux de transport. Les progrès des technologies émergentes comme l'intelligence artificielle et les véhicules autonomes nécessitent également des montagnes de données.

1.5. *Les données sont utilisées pour faire croître la valeur des actifs*

Les bâtiments, les infrastructures, les véhicules et les machines par exemple sont des actifs qui se déprécient. Ils perdent de la valeur au fil du temps à force d'usure. Cependant, les technologies intelligentes qui collectent des données sur leur utilisation permettent de lutter contre le cycle normal de détérioration. L'intelligence artificielle combinée à l'Internet des objets rendra ces derniers adaptables et réactifs prolongeant ainsi leur durée de vie utile. Plutôt que de se déprécier, les actifs intelligents peuvent conserver et même gagner en valeur. S'ils ne prennent pas de valeur, les données peuvent au moins ralentir leur déclin.

2. EXTRACTION DES DONNÉES

Lorsque l'on parle de collecte ou d'exploitation des données, on s'imagine un processus neutre qui concerne des données objectives sur le monde qui nous entoure. Cependant, il s'agit souvent d'informations sur les individus : leur identité, leurs croyances, leurs comportements et d'autres informations personnelles. Pour les entreprises, une grande partie de la valeur produite par la technologie intelligente ne provient pas nécessairement de l'achat du bien, mais plutôt de son utilisation. Ou même, du simple fait de l'avoir à portée de main puisque de nombreux appareils enregistrent en permanence l'activité de leur propriétaire et les caractéristiques de leur environnement. Les données sont au cœur de la création de valeur des entreprises dans une société où le produit du travail tend à devenir immatériel : contenu culturel, savoir, service (blog, site web, vidéo YouTube, logiciel, etc.). Les marchandises matérielles renferment elles-mêmes de plus en plus de savoir valorisable dont elles ne sont finalement qu'une sorte de matérialisation. L'interaction avec les objets intelligents, en particulier ceux qui sont intégrés dans notre vie quotidienne, génère des quantités de données. Ces données intéressent les entreprises, mais également les États. En février 2016, le directeur américain du renseignement national de l'époque, James Clapper, a admis devant un panel du Sénat que les agences fédérales pourraient utiliser les objets connectés pour la surveillance de masse : « À l'avenir », a-t-il admis, « les services de renseignement pourraient utiliser l'Internet des objets à des fins d'identification, de surveillance, de contrôle, de localisation et de ciblage pour le recrutement, ou pour accéder aux réseaux ou aux identifiants des utilisateurs ».

Un exemple typique est le réfrigérateur. Le réfrigérateur ordinaire est un objet passif : il ne fait que conserver les aliments au froid. Le réfrigérateur intelligent est

un objet actif : il conserve les aliments au froid, mais il garde également la trace des choses comme vos marques préférées, les aliments que vous consommez et à quelle heure. Il sait quand vos aliments sont presque épuisés ou périmés et peut ensuite utiliser ces données pour vous envoyer des publicités ciblées, recommander des recettes parrainées, surveiller vos apports alimentaires et passer commande automatiquement. Le réfrigérateur intelligent peut également être utilisé à d'autres fins, notamment pour vous faire surveiller à distance par la police. C'est ainsi que fonctionne la logique de l'accumulation : elle transforme le réfrigérateur en une machine à produire, à collecter et à transmettre des données.

Des réactions commencent déjà à se manifester : en 2017, un fabricant d'appareils électroménagers américain, Whirlpool, a saisi l'État en lui demandant d'imposer des droits de douane à ses concurrents coréens, LG et Samsung, car ces entreprises vendent des appareils intelligents à bas prix, ce qui grignote la part de marché des entreprises comme Whirlpool. LG et Samsung sont en mesure d'agir ainsi parce qu'elles reconnaissent, comme l'a observé le *New Yorker*, que « la façon de gagner pour une entreprise axée sur les données, c'est de pousser les prix aussi bas que possible afin de construire une base de clients, d'améliorer le flux de données et d'en tirer profit à long terme ». Alors que Whirlpool tire ses profits de la vente des appareils électroménagers, LG et Samsung misent sur les données provenant des utilisateurs. Ainsi, un appareil intelligent n'est plus simplement une marchandise, mais il devient (peut-être principalement) un moyen de produire des données. Dans la même logique, les aspirateurs robotisés cartographient secrètement le domicile des utilisateurs afin que le fabricant puisse exploiter ces données. L'accumulation de données est à l'origine de nombreuses décisions clés concernant le développement technologique, la gouvernance politique et les modèles économiques.

Comme l'explique Shoshana Zuboff, dans le contexte de ce qu'elle appelle le capitalisme de surveillance, « la logique de l'accumulation organise la perception et façonne l'expression des affordances technologiques à la racine. C'est le contexte acquis de tout modèle économique. Ses hypothèses sont largement tacites et son pouvoir de façonner le champ des possibles est donc largement invisible. Elle définit les objectifs, les réussites, les échecs et les problèmes. Elle détermine ce qui est mesuré et ce qui n'est pas pris en compte, la manière dont les ressources et les personnes sont allouées et organisées, qui est valorisé dans quels rôles, quelles activités sont entreprises et dans quel but. La logique d'accumulation produit ses propres relations sociales et, avec cela, ses conceptions et son utilisation de l'autorité et du pouvoir ».

Lorsque les données sont traitées comme une forme de capital, l'impératif d'en collecter autant que possible, à partir d'autant de sources que possible et par tous les moyens, intensifie les pratiques d'accumulation existantes et conduit à la création de nouvelles. En effet, à l'instar d'autres systèmes d'extraction de l'histoire du capitalisme – l'accaparement des terres et l'exploitation des ressources –, de nombreuses pratiques d'accumulation de données désormais courantes sont des formes d'extraction plus énergiques, dans lesquelles les données sont prises sans le consentement des producteurs et des sources et ce, sans compensation équitable.

Le terme de mégadonnées recouvre un ensemble de données hétérogènes – pour ne pas dire hétéroclites. Il est d'usage de les décrire par le symbole « 3V » correspondant à « vitesse » (fréquence d'actualisation élevée), « variété » (images, données de téléphonie mobile, données issues de capteurs, textes, etc.) et « volume » (masse d'informations considérable) qui en résulte étant considérable. D'autres chercheurs ont tenté d'ajouter les termes de véracité ou de complexité concernant les données collectées. Ils ne sont cependant pas parvenus à une description exhaustive qui ne laisse pas de côté le rôle des capacités, notamment technologiques et humaines, sans lesquelles ces données demeureraient inertes, ni celui des acteurs et des enjeux d'économie politique. On parle dans ce cas des « 3C » : « crumbs » (miettes en anglais) pour évoquer l'idée de traces numériques laissées par les utilisateurs, « capacités » (humaines, technologiques, techniques, institutionnelles) et « communauté » pour évoquer l'émergence et les interactions d'acteurs divers tels que les producteurs, les collecteurs et les analystes, au sein de ce nouvel écosystème. En effet, dans cette masse de données l'enjeu est de collecter les informations qui font du sens et de disposer de la capacité à les faire parler pour prendre des décisions, en extraire de la valeur ainsi que de nouveaux concepts techniques, commerciaux ou organisationnels. La maîtrise technique passe par la maîtrise des compétences nécessaires, un enjeu de taille qui concerne autant la collecte et la création de données que leur sélection, leur manipulation, leur analyse et leur utilisation.

La terminologie utilisée pour décrire la manière dont les données sont accumulées – en particulier les données sur les personnes – occulte le fait que ces données sont souvent acquises de manière cachée à des fins inconnues par les cibles. La question du consentement est relativement simple. La manière dont les entreprises technologiques traitent le consentement est problématique, ce n'est un secret pour personne ; c'est d'ailleurs une question souvent soulevée par les journalistes et les universitaires. Lorsque les entreprises cherchent à obtenir le consentement pour enregistrer, utiliser et/ou vendre les données d'une personne, cela se fait généralement sous la forme d'un contrat.

Dans un monde idéal, le consommateur serait informé de l'exploitation et du stockage de ses données et procéderait à un arbitrage entre le bénéfice apporté par les nouveaux services et la perte de confidentialité concernant sa vie privée. Cependant, même si dans de nombreux cas les individus doivent accepter que leurs données soient accumulées, cet accord ressemble peu à la définition du consentement. Sans même parler de consentement éclairé. Lorsqu'une chose est prise sans consentement, on appelle ça un vol. Que le vol porte sur un bien immatériel, ici des informations sur un individu, plutôt qu'un bien matériel, n'enlève rien au problème éthique posé par l'acte. La question de la compensation équitable est plus compliquée en grande partie parce qu'il est difficile de donner un juste prix à des informations personnelles. Les données sont évaluées différemment par différentes entreprises. Leur valeur augmente également de manière non linéaire en fonction de leur quantité. Plus une banque de données est importante et diversifiée, plus on peut en tirer d'informations. Ainsi, les données concernant un individu peuvent ne pas être facilement converties en capital économique, mais les données agrégées de centaines, de milliers ou de millions d'individus peuvent avoir une valeur immense.

Même s'il est difficile de fixer le prix des données, il est possible de juger l'équité de la compensation d'au moins deux façons : (1) quel est le type de compensation offert, le cas échéant, et (2) quelle est la différence entre la compensation et la valeur qui en est tirée *in fine* ? La compensation prend le plus souvent la forme d'un accès à des services, par exemple la plateforme de Facebook et le moteur de recherche de Google. Plutôt que de faire payer l'utilisation du service, l'entreprise collecte des données en guise de paiement. La valeur de ces données pour sa part est énorme, à tel point que certaines des entreprises les plus riches du monde, comme Facebook et Google, sont bâties sur le capital de données.

On estime que l'industrie du courtage de données génère 200 milliards de dollars de revenus annuels. À eux seuls, les trois plus grands courtiers en données – Experian, Equifax et Transunion – rapportent chacun des milliards de dollars par an. Même dans le cas des courtiers relativement modestes, la différence entre la valeur générée par les données et ce qu'ils offrent en échange est considérable. En outre, d'autres grands secteurs comme la finance, l'assurance et l'industrie manufacturière s'appuient de plus en plus sur le capital de données pour générer de la valeur. Ces entreprises accumulent des milliards de dollars de plus-value à partir du « travail numérique » effectué par des individus, tout en ne payant que peu, voire rien, en retour.

Il est important de noter que des critères individuels, par exemple l'identité et la classe sociale des individus, peuvent parfois influencer la nature de l'extraction de données. Certains groupes font l'objet de davantage de collecte que d'autres, notamment les personnes plus démunies que les pouvoirs publics et les institutions financières suivent quelquefois plus, ou les personnes de couleur qui peuvent passer au travers des algorithmes qui sont majoritairement entraînés avec des visages d'hommes blancs.

Si l'univers est vu comme une réserve potentiellement infinie de données, cela signifie que l'accumulation et la circulation des données n'ont pas de fin. L'impératif de collecte influence de nombreuses décisions concernant les modèles économiques, la gouvernance politique ou le développement technologique. Les données personnelles sont prélevées avec peu de considération pour le consentement et la compensation.

Aussi importantes que soient ces questions, elles masquent les problèmes systémiques d'inégalité et d'exploitation qui sont endémiques à l'économie politique contemporaine des données. En outre, voir les différentes pratiques de collecte de données comme une extraction permet de jeter les bases des réponses politiques et juridiques à cette accumulation effrénée et invasive. Ces réponses peuvent se présenter sous la forme de réglementations – essentiellement de contrôle de capitaux – visant les données que les entreprises peuvent collecter, la manière de le faire, les lieux où elles peuvent les envoyer et les stocker et la quantité. Les pouvoirs publics peuvent également prendre à leur charge certains aspects de la gestion des données et les considérer comme un élément de l'infrastructure d'un pays.

Le fait de suggérer aux États de prendre en charge certaines parties de l'économie des données et de briser des entreprises monopolistiques comme Google peut être vu comme la démonstration de la puissance des mégadonnées, qui se trouvent aujourd'hui

au même niveau que l'industrie du pétrole ou le secteur de la finance. Cela illustre la nécessité de poursuivre une réflexion critique sur l'économie politique des données, ainsi que sur les réformes et les alternatives au capitalisme des données qui est un changement d'orientation ; une transition vers un nouveau type de capital et de nouvelles méthodes d'accumulation. Cette transition découle de l'un des régimes socioéconomiques qui a dominé ces dernières décennies : le capitalisme financier.

La transition vers une économie des données, comme la transition vers une économie de la finance avant elle, constitue une nouvelle frontière de l'accumulation et une nouvelle étape du capitalisme. Cette nouvelle économie n'en est encore qu'à ses débuts, mais la richesse et le pouvoir des capitalistes des données sont immenses et continuent de croître. Les théories et les méthodes utilisées pour analyser le capitalisme financier et les technologies de l'information doivent maintenant être synthétisées et appliquées à l'étude de la signification, de la pratique et des implications de l'économie des données en tant que régime économique-politique.

3. LES MÉGADONNÉES (BIG DATA)

Les mégadonnées sont de grands volumes d'informations très variées, générées, capturées et traitées à grande vitesse (Laney, 2001) et dont la quantité augmente de façon exponentielle. Aujourd'hui, il existe trois fois plus d'objets connectés que d'habitants sur la Terre, avec une croissance de 10 % par an. Ce traitement est difficile par les moyens classiques (Constantiou et Kallinikos, 2015), mais à l'aide de technologies analytiques avancées, les entreprises peuvent les utiliser pour développer des idées, des produits et des services innovants (Davenport *et al.*, 2012).

Les entreprises espèrent profiter des mégadonnées pour en tirer des avantages dans de nombreux domaines, notamment le commerce électronique, l'administration en ligne, la science, la santé et la sécurité (Chen *et al.*, 2012). Ce qui constitue de la valeur à leurs yeux dépend de leurs objectifs stratégiques (Ghoshal *et al.*, 2014). Il s'agit à la fois de valeur sociale et de valeur économique.

Concernant la valeur sociale, il peut s'agir de l'amélioration des domaines tels que l'éducation (Cech *et al.*, 2015), la santé (Raghupathi et Raghupathi, 2014) et la sécurité (Newell et Marabelli, 2015). Les pouvoirs publics peuvent par exemple utiliser les mégadonnées pour « améliorer la transparence, accroître l'implication des citoyens dans les affaires publiques, prévenir la fraude et la criminalité, améliorer la sécurité nationale et participer au bien-être des personnes par une meilleure éducation et de meilleurs soins » (Kim *et al.*, 2014, p. 81). On retrouve dans cette catégorie des avantages pour les individus, mais aussi pour la société, notamment sous la forme de croissance de l'emploi et de productivité (Loebbecke et Picot, 2015).

La valeur économique produite par les mégadonnées peut se traduire pour une entreprise par une augmentation des bénéfices, par la croissance de l'activité et par un avantage concurrentiel (Davenport, 2006 ; Davis, 2014 ; Tyagi, 2003). Les entreprises qui s'appuient sur les mégadonnées pour orienter leur stratégie et leur activité

quotidienne obtiennent de meilleurs résultats financiers que celles qui ne le font pas (LaValle et al., 2011 ; McAfee et Brynjolfsson, 2012).

En général, les mégadonnées sont perçues comme une source de services et d'opportunités commerciales innovants (Davenport et al., 2012 ; Davenport et Kudyba, 2016 ; McAfee et Brynjolfsson, 2012). En outre, on estime qu'elles permettent de gagner en efficacité, par exemple en optimisant les flux de la chaîne d'approvisionnement, en fixant le prix le plus rentable pour les produits et les services, en sélectionnant les bonnes personnes pour certaines tâches et certains emplois, en minimisant les erreurs et les problèmes de qualité, et en améliorant les relations avec les clients (Chen et al., 2012 ; Davenport, 2006 ; McAfee et Brynjolfsson, 2012). En outre, les mégadonnées peuvent permettre d'améliorer les prises de décision et d'en tirer une valeur économique et sociale supplémentaire (Sharma et al., 2014) par l'élaboration de stratégies plus éclairées (Constantiou et Kallinikos, 2015). C'est le cas notamment en basculant dans une optique de pilotage par les données et centrée sur le client dans laquelle les entreprises s'organisent autour des données et de la connaissance de leur clientèle pour la fidéliser et pour la faire croître (Alain Levy, *Big, Fast Et Open Data*, 2014).

Ainsi, la littérature universitaire et professionnelle démontre les opportunités que les mégadonnées offrent aux entreprises (Clarke, 2016). Cependant, le gain de valeur n'est pas garanti et certaines entreprises peuvent surestimer leur capacité à en bénéficier (Ransbotham et al., 2016 ; Ross et al., 2013).

4. APPROCHE INDUCTIVE ET DÉDUCTIVE DE L'ANALYSE DES MÉGADONNÉES

Les mégadonnées proviennent de différentes sources : de l'intérieur de l'entreprise avec, entre autres, les progiciels de gestion intégrés (*ERP* en anglais) et les transactions ; et de l'extérieur avec les données fournies par des tiers ou générées par les utilisateurs, les données publiques ou encore celles issues des capteurs (par exemple, Zuboff, 2015). Par conséquent, elles sont souvent produites et collectées à d'autres fins que celles pour lesquelles elles sont utilisées (Constantiou et Kallinikos, 2015 ; Newell et Marabelli, 2015).

En fonction de la variété (Kim, 2015) et de la granularité (c'est-à-dire le niveau de détail) des données (Yoo, 2015), il peut être difficile de prévoir *ex ante* quelles informations peuvent être obtenues à partir de quelles sources (Aaltonen et Tempini, 2014 ; Constantiou et Kallinikos, 2015). Les mégadonnées poussent à la collecte sans but et favorisent une approche ascendante et inductive de cette collecte ainsi que de l'exploration et de l'analyse (Constantiou et Kallinikos, 2015 ; Olbrich, 2014 ; Van den Broek et Van Veenstra, 2015). Une telle approche part des données et cherche ensuite à générer une explication théorique. Par exemple, Madsen (2015) a réalisé une étude sur la façon dont les caractéristiques technologiques de l'analytique sociale numérique (un sous-ensemble de l'analytique des mégadonnées) influençaient le travail de projet. Cette étude illustre, au travers du cas d'un chef de projet, le fait que les analystes peuvent donner du sens aux mégadonnées en les regroupant dans des catégories prédéterminées, ou en choisissant d'accepter des

Un éclairage inédit sur le nouveau capitalisme des données et ses enjeux

Lionel Melka est le responsable de la recherche de la société de gestion Homa Capital.

Il enseigne la finance à l'Université Paris-Dauphine depuis plus de quinze ans, est investisseur dans plusieurs startup à impact et conseille quelques grandes familles industrielles dans leurs investissements et leur allocation d'actifs.

Il est secrétaire général d'une association d'aide aux personnes réfugiées et Président de la Fondation Copernic.

Rachel Delacour est CEO et co-fondatrice de Sweep, une société innovante permettant aux entreprises de réduire leur empreinte carbone, et est une personnalité emblématique de la French tech.

La transition numérique provoque un bouleversement de l'économie mondiale. Dans la plupart des secteurs, la collecte, le traitement et l'exploitation des données permettent l'émergence de nouveaux modèles d'affaires.

Google, TikTok, PayPal, Shopify, Tinder, BlaBlaCar, Uber, Palantir, IQVIA : comment ces sociétés dont le modèle d'affaires est basé sur l'exploitation des données ont-elles crû de manière si fulgurante ? Comment parviennent-elles à dominer leur secteur ?

L'essor de cette nouvelle économie autour des données est marqué par la domination d'un principe de redevabilité qui investit tous les espaces de pouvoir, tant dans les démocraties que dans les entreprises. Pourtant, cette révolution a rarement fait l'objet d'une analyse rigoureuse, au-delà des controverses médiatiques.

Ce livre raconte cette transformation et explique l'émergence de ces nouveaux acteurs. Il allie une approche théorique et des cas pratiques permettant de comprendre cette révolution en cours. Il s'adresse donc aussi bien aux étudiants en économie et gestion (écoles de commerce, universités, etc.), qu'aux professionnels de la finance (dirigeants d'entreprises, gérants de fonds, etc.) et à tous ceux qui souhaitent mieux comprendre les bouleversements récents du capitalisme.

RESSOURCES NUMÉRIQUES



Grâce aux QR Codes, accédez à 10 nouvelles études de cas en ligne : LinkedIn, Tesla, Square, Inovie...

ISBN 978-2-8073-4815-8



deboeck
SUPÉRIEUR

www.deboecksuperieur.com