

Big Data : la révolution des données est en marche de Viktor Mayer-Schönberger et Kenneth Cukier

Smart : enquête sur les internets de Frédéric Martel

Marie-Ève Tremblay-Cléroux

Number 254, Fall 2015

La galaxie cybernétique

URI: <https://id.erudit.org/iderudit/79870ac>

[See table of contents](#)

Publisher(s)

Spirale magazine culturel inc.

ISSN

0225-9044 (print)

1923-3213 (digital)

[Explore this journal](#)

Cite this article

Tremblay-Cléroux, M.-È. (2015). *Big Data : la révolution des données est en marche* de Viktor Mayer-Schönberger et Kenneth Cukier / *Smart : enquête sur les internets* de Frédéric Martel. *Spirale*, (254), 31–33.

RÉVOLUTION NUMÉRIQUE : LA MISE EN DONNÉES DU MONDE

PAR MARIE-ÈVE TREMBLAY-CLÉROUX

BIG DATA : LA RÉVOLUTION DES DONNÉES EST EN MARCHÉ

de Viktor Mayer-Schönberger et Kenneth Cukier
Traduit de l'anglais par Hayet Dhifallah
Robert Laffont, 96 p.

SMART : ENQUÊTE SUR LES INTERNETS

de Frédéric Martel
Stock, 406 p.

En cette ère du numérique, les GAFA (acronyme regroupant Google, Amazon, Facebook et Apple) n'ont pas seulement réussi à tirer leur épingle du jeu : elles ont atteint une puissance inégalée. Nul besoin ici de les décrire, car leur présence dans la vie quotidienne occidentale relève de l'évidence. L'achat en ligne, l'utilisation du web, les médias sociaux et les téléphones intelligents rythment notre existence. Fers de lance de cette révolution, les GAFA sont engagées dans un processus de mondialisation accéléré. Mais jusqu'où ira cette globalisation du monde ?

De nos jours, les capacités de stockage numérique sont telles que chaque personne en Occident possède chez elle environ 320 fois l'équivalent des informations qui avaient été stockées dans la bibliothèque d'Alexandrie. Dans leur essai, Viktor Mayer-Schönberger et Kenneth Cukier s'intéressent à l'impact qu'a engendré l'augmentation draconienne du nombre de données numériques et à l'usage qu'en font les GAFA. Les *big data*, soit l'utilisation de l'ensemble des données disponibles sur un sujet en particulier plutôt que leur échantillonnage, ont entraîné un changement de paradigme scientifique. Le fantasme propre au XIX^e siècle, tout connaître, se manifeste en ce XXI^e siècle sous le désir de quantifier

le monde : « [le] monde ne nous apparaît plus telle une suite d'événements à interpréter comme des phénomènes naturels ou sociaux, mais comme un univers composé essentiellement d'informations ». *Big Data : la révolution des données est en marche* présente l'uniformisation du monde par sa mise en données. La disparition des modèles théoriques et des experts au profit des spécialistes en analyse de données, en intelligence artificielle, en mathématiques ou en statistiques y est annoncée.

D'une méthode qui vise l'exactitude parfaite et qui s'intéresse avant tout à la causalité, on serait passé à une autre, moins précise, qui se concentre sur la corrélation, la tendance. L'utilisation des *big data* s'inscrit parfaitement dans l'esprit du temps : instantanéité, rendement, application immédiate, maximisation des profits. Enthousiastes, les auteurs vantent la supériorité de l'analyse de données par corrélation, puisque « dans de plus en plus de cas, cette analyse vaut mieux qu'une longue réflexion en termes de causalité, qui nécessite de longues et coûteuses expériences, soigneusement contrôlées ». Les auteurs de l'essai adhèrent même à l'audacieuse thèse de Chris Anderson, l'ex-rédacteur en chef du magazine *Wired*, selon laquelle « le déluge de données rend la méthode scientifique obsolète ».

Marchandisation des données

Nos activités courantes réalisées grâce aux GAFA créent une multitude de données, qui leur sont très utiles sur le plan économique. Il s'agit en quelque sorte d'un travail gratuit que nous accomplissons à notre insu. Aujourd'hui, la valeur économique exacte de ces données demeure encore difficile à évaluer. Ainsi, en 2009, Google rachète l'application Captcha¹ développée dans les années quatre-vingt-dix

pour contrer les *spambots* et la transforme afin de l'utiliser dans son processus de numérisation Google Books. Baptisée ReCaptcha², cette nouvelle application lui fait gagner plus d'un milliard de dollars américains chaque année. En effet, des milliers de personnes déchiffrent quotidiennement les mots que les systèmes de reconnaissance optique ont échoué à décoder.

L'utilisation des *big data* chez Amazon explique le succès colossal du vendeur en ligne. Au début de son existence, Amazon employait une dizaine de critiques littéraires afin de personnaliser le contenu du site. Toutefois, dès que son système de recommandations personnelles informatisées basées sur les liens entre les produits a été mis au point, les critiques ont été renvoyés. Les données étaient bien plus efficaces pour vendre des livres que les humains. L'omniprésence des *big data* influence également les mégaproductions cinématographiques. L'immense base de données the-numbers.com sert spécifiquement à conseiller les producteurs de films et les institutions financières sur le choix des acteurs ou sur le budget. Ce site internet se spécialise dans les projections économiques basées sur la similarité entre les films, le budget, le genre et la méthode de production³. Une des conséquences de cette utilisation des données, son impact direct sur la créativité et la liberté artistique, est pourtant passée sous silence par les auteurs. Dans ce modèle, la dimension économique domine alors la dimension artistique : la viabilité du projet cinématographique repose uniquement sur son potentiel de rentabilité. Parce que difficilement calculable, un élément vital comme l'originalité du scénario ne peut être prise en compte.

Diversité des internets

Le portrait que dresse Frédéric Martel des GAFAs dans *Smart : enquête sur les internets* nuance le propos tenu dans l'essai *Big Data*. Après avoir étudié l'état du web aux États-Unis, en Chine, en Russie, en Amérique du Sud, en Europe, à Gaza et dans les pays musulmans, Martel affirme que la révolution numérique vécue actuellement ne débouche pas sur une globalisation généralisée, et encore moins sur une américanisation du monde. Pour ce faire, il

s'emploie à exposer qu'« *internet n'abolit pas les limites géographiques traditionnelles, ne dissout pas les identités culturelles, n'aplanit pas les différences linguistiques : il les consacre* ».

L'essai, qui enquête sur la situation du numérique dans une cinquantaine de pays, débute là où tout a commencé : la Silicon Valley. La baie de San Francisco, lieu de naissance de Google, Facebook et Apple, continue de rayonner grâce à l'état d'esprit qui y règne, « *aux intersections entre le monde de la recherche, celui de la finance et de l'entrepreneuriat* ». Modèle par excellence, le fonctionnement de la Silicon Valley est dupliqué avec plus ou moins de succès par des États partout dans le monde. En Chine, les ambitions mégalomanes des GAFAs se trouvent bloquées par le gouvernement, qui les a chassées de son territoire. De l'autre côté d'un rideau de fer numérique se dressent l'intranet chinois et ses clones. S'y déploient des versions copiées, et adaptées aux réalités chinoises, de la plupart des grandes applications étatsuniennes. Les 560 millions d'utilisateurs de Renren (Facebook), Youku (YouTube), QQ (MSN), Weibo (Twitter) et Baidu (Google), surveillés et censurés par les autorités, sont coupés du reste du monde numérique.

L'essai de Martel démontre comment « *la révolution numérique apparaît, au contraire, comme une territorialisation et une fragmentation : internet, c'est un "territoire"* ». Plusieurs lieux résistent à l'hégémonie de l'internet étatsunien. Le plus grand ghetto d'Afrique, Kibera au Kenya, a été boudé par Google Maps. Ici, les *big data* sont au service des gens. Map Kibera utilise les masses de données de la communauté afin de fournir une carte numérique mise à jour instantanément de ce bidonville d'environ 500 000 habitants. À Monterrey au Mexique, l'usage de Twitter a été subverti par la population. Alors que le Québec a connu #manifencours pendant le printemps érable, les résidents de Monterrey suivent #Monterreyfollow pour se tenir informés en temps réel des tueries, barrages policiers et autres dommages collatéraux de la guerre entre narcotrafiquants. L'utilisation d'un phénomène mondial comme celui des GAFAs est donc territorialisée ; les communautés locales se les approprient de façon différente.

De la collecte de données à la surveillance

Quand les voitures de Google Street View sillonnent les rues ou que notre iPhone est allumé, de grandes quantités d'information transitent vers Google ou Apple. Au-delà de la situation de surveillance constante dans laquelle nous nous trouvons, l'utilisation des *big data* complique non seulement la protection de la vie privée des usagers d'internet, mais ouvre également la porte à une pénalisation des individus en fonction de prévisions. En ce sens, l'essai *Big Data* confirme ce que plusieurs craignaient déjà, soit que « *les données relatives aux personnes sont fréquemment collectées de manière passive, sans effort ou presque de ces dernières, voire à leur insu.* »

Mayer-Schönberger et Cukier appréhendent, sur un ton légèrement paranoïaque, une dérive étatique dans l'usage des *big data* : qu'une dictature des données menant à un scénario similaire à celui du film *Minority Report* soit instaurée. Il ne s'agirait plus, dans cette dystopie, d'utiliser les données afin de faire du profilage (tels individus qui agissent de telle manière doivent être davantage surveillés), mais bien pour cibler précisément un individu en particulier et pour l'arrêter avant que le crime ne se produise. Plutôt que de pénaliser l'intention de commettre un crime (c'est déjà le cas), on pénaliserait l'appartenance à un groupe où la possibilité de perpétrer ledit crime serait très probable.

Or, Frédéric Martel dresse un portrait bien plus inquiétant encore. Il nous rappelle que « *les responsables de Facebook et de Google, dont le modèle économique repose largement sur la collecte des données personnelles, ont laissé entendre que la protection de la vie privée pourrait finalement être une "anomalie"* ». Suite à l'affaire Snowden, nombre de citoyens ont réalisé qu'il n'existait pas de protection minimale des données aux États-Unis. Le secret du

succès des GAFAs repose justement sur l'utilisation des *big data* à des fins commerciales, menant aux abus que nous connaissons. Pourtant, aux dires de Gary Reback dit « *le bouledogue de l'antitrust* », l'un des rares avocats qui ose se frotter aux géants, aucune organisation étatsunienne n'a le pouvoir ou la volonté de s'en prendre à Google.

Nous en sommes là

L'ouvrage de Martel pallie les lacunes de *Big Data : la révolution des données est en marche*. L'enquête sur les *internets* permet de saisir concrètement les multiples variantes du web, et ce selon le contexte politique, social et économique. Alors que les auteurs de *Big Data* valorisent (parfois excessivement) le phénomène de la mise en données du monde, Martel présente les impacts de cette mise en données sur la société et effectue une critique de l'intégration de la vie quotidienne des gens (transformée en *big data*) au marché.

Même si le futur demeure difficile à prédire, ces deux essais mettent en lumière la prépondérance que les *big data* ont déjà acquise. De plus, la thèse originale de Martel offre la possibilité d'attaquer de front l'idée qu'internet tend à uniformiser le monde en démontrant comment les différentes cultures adaptent les GAFAs à leurs réalités locales sans y dissoudre leurs identités. Le véritable danger du numérique ne se situe donc pas dans la potentielle hégémonie de l'anglais sur le web ou le règne mondial des GAFAs, mais dans le traitement des données personnelles. Une épée de Damoclès pend au-dessus de nos têtes : le fait de concentrer autant d'informations sensibles entre les mains des gouvernements ou d'entreprises comme Google pourrait conduire à une situation explosive à moyen ou long terme. Quant à la naissance des identités numériques et la question de leur propriété, elles risquent d'être au cœur des débats au cours de la prochaine décennie. ■

¹ Popularisé dans les années 2000 pour contrer les courriers électroniques non sollicités, un Captcha est un test informatique utilisé fréquemment afin de différencier un être humain d'un ordinateur. Ce test peut prendre la forme d'une série de lettres et de chiffres collés les uns aux autres, d'un mot qui apparaît sur un fond embrouillé ou rayé par une ligne brisée.

² Le ReCaptcha présente deux éléments distincts à l'utilisateur : le premier est un Captcha traditionnel, dont la réponse est connue par le système ; le second est un terme que le système de reconnaissance optique de Google n'a pas réussi à identifier. Google se sert donc des aptitudes cognitives des utilisateurs afin de déchiffrer ce mot.

³ Voir à cet effet la section du site sur la méthode de recherche utilisée : <<http://www.the-numbers.com/research-analysis>>.