



HAL
open science

Le Big Data : une mise sous curatelle technico-financière

Paul Naegel

► **To cite this version:**

| Paul Naegel. Le Big Data : une mise sous curatelle technico-financière. 2017. hal-01451389

HAL Id: hal-01451389

<https://hal.science/hal-01451389>

Preprint submitted on 1 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Le Big Data : une mise sous curatelle technico-financière

par
Paul Naegel¹

Résumé

Le Big Data, en arrière-plan de ses promesses de nouvel eldorado, grâce aux mégadonnées numériques, cache une mise sous *curatelle technico-financière*. Nous empruntons cette expression au psychanalyste Roland Gori. Le mot curatelle désigne une mesure établie par justice, à destination d'une personne supposée incapable de gérer ses intérêts. Cette mesure est en fait appliquée à tout internaute, c'est-à-dire toute personne – physique ou morale - qui alimente consciemment, ou à *l'insu de son plein gré*, le Big Data, par le simple fait d'une connexion au réseau Internet.

Le Big Data est un des constituants du système technicien, et à ce titre fait penser au tableau de Goya : *Saturne, dévorant un de ses fils*. Et cela surtout depuis le déploiement de la crise financière mondiale ouverte en 2008. La critique du *système technicien* n'est pas nouvelle ; elle a été faite dès 1977 par Jacques Ellul, à une époque où les ordinateurs étaient encore des fœtus par rapport à ce qui fonctionne de nos jours dans plus de 30.000 *data center*, pour la plupart installés physiquement aux USA. C'est là que le Big Data est logé, et où, à travers ceux qui détiennent ces coffres forts, il exerce sa curatelle, autrement dit sa mise sous tutelle de ceux qui le nourrissent gratuitement.

Cette curatelle met en péril la démocratie, car ce sont désormais les *data* qui se substituent aux responsables politiques, supposés élus par les peuples, pour décider *de ce qui est bon pour nous*. Et il semble que comme l'aimait à le dire Margareth Thatcher : *There Is No Alternative (TINA)*.

Nous pensons qu'il existe des moyens d'échapper à la *curatelle* du *Big Data*, que nous tentons de suggérer dans cet article.

Abstract

Big Data, in the background of his promises of new El Dorado, thanks to big digital data, places the world under a hidden technical and financial curatorship. We have borrowed this expression from the psychoanalyst Roland Gori. Curatorship means a measure established by justice, destined for a person assumed unable to manage its interests. This measure is in fact applied to all Internet users, i.e. any person - physical or moral - that powers consciously, or unbeknownst to voluntarily, Big Data, by the simple connection to the Internet.

Big Data is one of the components of the technical system, and as such makes think to a painting of Goya: Saturn devouring one of his sons. And this particularly since the deployment of the global financial crisis opened in 2008. The

¹ Centre François Viète – Université de Nantes

criticism of the *technician system* is not new; It was made in 1977 by Jacques Ellul, at a time when computers were still fetuses, compared to what works today in more than 30,000 data center, for the most part installed physically in the USA. This is where the Big Data is housed, and where, through those who hold these safes, he exercises his curatorship, in other words his tutelage, of those who feed him for free.

This curatorship is endangering democracy, because they are now data that substitute for politicians, supposed elected by the people, to decide what is good for us. And it seems that as Margareth Thatcher liked to say: *There Is No Alternative (TINA)*.

We believe that there are ways to escape the guardianship of Big Data, wich we are trying to suggest in this article.

Introduction

Le phénomène *Big Data* est partout, et il nous cerne de toutes parts, sauf à espérer s'isoler du monde contemporain. Consciemment, ou à *l'insu de son plein gré*, tout internaute qui se connecte à un site, quel qu'il soit, alimente en permanence et gratuitement le monstre qu'est devenu le *Big Data*. Car à l'arrière-plan des promesses de nouvel eldorado – dans la sphère marchande mais aussi dans le domaine scientifique – le *Big Data* nous met sous la curatelle du système technico-financier, lequel remplace pour l'essentiel l'économie réelle du XX^e siècle.

Une curatelle est ce qui est imposé par voie de justice – c'est-à-dire en vertu de la Loi – à une personne supposée incapable d'agir par elle-même dans son propre intérêt. Dans la curatelle par le système technico-financier, il n'y a plus d'autre loi que celle du *Big Data*, c'est-à-dire des métadonnées qu'il contient. Et cette loi est celle de l'optimisation des résultats financiers pour les actionnaires de toutes sortes, dont les propriétaires des 30.000 data center dans lesquels est logé le *Big Data*.

Ce sont les algorithmes, appliqués aux mégadonnées, qui font savoir aux internautes ce qui est bon pour eux. Même lorsqu'il est question de choisir des candidats pour exercer des fonctions politiques. Lesquelles sont réduites à un rôle de figuration, et à la production d'éléments de langage visant à affirmer que *tout va bien* – ou *mal*, selon les moments - et que nous sommes encore en démocratie.

Le *Big Data* n'est pas fondé sur une *idéologie*, mot devenu une insulte, alors qu'il s'agit, étymologiquement, d'une science qui a pour objet l'étude des idées. Il est donc possible d'opposer des idéologies à travers des controverses. Le *Big Data* n'a pas pour fondement une idéologie particulière, mais une *croyance* à caractère général. Les croyances ne se discutent pas, car elles se dispensent de la validation par la raison, et reposent sur des convictions. Il est donc inutile d'essayer d'écarter un croyant de ses convictions. Et il y foule de croyants dans le phénomène du *Big Data*.

La critique du système décrit dès 1977 par Jacques Ellul dans le *système technicien*² n'a pas convaincu à l'époque, alors qu'elle contenait en germe le *Big Data*. On sait que nul n'est prophète en son pays, et que les prophètes risquent souvent leur vie !

² ELLUL J., Le système technicien, Paris, Cherche Midi, 2012, 334p. – Réédition.

Elles sont loin derrière nous, désormais, les révolutions industrielles, à commencer par celle datée de la fin du XVIII^e siècle, qui débuta en Grande Bretagne. Et les britanniques tirent maintenant l'essentiel de leur richesse non plus de l'exportation de produits manufacturés, mais des transactions financières réalisées par la City de Londres. L'économie de marché s'est non seulement imposée à quasiment tous les pays du monde, mais elle a entraîné la quantification et la numérisation de toutes ses transactions en termes monétaires.

Là encore, le *Big Data* règne en maître : des millions de transactions sur les marchés financiers sont gérés de manière automatique par les ordinateurs installés au plus près des salles de marché : chaque nanoseconde compte ; c'est le *high-frequency trading*, dans lequel les humains n'ont ni le temps, ni la possibilité d'intervenir. D'autant que les *produits dérivés*, qui sont légion, et de plus en plus sophistiqués, ont conduit à des titrisations librement négociables, mais sans plus aucun lien avec un créancier identifiable. Ce qui a donné lieu aux *produits financiers toxiques*, mis en dernier ressort à la charge des États en tant que *dette souveraine*, et ainsi transformés en impôts pour les contribuables.

En un mot comme en mille, le système technico-financier est *data driven*, c'est-à-dire qu'on est passé, comme l'affirme Alain Supiot, *d'un gouvernement par les Lois à une gouvernance par les nombres*³

Des capteurs à tous les étages pour nourrir le *Big Data*

Il fût un temps où les immeubles parisiens affichaient fièrement sur leurs murs des plaques indiquant : *eau et gaz à tous les étages*. A la place, on pourrait signaler de nos jours les immeubles dans lesquels on peut accéder à la fibre optique. Autant se passer du *Big Data* si on n'a pas cette précieuse liaison à Internet, avec son très haut débit. Mais à y regarder de près, ces réseaux de fibres optiques ressemblent beaucoup aux réseaux de distribution d'eau et de gaz, toujours en activité, et sont comme eux gérés centralement.

La nouveauté est dans les réseaux sans *fil à la patte*. Nous évoquons ici les *mobiles* de toutes sortes, à commencer par les incontournables *smartphones*⁴ que les jeunes et moins jeunes de la *génération « Y »*⁵ ont en permanence à portée de main. Certains dorment avec, de peur de manquer un message ou un appel. Observez dans un transport en commun, voire dans une salle d'attente de médecin, la frénésie avec laquelle ces appareils sont littéralement *caressés* par leur propriétaire. Pour le plus grand bien du *Big Data*, mais ils ne le savent pas, où s'en moquent. Le moment n'est pas loin où il n'y aura plus personne qui utilisera un ordinateur portable pour accéder à Internet, envoyer des courriels, consulter les nouvelles, la météo et ainsi de suite.

³ Les cours d'Alain Supiot au Collège de France analysent et dénoncent l'asservissement du Droit et de la Loi à la gouvernance par les nombres qui s'est enracinée dans les sociétés contemporaines au point d'en saper les fondements démocratiques.

⁴ Téléphones mobiles connectés sans fil à un opérateur, et munis d'écrans tactiles au lieu de claviers mécaniques.

⁵ Les fils partant des écouteurs placés dans les oreilles forment un « Y » pour rejoindre un lecteur de type MP3 porté en bandoulière.

Ce n'est qu'un début : nous vivrons bientôt avec, et au milieu, d'autres *objets connectés* : nos capteurs de données physiologiques (tensiomètres, dosimètres d'insuline, etc.), nos appareils domestiques tels que les réfrigérateurs et même leurs contenus, munis de puces *RFID*⁶, pour nous aider à regarnir les stocks depuis notre super marché favori. Certains appareils d'assistance médicale, notamment à la respiration chez les personnes affectées par des apnées du sommeil, sont déjà *connectés* via Internet à des centres médicaux. Les pneumologues peuvent ainsi savoir si le patient suit leur prescription d'usage, faute de quoi la prise en charge financière peut être suspendue ou annulée.

Les véhicules, notamment ceux à conduite dite *autonome*, sont également des objets fortement connectés, qui sont un aliment de choix pour le *Big Data*, lequel enregistrera au fur et à mesure leurs itinéraires à travers le système GPS. Déjà les avions en vol envoient via des satellites et en temps réel des informations sur l'état de leurs moteurs à ceux qui les ont fournis, dans un but de maintenance préventive.

Les corrélations : alpha et oméga des traitements du *Big Data*

L'histoire des statistiques est née avec la civilisation, et la science statistique semble avoir existé dès la naissance des premières structures sociales. Elle a commencé par des recueils de données, avant que n'apparaisse, au XVIII^e, un rôle *prévisionnel*, avec la construction des premières tables de mortalité. En 1746, Antoine Deparcieux publia un ouvrage qui fait encore référence pour les statisticiens⁷. Les statistiques ont été munies d'un outillage mathématique avec Fermat, Pascal, Bayes, Condorcet et Laplace. Mais c'est à Adolphe Quetelet que l'on doit l'idée que la statistique est une *science* s'appuyant sur les *probabilités*.

On vient de rappeler qu'à la collecte de données se sont adjointes des *théories* permettant d'en faire un usage autre que le simple archivage. Avec le développement de l'informatique, après la seconde guerre mondiale, les possibilités de stockage et de traitement des données statistiques se sont amplifiées. L'évolution de la technologie et la mise en service des réseaux reliant les ordinateurs à travers le monde a ouvert la voie au phénomène *Big Data*.

Il n'y a pas de définition de ce phénomène qui fasse consensus, chacun pouvant apporter la sienne en fonction de ses intérêts et croyances. Mais deux attitudes se dégagent pourtant à la lecture des publications relatives au *Big Data*, notamment académiques : les unes sont un plaidoyer, parfois radical, proclament qu'il constitue une *révolution scientifique* ; les autres sont souvent des critiques radicales de la *dictature des données* dans tous les domaines de la vie économique et sociale.

⁶ La RFID est un système d'identification par radiofréquence qui permet d'écrire, de stocker et de relire des informations sur des étiquettes électroniques intégrées aux produits à tracer.

⁷ DEPARCIEUX A., *Essai sur les probabilités de la durée de vie humaine*, 1746

Parmi les thuriféraires de l'exclusion de toute théorie dans l'analyse des contenus du *Big Data*, citons un article de Chris Anderson⁸. En plus consistant, parmi les promoteurs du *Big Data*, on peut lire l'ouvrage de Viktor Mayer-Schönberger et Kenneth Cukier⁹. On y trouve des affirmations péremptoires telles que : *Now, the quantitative change has produced a qualitative change* (p.10), ou encore : *Big Data is about 'what', not 'why'. We don't always need to know the cause of a phenomenon; rather, we can let data speak for itself* (p.14). Le ton étant donné, on ne s'étonnera pas de trouver un chapitre entier de l'ouvrage consacré à la question des *corrélations* (pp.50-72). On peut lire dans ce chapitre cette *prédiction* : *In the future, big-data correlations will routinely be used to disprove (réfuter) our causal intuitions, showing that often there is little if any statistical connection between the effect and its supposed cause* (p.65)

Mais assez curieusement, les auteurs¹⁰ réfutent l'article de Chris Anderson que nous avons cité plus haut, comme étant trop radical, avec sa phrase choc : *With enough data, the numbers speak for themselves. Petabytes allows us to say: 'Correlation is enough'*. Ils considèrent que la théorie n'est pas absente du *Big Data*, étant lui-même fondé sur de la théorie, puisqu'il emploie des théories statistiques et des mathématiques, et par moments la science informatique. Mais pour en finir avec cette pseudo controverse, Mayer-Schönberger et Cukier affirment quand même : *Yes, there are no theories [dans le Big Data] about the causal dynamics of a particular phenomenon*.

Il s'agit bien, finalement, par le biais des corrélations pouvant être – ou non – trouvées entre les constituants du *Big Data*, d'un plaidoyer pour une société *data driven* [conduite par les données]. A cet égard, il est intéressant de lire Alain Supiot – ou d'écouter ses leçons au Collège de France – sur le thème : *Du gouvernement par les Lois à la gouvernance par les nombres*¹¹

Comment échapper à la curatelle du *Big Data* ?

Sauf à faire l'éloge de la fuite et de la déconnection totale de tout réseau, et à en revenir aux coursiers *porteurs de messages* en lieu et place des courriels¹², nous serons peu ou prou soumis de plus en plus à la curatelle du *Big Data*.

Est-ce à dire que nous devons dévorer bouche bée tout ce qui en provient à partir de la seule recherche de corrélations entre ses contenus ? Il est clair que nous n'avons pas la connaissance précise, et encore moins la maîtrise, sur les algorithmes utilisés par un moteur de recherche comme celui de Google¹³.

⁸ ANDERSON C., "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete", Wired Magazine, 23 juin 2008

⁹ Mayer-Schonberger V., Cukier K., *Big data, a revolution that will transform how we live, work and think*, New York, 2013, 242p.

¹⁰ Ibid., 2013, p.71

¹¹ SUPIOT A., *La Gouvernance par les nombres*, Cours au Collège de France en 2012-2014, Paris, Fayard, 2015, 512p.

¹² C'est la méthode que recommande, en lieu et place des mails pour toute communication importante, le nouveau président des USA, Donald Trump.

¹³ En l'espèce *MapReduce*, algorithme avec lequel sont effectués des calculs parallèles, et souvent distribués, de données potentiellement très volumineuses, typiquement supérieures en taille à 1 téraoctet. On peut en trouver des descriptions sommaires dans la littérature spécialisée.

En écartant les objectifs des thuriféraires pour lesquels le *Big Data* signifie essentiellement *Big Money*, le chercheur peut s'intéresser à des travaux qui constituent des avancées intéressantes dans les domaines suivants :

- La reconnaissance des formes ;
- L'extraction de connaissance à partir de documents textuels.

Pour la reconnaissance des formes, les réseaux de neurones artificiels dans leurs versions les plus récentes – et qui sont soumis par leurs auteurs à des concours comparatifs quant à leurs résultats - on peut citer parmi d'autres les travaux du français Yann LeCun. On peut suivre ses leçons prononcées au Collège de France, où il a été titulaire de la chaire *Informatique et Sciences numériques*¹⁴ en 2015-2016. Par analogie avec le système de vision des chats et singes étudié par Hubert et Wiesel¹⁵, LeCun¹⁶ et ses collègues ont développé des réseaux neuronaux multicouches dans lesquels on peut observer des généralisations à partir de figures élémentaires telles que celles colletées dans les cellules de la rétine, et transmises aux *colonnes de neurones spécialisés* qui se trouvent dans le cerveau occipital des mammifères. Il ne s'agit pas encore d'extractions de connaissances de nature conceptuelle, mais ces travaux, en plein essor, font partie de ceux consacrés à l'apprentissage par les machines (*machine learning*),

S'agissant de l'extraction de connaissance à partir de documents textuels – et plus généralement de ce qu'il est convenu de nommer les *humanités numériques* – on peut examiner une publication de Daniel Jurafsky¹⁷. Le modèle proposé est de type vectoriel, et permet d'obtenir, selon les auteurs, à partir de documents multiples sans lien apparent entre eux, mais numérisés et contenus dans un *Big Data*, non seulement des réponses à des questions posées en langage naturel, mais aussi de produire des résumés et des textes de synthèse d'articles, voire de transcriptions de réunions.

Dans les deux types de travaux indiqués sommairement ci-dessus, il est nécessaire, pour valider les résultats exhibés, de disposer de grandes quantités de données, qui serviront pour moitié à *l'apprentissage* des systèmes proposés, et pour l'autre moitié à les tester. Le résultat de ces tests se traduira par un pourcentage plus ou moins élevé de bons résultats. Typiquement, un système de reconnaissance de formes doit être capable de reconnaître, comme étant une voiture, une automobile qu'il n'a pas rencontrée pendant la phase d'apprentissage, et cela avec une marge d'erreur aussi faible que possible.

Conclusions

Nous avons tenté, dans ce court article, et sans entrer dans des détails techniques, de montrer que le phénomène *Big Data* laisse dans l'ombre une *curatelle technico-financière* – c'est-à-dire une mise sous tutelle - à laquelle sont soumis tous

¹⁴ <http://www.college-de-france.fr/site/yann-lecun/inaugural-lecture-2016-02-04-18h00.htm>

¹⁵ HUBEL D et WIESEL T., "Receptive fields and functional architecture of monkey striate cortex", J Physiol. (1968) : 195, pp. 215-243

¹⁶ LECUN Y., BENGIO Y., "Deep learning", Nature (2015), 521, pp.436-444

¹⁷ JURAFSKY D. et MARTIN JH, *Speech and language processing: An introduction to natural language processing, computational linguistics and speech recognition*, Prentice Hall, Englewood Cliffs, 2007, 950p.

ceux – individus ou organisations - qui se connectent - volontairement ou non – à un réseau. Ce qui a pour conséquence une augmentation exponentielle du contenu des centres de stockage (plus de 30.000 à travers le monde à l'heure où nous écrivons). Ces *data center*, qui hébergent les *Big Data*, ne sont pas accessibles sans frais et en toute transparence par ceux-là mêmes qui les alimentent pourtant gratuitement, car ils sont exploités pour l'essentiel, à des fins marchandes, par leurs propriétaires.

Une société *data driven* – c'est-à-dire *conduite par les nombres du Big Data* – s'est ainsi mise en place à l'insu de notre plein gré. C'est-à-dire une société qui n'est plus gouvernée par des lois, mais est celle de la gouvernance par les nombres. Ce qui conduit à l'interroger sur ce qui subsiste de démocratie et de pouvoir politique.

Ayant volontairement laissé de côté les bienfaits supposés du *Big Data*, nous avons indiqué des travaux à caractère scientifique - reconnaissance de formes, analyses de documents textuels – qui font l'objet de recherches prometteuses avec déjà des résultats intéressants et vérifiés. Les questions éthiques que posent la collecte massive d'informations, et les pratiques qu'elle peut induire, sont cependant loin d'avoir reçu des réponses satisfaisantes.

Bibliographie

Nous reprenons ci-après, dans l'ordre alphabétique des premiers auteurs, les ouvrages et articles cités dans notre texte.

- ANDERSON C., "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete", *Wired Magazine*, (2008), vol. 16 n° 07.
- DEPARCIEUX A., *Essai sur les probabilités de la durée de vie humaine*, 1746
- ELLUL J., *Le système technicien*, Paris, Cherche Midi, 2012, 334p. – Réédition
- Mayer-Schonberger V., Cukier K., *Big data, a revolution that will transform how we live, work and think*, New York, 2013, 242p.
- HUBEL D et WIESEL T., "Receptive fields and functional architecture of monkey striate cortex", *J Physiol.* (1968) : 195, pp. 215-243
- JURAFSKY D. et MARTIN JH, *Speech and language processing: An introduction to natural language processing, computational linguistics and speech recognition*, Prentice Hall, Englewood Cliffs, 2007, 950p.
- LECUN Y., BENGIO Y., "Deep learning", *Nature* (2015), 521, pp.436-444
- SUPIOT A., *La Gouvernance par les nombres*, Cours au Collège de France en 2012-2014, Paris, Fayard, 2015, 512p.