



HAL
open science

Une brève histoire d'Internet et le rôle de la Data dans la société contemporaine.

Bruno Teboul

► **To cite this version:**

Bruno Teboul. Une brève histoire d'Internet et le rôle de la Data dans la société contemporaine.. 2016. hal-01423627

HAL Id: hal-01423627

<https://hal.science/hal-01423627v1>

Preprint submitted on 6 Jan 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.

Paris, le 15 décembre 2016.

Texte issu d'une conférence donnée pour UTOPIES le 15 décembre 2016 à Paris.

Une brève histoire d'Internet et le rôle de la Data dans la société contemporaine.

Bruno Teboul^{1 2 3}

¹ Université Paris Dauphine (PSL), Laboratoire DRM, Dauphine Recherches en Management.

² Membre de la Gouvernance de la Chaire Data Scientist de l'École Polytechnique.

³ Groupe Keyrus, Senior Vice-President Science & Innovation.

Résumé

Cet article se propose de remonter aux origines d'Internet, sans prétendre en reconstruire l'histoire complète, mais en se focalisant sur les grands tendances et les facteurs explicatifs les plus saillants, communément admis dans la littérature; et afin de mieux comprendre le phénomène et son lien avec l'avènement du « Big Data » (ou *Mégadonnées*). La première partie de l'article aborde l'origine d'Internet et le rôle et l'influence de la *Cybernétique*, de la *Contre-Culture*, du *Creative Ethos* et de la *Creative Class*. Nous décrivons dans ce contexte, le passage du projet « Arpanet » à la naissance du « World Wide Web », et ses conséquences sur l'explosion des données numériques: que nous résumons par la formule « *du Dataclysme aux traitements des Mégadonnées* ». Dans la seconde partie de l'article, nous tentons de donner une définition aux données (Data) et d'attribuer un rôle à la Data. Nous nous inspirons ici de la *philosophie de l'information* et rappelons la *définition diaphorique d'une donnée*. La troisième partie de l'article analyse les conséquences des Mégadonnées sur la société contemporaine: telles que les limites à l'anonymisation des données, les dangers d'une tentative de « *réduction datamétrique* » de la personne, ou encore la critique de la « *société du calcul* » et de la « *gouvernance algorithmique* » à l'ère de la « *data-driven society* ».

Abstract

This paper proposes to go back to the origins of the Internet, without pretending to reconstruct its complete history, but focusing on the major trends and the most salient explanatory factors commonly accepted in literature. And to better understand the phenomenon and its link with the advent of the "Big Data". The first part of the article discusses the origin of the Internet and the role and influence of *Cybernetics*, *Counter-Culture*, *Creative Ethos* and *Creative Class*. We describe in this context the transition from the "Arpanet" project to the birth of the "World Wide Web", and its consequences on the explosion of digital data: which we summarize by the formula "*From Dataclysm to Big Data treatments and analysis*". In the second part of the article, we try to give a definition to the data and assign a role to the Data. We draw here from *the philosophy of information* and recall the *Diaphoric Definition of a Data*. The third part of the article analyzes the consequences of Big Data on contemporary society: such as limits to the anonymization of data, the dangers of an attempt at "*datametric reduction*" of individual person, or the criticism of a "*computation society*" and an "*algorithmic governance*" in the era of the "*data-driven society*".

1] D'Internet aux Mégadonnées...

1.1 A l'origine d'Internet:le rôle et l'influence de laCybernétique...

Comme l'indique Monique Dagnaud¹, on peut considérer qu'Internet est né aux Etats-Unis, en Californie dans un contexte très particulier où se sont croisés et mélangés plusieurs facteurs à l'origine de la naissance conjugée de la théorie de l'information et de la communication (dans les années 1940), de la cybernétique, de l'apparition des premiers ordinateurs personnels (vers 1960-1970), d'une certaine contre-culture que l'auteure résume parfaitement en trois points: « *free, free speech, free of charge* ». Libre (*free*) comme l'idée de l'autonomie de l'individu sublimée grâce à la micro-informatique notamment, la liberté de circulation de l'information (*free speech*) est rendue possible par les autoroutes de l'information et la promesse d'une communication en réseau et sans limite; le partage désintéressé du contenu informationnel lui-même (*free of charge*).

Comme le rappelle également Monique Dagnaud, « *Norbert Wiener publia un ouvrage en 1950, Cybernétique et Société, consacré aux moyens d'éviter une dérive autoritaire des pays démocratiques [...]. Cette vision imprégna la contre-culture des années 1960* ».²

Héritier de la première et de la seconde « *Cybernétique* »³, *Internet* est d'abord pensé comme un système sociotechnique qui veut révolutionner l'ordre social pour échapper à l'aliénation d'une société automatisée, inhumaine en proie à la guerre afin de redonner à la culture et à la diffusion des savoirs un rôle civique et politique. Où les machines seraient pensantes, bienveillantes pour tous les humains. Le terme « *Cybernétique* » a été forgé à partir du grec « *kubernêsis* », qui signifie, au sens propre, « *action de manœuvrer un vaisseau* », et au sens figuré, « *action de diriger, de gouverner* ». Utilisé pour la première fois, par Ampère pour désigner l'art du gouvernement, comme l'indique Paul Mengal: « *c'est dans ce sens que le mot cybernétique fait son apparition dans la langue française sous la plume d'André-Marie Ampère (1775-1836), physicien, fondateur de l'électrodynamique. En dehors de ses intérêts pour la physique et la chimie, Ampère s'est également passionné pour la philosophie des sciences et c'est dans un ouvrage intitulé Essai sur la philosophie des sciences, ou Exposition analytique d'une classification naturelle de toutes les connaissances humaines que l'on trouve le mot cybernétique* »⁴.

Le concept de « *cybernétique* » a été mis en circulation, dans sa signification actuelle, par le célèbre ouvrage de Norbert Wiener: *Cybernetics or Control and Communication in the Animal and the Machine* de 1948. Dans cet ouvrage, Weiner définit la cybernétique comme la « *science constituée par l'ensemble des théories relatives aux communications et à la régulation dans l'être vivant et la machine* ». C'est ainsi qu'il nous introduit non seulement au

¹ Monique Dagnaud, Le Modèle Californien, Comment l'esprit collaboratif change le monde, Mai 2016, Odile Jacob.

² Ibidem, page 21.

³ Marvin Minsky présente dans *Semantic Information Processing*, en 1969 la première cybernétique comme un tronc commun qui se serait divisé en trois branches: la « *simulation cognitive* » à la Allen Newell et Herbert Simon, l'« *intelligence artificielle* » et la « *seconde cybernétique* » ou théorie des systèmes auto-organiseurs.

⁴ Mengal, Paul. « *Cybernétique, histoire d'un mot* ». *Res Publica* 18 (1998): 1946-1953.

nouveau paradigme cybernétique, mais aussi à un discours nouveau, une véritable critique du modernisme et de l'idée occidentale du progrès.

Dans cet ouvrage fondateur, Wiener fait la synthèse de toute une série de recherches poursuivies dans le domaine des mathématiques pures (théorie de la prédiction statistique), dans le domaine de la technologie (machines à calculer, télécommunications), dans ceux de la biologie et de la psychologie, et jette les bases d'une science nouvelle, à support mathématique, destinée à couvrir tous les phénomènes qui, d'une manière ou d'une autre, mettent en jeu des mécanismes de traitement de l'information. C'est avec la création des ordinateurs et des questionnements qui vont s'ensuivre que va naître la cybernétique. Le mérite de Wiener a surtout été de poser le premier la question de l'homme à travers le paradigme de l'information. En effet, pour lui l'homme est réductible à un ensemble d'informations communes à tous et à quelques informations particulières à chacun.

Dès lors, l'information et son transport prennent tout leur sens et la connexion de l'homme et de la machine devient à l'évidence une étape de cette informatisation de l'homme. La cybernétique part de l'idée que l'organisation de la machine comme un cerveau permettrait à terme de produire une intelligence artificielle. L'ordinateur comme machine artificielle disposant d'une mémoire virtuellement infinie et d'une capacité d'apprentissage devait donc lui permettre de gérer à terme plus rationnellement que toute autre machine, un système, une entreprise voire une société.

Cette logique s'inscrit pleinement dans le cadre d'une obsession de Wiener: l'entropie. L'entropie est selon lui, la tendance de la nature à défaire et à désorganiser ce qui était organisé et cohérent, la mort n'étant que l'ultime étape de celle-ci. Les machines organisent et permettent la communication qui est le seul moyen de lutter contre celle-ci. La multiplication des messages et donc de la communication devant permettre d'accroître la cohérence et représente à l'évidence l'axe fondamental selon lequel devait se développer les sociétés.

Cette vision s'élève contre le secret ou la commercialisation de l'information et présuppose un libre et vaste accès aux informations ce qui représente aujourd'hui encore une vision assez utopique. C'est pourtant de cette pensée que se réclament les hackers, pirates d'un nouvel océan de données. Selon eux, le libre accès est une donnée fondamentale du réseau et doit être préservé, par leur action parfois. Cette société de réseaux de données, des autoroutes de l'information, Wiener l'appelait de ses vœux car elle aurait dû permettre une véritable fluidité, qui seule pourrait assurer à la société une stabilité nouvelle.

C'est ainsi que Fred Turner souligne le rôle fondamental de la cybernétique et notamment de la pensée de Norbert Wiener chez Stewart Brand et son rôle dans l'avènement d'Internet: *« qui conçoit les objets techniques et les humains comme un même système sociotechnique. Dessinant un monde où l'information modèle l'ordre social, la cybernétique a fondé la réflexion du jeune Stewart Brand sur la société et les médias. Avec la perspective terrifiante de l'apocalypse nucléaire et entourée d'appareils, la génération de Stewart Brand percevait la société comme une machine automatisée pour laquelle il faudrait « peser de tout [son] corps sur ses engrenages » selon l'expression de Mario Savio. Face à la crainte de l'aliénation de la culture adulte, deux mouvements sociaux distincts ont émergé de cette période : la Nouvelle Gauche, qui lutte pour les droits civiques et se mobilise autour de l'action politique, et les Nouveaux Communistes, un mouvement introverti tourné vers la conscience et*

l'intimité. De manière surprenante, ces derniers vont adopter les pratiques sociales collaboratives et l'euphorie technologique des laboratoires de recherche militaire. La cybernétique permettait aux habitants des communautés alternatives de rêver d'un mode de vie « sous la haute surveillance de machines pleines d'amour et de grâce »⁵.

1.2 A l'origine d'Internet:Contre-Culture, Creative Ethos etCreative Class...

C'est dans les années 1970, que le concept de contre-culture fut inventé, à une époque donc où des mouvements de contestations naissent, incarnés par les « hippies, les beatniks, les militants pacifistes, les étudiants, la Nouvelle Gauche américaine, les activistes noirs, les écologistes, les féministes, etc. L'impulsion protestataire est alors générale et ébranle tous les cadres de la société traditionnelle. Plutôt qu'une simple opposition systématique à la société, cherchant des solutions réformatrices, la dissidence des années soixante imagine un autre modèle complet. Refusant de s'accommoder ou de se résigner à l'ordre en place, cette minorité radicale manifeste sa volonté d'en finir avec lui et de lui substituer un système fondé sur des logiques de liberté, de gratuité, d'égalité, de solidarité et de pleine démocratie »⁶.

La théorie de Richard Florida est inspirée des travaux de Robert Lucas (Prix Nobel d'économie en 1995, « *anticipations rationnelles* ») et Jane Jacobs (philosophe de l'urbanisme), qui stipulent que c'est le *capital humain* des classes créatives qui enrichit les régions, et non pas les entreprises technologiques qui s'y installent. Ces entreprises technologiques ne feraient que suivre les classes créatives. Ainsi elles choisiraient même de s'installer dans une région, comme la Silicon Valley, suivant trois critères fondamentaux: les talents (le taux de personnes éduquées), les technologies (dynamisme de la région dans les industries technologiques) et la tolérance (indiquée par le nombre d'artistes, musiciens et la taille de la communauté gay).

Dans *The Rise Of the Creative Class*, Richard Florida réfléchit à ce qui pousse la classe créative (les ingénieurs, designers, artistes, architectes...) à se concentrer dans certaines villes d'Amérique du Nord au détriment d'autres. L'ouvrage de Richard Florida insiste aussi sur la dimension multiculturelle des régions dynamiques, établissant une relation directe entre le dynamisme de la scène musicale ou de la communauté gay, le taux d'immigration ou la variété des communautés ethniques et la richesse générée par les sociétés technologiques. Dans le cadre de cette réflexion, il remonte à la notion de « *creative ethos* », cette pulsion créative qui innerve nos sociétés de la connaissance et qui est notre source de richesse.

Selon Florida, le « *creative ethos* » contemporain naît de la rencontre entre a) la contre-culture hippie qui s'est installée à San Francisco et b) l'éthique protestante du travail (*protestant work ethic*). Cette jonction improbable se produit à la fin des années soixante en

⁵Samuel Goëta, « Fred Turner, Aux sources de l'utopie numérique. De la contre-culture à la cyberculture, Stewart Brand un homme d'influence », Questions de communication, 2013.

⁶Steven Jezou-Vannier, Contre-culture(s) – Des Anonymous à Prométhée, Attitudes, Le mot et le reste, 2013.

Californie. C'est dans cet esprit qu'apparaissent des technophiles et entrepreneurs aussi inventifs que Steve Jobs, Paul Allen, Steve Wozniak ou encore Gordon Moore.

1.3 De "Arpanet" au "World Wide Web" ...

C'est dans ce contexte très particulier qu'Internet va naître et se développer aux Etats-Unis, comme une nouvelle utopie capable de forger une nouvelle société; un nouvel espoir pour des générations attachées aux idées de liberté, de paix, de collectif, fascinés par les promesses de la communication technologique. Ces technologies elles-mêmes soutenues par les programmes gouvernementaux de l'époque comme l'atteste la création du Information Processing Techniques Office (IPTO), nommé à l'origine « Command and Control Research », une section du Département de la Défense américaine, la très active DARPA (Defense Advanced Research Projects Agency).

« La naissance d'Internet s'est effectuée sous la juridiction d'une institution gouvernementale tout à fait spécifique créée à l'intérieur du DoD dans les années soixante... Cette institution est le bureau des techniques de traitement de l'information (IPTO)⁷. Durant ses 24 années d'existence (1962-1986), l'IPTO a rendu possible un grand nombre de percées technologiques qui ont fondamentalement changé la nature de l'informatique. Ces percées technologiques comprennent l'informatique conversationnelle, le temps partagé, les interfaces graphiques interactives, la commutation par paquet et la création de l'ARPANET et, peut-être le plus important, la création d'Internet. Le fondateur de ce bureau fût JCR Licklider qui créa IPTO en 1962. Durant son premier mandat de directeur, il a été capable d'imprimer une direction vers un programme de recherche informatique étendu et visionnaire qui incluait la création de la communauté de chercheurs de l'IPTO. Licklider a appelé cette communauté scientifique hors normes qui collaboraient les uns avec les autres l'« Intergalactic Network »⁸.

L'*Intergalactic Network*, comme communauté scientifique et projet, retentit aujourd'hui en véritable prémonition du web annoncée par Licklider. On sait désormais que l'Internet naîtra vingt ans plus tard. Alors attachons-nous maintenant à comprendre comment s'est articulé le passage d'*Internet* au *World Wide Web*.

Selon Wikipedia, « le terme d'origine américaine « Internet » est dérivé du concept d'*interneting* (en français: « interconnecter des réseaux ») dont la première utilisation documentée remonte à octobre 1972 par Robert E. Kahn⁹ au cours de la première ICC (International Conference on Computer Communications) à Washington. Les origines exactes du terme « Internet » restent à déterminer. Toutefois, c'est le 1er janvier 1983 que le nom « Internet », fut utilisé officiellement pour désigner l'ensemble d'ARPANET et plusieurs réseaux

⁷The Information Processing Techniques Office (IPTO), originally "Command and Control Research, was part of the Defense Advanced Research Projects Agency of the United States Department of Defense.

⁸Ronda Hauben, « À la recherche des pères fondateurs d'Internet. Pourquoi a-t-on besoin d'une histoire d'Internet? », *Multitudes* 2003/1 (no 11), p. 193-199.

⁹<http://www.cnri.reston.va.us/bios/kahn.html>

*informatiques*¹⁰ ». En France, un projet nommé Cyclades, inspiré par ARPANET fut dirigé par Louis Pouzin. Pouzin et Kahn indiquent que le protocole TCP/IP fut inspiré par le réseau Cyclades français, poussé par la CII et sa Distributed System Architecture: on parle alors déjà d'architecture de calcul distribuée.

L'internet précède de plusieurs années l'avènement du « *World Wide Web (WWW)* » qui signifie littéralement la « *toile (d'araignée) mondiale* ». Le *Webest* traduit en français par l'expression « *la Toile* », qui est un système *hypertexte* public fonctionnant sur Internet.

*« Le Web permet de consulter, avec un navigateur, des pages accessibles sur des sites. L'image de la toile d'araignée vient des hyperliens qui lient les pages web entre elles. Le Web n'est qu'une des applications d'Internet; distincte d'autres applications comme le courrier électronique, la messagerie instantanée, et le partage de fichiers en pair à pair. Le Web a été inventé par Tim Berners-Lee et Robert Cailliau plusieurs années après Internet, mais c'est lui qui a rendu les médias grand public attentifs à Internet. Depuis, le Web est fréquemment confondu avec Internet; en particulier, le mot Toile est souvent utilisé dans les textes non techniques sans qu'il désigne clairement le Web ni Internet »*¹¹.

Après ce bref rappel historique de l'avènement d'Internet et du Web, nous allons maintenant tenter d'établir les liens entre l'Internet, l'explosion du Web et la prolifération des données comme conséquence.

1.4 L'explosion des données numériques: du *Dataclysm* aux traitements des *Mégadonnées*...

En effet, l'accélération exponentielle du réseau des réseaux a généré un phénomène controversé, mais bien réel: les « *mégadonnées* » que l'on nomme communément « *Big Data* ». L'expression « *Big Data* » date de 1997 selon l'Association for Computing Machinery. En 2001, Doug Laney, grand analyste du cabinet Meta Group (devenu Gartner) décrivait les *big data* selon le principe des « trois V »:

- le Volume de données de plus en massif
- la Variété de ces données qui peuvent être brutes, non structurées ou semi-structurées
- la Vitesse qui désigne le fait que ces données sont produites, récoltées et analysées en temps réel

Depuis, 3 autres V ont été ajoutés tels que la Visibilité des données, la Vérité des données et la Valeur des données tant le phénomène s'accélère et est difficile à appréhender même pour les scientifiques des données: les *data scientists*¹².

¹⁰http://www.circleid.com/posts/a_closer_look_at_the_controversy_over_the_internets_birthday_you_decide

¹¹https://fr.wikipedia.org/wiki/World_Wide_Web.

¹² CF l'ouvrage « *La Donnée n'est pas donnée* », Actes du Colloque Ecole Militaire, publié chez Kawa en juillet 2016 et notamment l'article de Bruno Teboul et Thierry Berthier « *Valeur et Vérité de la donnée: Enjeux pour l'entreprise et défis pour le Data Scientist* ».

Effectivement, depuis l'avènement d'internet et du web, la numérisation du monde a accompli son destin planétaire. Cette numérisation est le fruit de fulgurantes avancées et découvertes scientifiques, de ruptures technologiques sans précédent qui ont rendu possibles la virtualisation, la dématérialisation des biens, des services, des marchandises, des échanges et ont ainsi modifiés le paysage économique aux Etats-Unis, en Europe et en Asie grâce notamment au développement des technologies de l'artificiel et de l'immatériel.

Cette numérisation galopante, omniprésente dite « pervasive » paraît sans limite et explique la prolifération de données laissées sur le web par désormais 3 milliards d'internautes en moins d'un demi-siècle. Ces données sont protéiformes, à 80% elles dites sont *non-structurées* (textes, photos, vidéos...), ou non relationnelles, non rationnelles, difficiles à stocker dans des bases de données classiques (*d'où le développement et l'utilisation de nouveaux protocoles de bases de données dits noSQL*).

La croissance exponentielle des capteurs intelligents (« *smart sensors et smart devices* ») l'explosion du mobile et de ses applications, des réseaux sociaux (« *user generated content* » notamment), la tendance à « *l'autométrisation* » (« *quantified-self* »), ont contribué à former la « *datafication* » du monde, ainsi que la « *plateformisation* » de l'économie (« *microtasking* », « *digital labor* »). *Datafication* et *algorithmisation* résultent d'un changement de paradigme lié à l'exploitation des mégadonnées: le déluge de données fait l'objet de traitement par des algorithmes de collecte, de catégorisation, d'analyse, de visualisation des données massives et multi-structurées (mégadonnées).

La *datafication* et *l'algorithmisation* du monde provoquent des bouleversements sans précédents sur le plan économique et social. Cette véritable mutation économique et sociétale provoque une évolution du capitalisme lui-même. En effet, avec le développement de nos économies contemporaines de plus en plus fondées sur l'information sous toutes ses formes, les données sont devenues la nouvelle matière première et constituent la forme de capital la plus recherchée: nous assistons à l'évolution du capitalisme industriel, entrepreneurial, puis financier, vers un *capitalisme cognitif* pour reprendre l'expression de Yann Moulier Boutang¹³.

L'œuvre remarquable de Moulier Boutang permet de comprendre comment le capitalisme, fondé sur l'échange et la production bascule dans une économie basée sur des mécanismes et des processus de *pollinisation*, de *contribution* et de *partage*, à l'ère de la donnée, de l'information et de la connaissance.

Le phénomène *Big data* s'est imposé en raison de l'augmentation exponentielle des données numériques, comme le montre la revue *Nature* qui, en 2008, popularisa définitivement l'expression « *Big Data*¹⁴ ».

« *L'époque des big data constitue un défi pour notre mode de vie et modifie notre interaction avec le monde. Son plus grand impact ? Quand la société va se rendre*

¹³ Voir notamment Yann Moulier Boutang « Le Capitalisme cognitif » 2011 et « L'économiste et l'abeille » 2014.

¹⁴ « The Internet search firm Google was incorporated just 10 years ago this week. Going from a collection of donated servers housed under a desk to a global network of dedicated data centers processing information by the petabyte, Google's growth mirrors that of the production and exploration of data in research. All of which makes this an apt moment for this special issue of *Nature*, which examines what big data sets mean for contemporary science.» (Editorial, « Community cleverness required », *Nature*, 455 (2008).

compte qu'elle doit mettre un bémol à son obsession de la causalité et se fonder sur de simples corrélations : il ne s'agit plus de connaître le pourquoi, mais seulement le quoi. Des siècles de pratiques vont être ainsi bouleversés. Car la nouvelle démarche remet en question la façon dont nous concevons fondamentalement la prise de décision et l'appréhension de la réalité »¹⁵.

Ces immenses flots de données, cet océan de data ou « *dataclysm* » croît toujours plus, à mesure que le cyberspace s'étend, le volume d'informations numérisées n'en finit plus d'augmenter. S'il paraît évident que le nombre d'internautes a explosé en deux décennies, les chiffres sont tous aussi éloquents. D'à peine trois millions en 1990, les internautes sont passés à environ deux milliards en 2010, sont trois milliards cette année (2016) et nous prévoyons près de cinq milliards d'internautes en 2020. Cette masse d'utilisateurs est également un immense réservoir de producteurs de contenus, c'est à dire de données, d'informations et de connaissances.

Selon l'institut IDC, 1,8 zettaoctet de données (Zo, 10²¹ octets) a été créé en 2011. « *L'information disponible à la surface de notre planète en 2020 devrait tourner autour des 40 Zo... Mais ces estimations sont rendues fausses d'année en année par les nouveaux usages* », précise Jean-Yves Pronier¹⁶, directeur marketing de EMC. Les données croissent vraiment de manière exponentielle... Mais quelle définition pourrions-nous donner au concept de donnée afin de mieux en comprendre le rôle et les enjeux dans notre société contemporaine ?

2] Définition et rôles de la data

2.1 Premières tentatives de définitions de la Data.

« Datum vient du latin, dont le pluriel est data. Cependant, en anglais, suite aux siècles, on utilise data comme nom indéénombrable et donc le verbe qui le suit est au singulier. Certains latinistes préfèrent utiliser le verbe au pluriel »¹⁷.

Le Dictionnaire de l'information¹⁸ définit la donnée comme ce qui est donné, déterminé, connu ou reconnu. Hannson (in Foray et coll., 2002) part de cette forme la plus rudimentaire de l'information qu'est la donnée pour établir cette double distinction : « Il faut que les données soient assimilables pour pouvoir constituer de l'information et qu'elles soient assimilées pour constituer du savoir ».

Bohn (1994)¹⁹ opère quant à lui, une distinction entre donnée, information et connaissance. Les données sont des éléments provenant des capteurs, elles sont relatives au niveau mesuré

¹⁵Victor MAYER-SCHÖNBERGER, Kenneth CUKIER, *Big Data. A Revolution that Will Transform How We Live, Work, and Think*, op. cit. [traduction de H. Dhifallah, Kindle, Empl. 133-137].

¹⁶http://www.liberation.fr/economie/2012/12/03/donnees-le-vertige_864585

¹⁷<https://fr.wiktionary.org/wiki/datum>

¹⁸ Dictionnaire de l'information, Editions Armand Collin (3^{ème} édition), 2008

¹⁹Bohn, R.E. (1994), « Measuring and Managing Technical Knowledge », Sloan Management Review, Fall 94 pp.61-73

d'une variable quelconque. L'information consiste en des données organisées dans une structure donnée et qui, placée dans un contexte, est dotée de sens. La connaissance va plus loin: elle permet de faire des prédictions, d'établir des liens de causalité ou de prendre des décisions à propos du « Que faire ? ».

Prax²⁰ définit la donnée comme un « *fait discret et objectif brut, résultant d'une observation, d'une acquisition ou d'une mesure effectuée par un instrument naturel ou artificiel. Cette donnée peut-être qualitative (le ciel est bleu) ou quantitative (la température est de 20°). Il n'y a normalement pas d'intention ni de projet dans la donnée, c'est ce qui lui confère son caractère d'objectivité* ». En mathématique, la donnée est une quantité connue qui sert à découvrir les inconnues, comme les données d'un problème. En informatique, on considère la donnée comme une représentation d'une information sous une forme conventionnelle destinée à faciliter son traitement.

2.2 Philosophie de l'information et définition diaphorique de la donnée.

Luciano Floridi, Professeur de Philosophie et Directeur de recherche au OxfordInternet Institute s'interroge quant à la possibilité de définir sémantiquement la donnée et étudie le rôle et la définition de l'information fondée sur les données: « *a data-based definition of information* »²¹. « *Cependant, il est à noter que le concept de données paraît à première vue assez imprécis, risquant de déplacer la difficulté au lieu de la lever. Luciano Floridi estime pourtant qu'il est moins riche et glissant que celui d'information et donc plus facile à manipuler. Aussi, une définition fondée sur les données fournit un bon point de départ* »²².

Selon la *Définition Générale de l'Information* (General Definition of Information) à laquelle se réfère également Luciano Floridi, il n'y a pas d'information sans donnée, sans data. La donnée n'est pas une information, contrairement à ce qu'on peut lire ici ou là. Certes, une information peut être transformée en donnée mais, comme l'écrit Pierre Delort, « *l'information est une séquence de signes pouvant être interprétés, tandis que la donnée est une valeur variable* »²³. L'auteur fait ici implicitement référence à la *General Definition of Information* (GDI), qui envisage l'information comme un composé de données et de sens (« *Data + meaning* »). Elle se transcrit en « *donnée + signifiante* », l'anglais *meaning* étant traduit ici par signifiante. C'est exactement la théorie qu'élabore Luciano Floridi avec l'énoncé suivant: « *la DGI doit considérer que σ est un exemple d'information, doté d'un contenu sémantique, si et seulement si* »:

– DGI 1 : σ consiste en une ou plusieurs données

²⁰Prax, J-Y., *Le manuel du knowledge management, une approche de deuxième génération*, Dunod, Paris, 2003, p.60

²¹Luciano Floridi, « *Semantic Conceptions of Information* », *Stanford Encyclopaedia of Philosophy*, 2005. Article accessible en ligne à l'adresse suivante: <http://plato.stanford.edu/entries/information-semantic/>

²²Sylvie Leleu-Merviel, Philippe Useille. *Quelques révisions du concept d'information*. Hermes. Problématiques émergentes dans les sciences de l'information, Lavoisier, pp.25-56, 2008, *Traité des sciences et techniques de l'information*, 978-2-7462-2110-9. <hal-00695777>.

²³Voir Pierre DELORT, *Le Big data, op. cit.*, Kindle, Empl. 57-58.

– DGI 2 : la donnée σ est formée correctement

– DGI 3 : la donnée σ , formée correctement, est porteuse de signifiante

Ainsi, il y aurait trois réquisits pour l'existence d'une information : a) il faudrait une ou plusieurs données ; b) que ces données soient bien formées (« *well-formed* »), c'est-à-dire assemblées selon certaines règles ; c) et qu'elles soient porteuses de sens (« *Meaningful* »), c'est-à-dire à même d'être interprétées ou, si l'on préfère, aptes à être traduites, ou exprimées autrement. La donnée correspond, quant à elle, à ce qui n'est pas uniforme (à ce qui est « *diaphora*/différent ») ; bref il s'agit d'une variable: « *une donnée est un fait supposé qui procède d'une différence ou d'un manque d'uniformité dans un contexte*²⁴ ». Il s'ensuit, comme on a déjà pu le remarquer, que la donnée est par conséquent une entité relationnelle²⁵.

Luciano Floridi, est le philosophe contemporain qui a le mieux contribué à l'essor de la *philosophie de l'information*. Son éclairage théorique permet de comprendre comme une donnée est réductible à un manque d'uniformité. La data est une « *diaphora* » au sens grec, ce qui signifie qu'elle est avant tout une « *différence* ». Floridi poursuit en indiquant que la Définition Diaphorique de la Data(DDD) peut être appliquée suivant trois niveaux, l'un rendant possible le suivant:

- Il distingue d'abord la Data comme « *diaphora de re* », c'est-à-dire comme manque d'uniformité dans le monde réel. Il n'existe pas de nom spécifique pour de telles «données dans la nature». Une suggestion possible est de se référer aux données comme des « *dédomen* » (traduction en grec ancien de «données»). On peut noter également d'un point étymologique, que le mot «données» vient de la traduction latine d'une œuvre d'Euclide « *Dedomena* ». On ne peut connaître directement la donnée, mais simplement l'inférer à partir de l'expérience.

Luciano Floridi explique que les « *Dedomena* » sont des données pures ou des données proto-épistémiques, c'est-à-dire des données avant qu'elles ne soient épistémiquement interprétées. En tant que «*fractures dans la fabrique de l'être*», elles ne peuvent être posées que comme un ancrage externe à notre information, car les « *dédomènes* » ne sont jamais accessibles ni élaborés indépendamment d'un certain niveau d'abstraction. Elles peuvent être reconstruites comme des exigences ontologiques, comme le « *noumène* » chez Kant²⁶ ou la « *substance* » chez Locke. Elles ne sont pas épistémiquement expérimentées, mais leur présence est empiriquement déduite de (et requise par) l'expérience.

- Les données définies comme « *diaphora de signo* », c'est-à-dire révélant des manques d'uniformité dans la perception de deux états physiques, tels qu'une charge

²⁴« A datum is a putative fact regarding some difference or lack of uniformity within some context » (Luciano FLORIDI, « Semantic Conceptions of Information », *art. cit.*, p. 9.)

²⁵« A datum is a relational entity » (Luciano FLORIDI, *ibid.*, p. 11). – Pour une discussion de la GDI, voir Sylvie LELEU-MERVIEL, Philippe USEILLE, « Quelques révisions du concept d'information », dans Fabrice PAPY (dir.), *Problématiques émergentes dans les sciences de l'information*, Paris, Lavoisier-Hermès, 2008, p. 25-56.

²⁶Le *noumène* dans la philosophie kantienne, désigne la chose en soi, la réalité absolue qui demeure hors de portée quand le phénomène renvoie à ce qui est saisi par les sens.

supérieure ou inférieure dans une batterie, une variation dans le signal électrique durant une conversation téléphonique.

- Les données comme « *diaphora de dicto* », ce sont les données qui s'expliquent par manque d'uniformité entre deux symboles, par exemple les lettres A et B dans l'alphabet latin.

Nous avons produit un tableau de synthèse des différentes DDD selon Floridi:

Les 3 niveaux d'application de la définition diaphorique de la donnée (DDD) d'après Luciano Floridi - Oxford 2005			
DDD 1	Diaphora de re	Données proto-épistémiques	Inférées par l'expérience
DDD 2	Diaphora de signo	Absence d'uniformité entre deux états physiques	Les variations d'un signal électrique au cours d'une conversation téléphonique
DDD 3	Diaphora de dicto	Absence d'uniformité entre deux symboles	A diffère de B dans l'alphabet latin

Les 3 niveaux d'application de la DDD selon L. Floridi (2005).

« Selon Floridi, les trois niveaux diaphoriques de données correspondent par hypothèse: chaque niveau rend possible le suivant²⁷. Mais comment passer d'un niveau à l'autre ? Cette question se pose dans une perspective bien précise: celle qui consiste à étudier comment l'information aide à construire l'expérience à partir d'un vécu indifférencié. La DDD1 (diaphora de re, c'est-à-dire différences nouménales) « rend possibles » des signaux de type DDD2 (diaphora de signo, c'est-à-dire différences perçues entre des états physiques à l'intérieur de la sémiose). Les signaux de DDD2 permettent alors de coder ces différences sous forme de symboles en DDD3 (diaphora de dicto, c'est-à-dire différences perçues entre des symboles: les DDD3 renvoient à un système explicite de symboles). A ce stade, la donnée devient une entité symbolique qui code une différence »²⁸.

Cette entité symbolique qui code une différence, est aussi « un point sur lequel on fonde un raisonnement, des suppositions, des constatations, des probabilités, qui étant indiscutables ou indiscutées, servent de base à une recherche, à un examen quelconque »²⁹. C'est ainsi que l'on peut parler de données scientifiques, de données techniques, de données physiques, de données médicales, de données génomiques...

²⁷« *Dedomena in (1) may be either identical with, or what makes possible signals in (2), and signals in (2) are what make possible the coding of symbols in (3)* ».

²⁸Sylvie Leleu-Merviel, Philippe Useille. Quelques révisions du concept d'information. Hermes. Problématiques émergentes dans les sciences de l'information, Lavoisier, pp.25-56, 2008. Traité des sciences et techniques de l'information, 978-2-7462-2110-9. <hal-00695777>

²⁹<https://fr.wiktionary.org/wiki/donn%C3%A9e#fr>

3]Les enjeux et conséquences des Mégadonnées sur la société contemporaine

Nous ne pourrions pas dans le cadre de cet article détailler l'ensemble des enjeux et des conséquences que pose le phénomène « Mégadonnées » sur notre société, mais nous allons nous concentrer sur quelques sujets qui nous paraissent être suffisamment conséquents et sans équivoques dans la littérature tel que le problème de la limite de l'anonymisation des données et son corollaire sur la « *data privacy* », les conséquences d'un « *réductionnisme datamétrique* »; le passage d'une vision « *post-orwellienne* » à une critique de la société « *post-foucauldienne* » ou encore « *post-deleuzienne* »: où les Data, le calcul, les algorithmes et les machines auto-apprenantes et augmentées, nous exposent à une certaine « *gouvernementalité algorithmique* »...

3.1 Les limites de l'anonymisation des données...

Les *Mégadonnées* mettent en exergue un problème nouveau, celui posé par le traitement des données massives et multi-structurées via des algorithmes à même de révéler des données parfois confidentielles et souvent sensibles sur les individus, qu'ils soient consommateurs, patients ou citoyens. Le défi mis en lumière depuis maintenant plus d'une décennie par de nombreux chercheurs en sciences des données et en droit notamment, nous montre combien l'enjeu de la préservation de la confidentialité des données sensibles et de la vie privée est prégnant, tant il est bousculé par les géants du web aujourd'hui.

Les techniques utilisées pour préserver la vie privée des individus font référence à ce qu'il convient d'appeler un processus d'anonymisation des données et peut être ainsi défini: « *l'anonymisation de données (a fortiori personnelles) consiste à modifier le contenu ou la structure de ces données afin de rendre très difficile ou impossible la « réidentification »³⁰ des personnes (physiques ou morales) ou des entités concernées (ce qui implique de bien définir ce que signifie dans ce cadre le concept d'identifiabilité). Les anglophones parlent aussi parfois de De-Identification(DE-ID)³¹ ».*

L'anonymisation et ses procédés semblent mis à mal depuis plusieurs années si l'on croit certains auteurs comme Pete Warden qui estime qu'il est quasiment impossible et impensable de réellement « anonymiser » certaines données et qu'il va falloir apprendre à travailler avec cette limite. En effet, d'éminents chercheurs ont montré qu'un individu pouvait être retrouvé dans une base de données médicales en connaissant son sexe, son code postal et sa date de naissance³², mais également dans une base de donnée

³⁰Hull SC, Wilfond BS (2008), *What does it mean to be identifiable*. Am J Bioeth. 2008 Oct ; 8(10) :W7-8.

³¹Neamatullah, I., Douglass, M. M., Lehman, L.-W. H., Reisner, A., Villarroel, M., Long, W. J., Szolovits, P., Moody, G. B., Mark, R. G., and Clifford, G. D. (2008). *Automated de-identification of free-text medical records*. BMC Medical Informatics and Decision Making, 8.

³²<http://arstechnica.com/tech-policy/2009/09/your-secrets-live-online-in-databases-of-ruin/>

téléphonique à partir de quatre points de géolocalisation³³ ou encore dans une base de données de carte bleues en connaissant quatre magasins et les jours où les personnes ont utilisés leur carte de paiement en question³⁴.

Selon Warden, prétendre qu'une donnée est totalement anonymisée peut en outre procurer un « *faux sentiment de sécurité* »³⁵. Il recommande de ne pas abandonner l'anonymisation (car elle freine et rend couteux le processus de reconstruction), mais il faut aussi diffuser plus intelligemment les données, en limitant si possible leur niveau de précision et en apprenant des experts (sociologues, médecins, économistes...) ou d'autres spécialistes qui travaillent sur ces questions depuis plusieurs années.

En comparant les données des uns aux autres, on peut en déduire votre propre comportement. « *Ce qui importe réellement est de savoir comment les gens sont reliés entre eux par les machines et comment, ensemble, ils créent un marché financier, un gouvernement, une société, et d'autres structures sociales* »³⁶

Notons que depuis 2015, dans une publication qui a fait beaucoup de bruit Yves Alexandre de Montjoye a démontré « *qu'à partir de l'étude des données de cartes bancaires, produites sur trois mois par 1,1 million de personnes, il suffit de quatre points «spatio-temporels* »³⁷(coordonnées géographiques, date et heure) pour retrouver l'identité de 90% des individus ». L'anonymisation des données, n'est donc plus possible pour protéger la vie privée et l'intimité des acteurs sur les médiums numériques. Les données sont autant de traces que nous laissons dans le cyberspace, sur les réseaux, sur les serveurs des opérateurs de paiement électronique.

Les traces numériques sont autant d'expressions ou de sentiments abandonnés consciemment ou inconsciemment par leurs auteurs, abandonnés au hasard et/ou de manière fortuite: on parle de « *shadow information* ». Ces « *informations fantômes* » sont formées par l'ensemble des données associées à sujet, à une personne, à un lieu, à un événement, et viennent ainsi alimenter le flux d'information, que chaque internaute laisse sur le web, les réseaux sociaux ou via leur mobile. Ces métadonnées, sont donc propices à l'analyse et l'extraction de sens sur le web et les réseaux sociaux, comme nous l'avons déjà évoqué.

Thierry Berthier, Mathématicien et chercheur à la Chaire Cyber-Défense de l'École Saint-Cyr a formé en 2013 le concept de « *projections algorithmiques* » pour définir « *la production de*

³³Yves-Alexandre de Montjoye, César A. Hidalgo, Michel Verleysen et Vincent D. Blondel, « Unique in the Crowd : The privacy bounds of human mobility », *Nature SRep*, no 3, 25 mars 2013.

³⁴Yves-Alexandre de Montjoye, « Unique in the shopping mall: On the reidentifiability of credit card metadata », *Science*, no 347, 30 janvier 2015.

³⁵Warden Pete "Why you can't really anonymize your data; It's time to accept and work within the limits of data anonymization". O'Reilly Media, Inc. 17 mai 2011.

³⁶ Hubert Guillaud, *D'autres outils et règles pour mieux contrôler les données*, billet de blog publié le 03/07/2013: <http://www.internetactu.net/2013/07/03/dautres-outils-et-regles-pour-mieux-controler-les-donnees/>

³⁷Yves-Alexandre de Montjoye, « Unique in the shopping mall: On the reidentifiability of credit card metadata », *Science*, no 347, 30 janvier 2015.

données et de métadonnées résultant des interactions d'un opérateur humain avec les systèmes qui l'entourent. Par son formalisme élémentaire, il permet de généraliser la notion floue de traces numériques produites volontairement ou non par un individu. Prenons l'exemple d'un message twitter rédigé par un utilisateur d'un réseau social. Son auteur croit n'émettre que 140 signes et être donc en mesure de les contrôler. Toutefois, plus de 4000 signes sont réellement envoyés. La différence tient à l'ensemble des métadonnées qui accompagne chaque message (compte de l'émetteur, adresse IP, date et lieu d'émission, etc.). Encore s'agit-il là d'une démarche volontaire de l'individu, d'une émission consciente»³⁸.

C'est aussi ce qui fait dire à l'un des plus grands spécialistes de la data science (reconnu mondialement), Sandy Alex Pentland (du MIT Lab), co-auteur de plusieurs papiers publiés dans *Science* et *Nature*, sur ces questions d'anonymisation des données pouvant être assez facilement reconstruites, donc retrouvées: « la puissance du Big Data, c'est de reposer sur les informations de comportement des gens plutôt que de reposer sur de l'information sur leurs convictions »³⁹.

Les Big Data ne reposent pas tant sur ce que vous publiez sur Facebook ou Twitter, ou à vos requêtes via Google, mais sur les données de localisation de votre téléphone ou celles de vos dépenses tracées grâce à votre carte de crédit. « Elles reposent surtout sur des miettes de données que vous laissez derrière vous quand vous vous déplacez dans le monde. Ces chapelures de données racontent l'histoire de votre vie. Elles disent ce que vous avez choisi de faire. (...) Les Big Data reposent surtout sur le comportement réel, et, en analysant ce type de données, les scientifiques peuvent dire une énorme quantité de choses sur vous. Ils peuvent dire si vous êtes le genre de personne qui va rembourser un prêt. Ils peuvent dire si vous êtes susceptible d'avoir un diabète. (...) Ils peuvent le dire parce que le genre de personnes que vous êtes est largement déterminé par le contexte social »⁴⁰.

Pentland indiquait déjà en 2013, dans une interview donnée à la revue *TheEdge* que nous étions bien face à un changement de paradigme qui nous impose de résoudre un nouveau dilemme entre les individus-citoyens, les acteurs traditionnels qui analysent et exploitent les données (banques, hôpitaux...) et les nouveaux acteurs, « pure players » qui agissent dans une « zone grise », celle de la dérégulation: « ceux qui ont les données les plus précieuses sont les banques, les compagnies de téléphone, les sociétés médicales, des industries hautement réglementées, qui ne peuvent pas vraiment tirer parti des données de la façon dont ils aimeraient à moins qu'ils obtiennent l'accès à la fois des consommateurs et des régulateurs. L'accord qu'ils vont devoir passer est de donner aux consommateurs le contrôle de leurs données en retour d'une possibilité leur permettant d'utiliser leurs données. Si vous traitez les données des personnes d'une manière responsable, les gens partageront volontairement leurs données. C'est la solution gagnant-gagnant au problème de la vie privée, et seules les entreprises qui ont grandi dans un environnement non réglementé ou les

³⁸Thierry Berthier, Projections algorithmiques et cyberspace. R2IE - revue internationale d'intelligence économique -Vol 5-2 2013 pp. 179-195.

³⁹Hubert Guillaud, *D'autres outils et règles pour mieux contrôler les données*, billet de blog publié le 03/07/2013: <http://www.internetactu.net/2013/07/03/dautres-outils-et-regles-pour-mieux-controler-les-donnees/>

⁴⁰ Hubert Guillaud, *D'autres outils et règles pour mieux contrôler les données*, billet de blog publié le 03/07/2013: <http://www.internetactu.net/2013/07/03/dautres-outils-et-regles-pour-mieux-controler-les-donnees/>

sociétés qui profitent de ce marché gris qui risque de se tarir y seront les plus fortement opposées »⁴¹.

C'est précisément le rôle du régulateur que de pallier à ce type de problème complexe et nouveau qui précède le droit et la réglementation des données en général et des données personnelles en particulier. Et c'est notamment ce que prévoit le nouveau Règlement Européen sur la protection des données, la « data privacy », la protection de la vie privée. Comme le rappelle la CNIL, Le nouveau règlement européen sur la protection des données personnelles est paru au journal officiel de l'Union européenne le 4 mai 2016 et entrera en application en 2018. L'adoption de ce texte doit permettre à l'Europe de s'adapter aux nouvelles réalités du numérique, à travers 3 axes principaux: le premier devant *renforcer les droits des personnes, notamment par la création d'un droit à la portabilité des données personnelles et de dispositions propres aux personnes mineures*, le second doit permettre de *responsabiliser les acteurs traitant des données (responsables de traitement et sous-traitants)* et le troisième devrait *crédibiliser la régulation grâce à une coopération renforcée entre les autorités de protection des données, qui pourront notamment adopter des décisions communes lorsque les traitements de données seront transnationaux et ainsi renforcer les sanctions*⁴².

3.2 De la « réduction datamétrique » des individus et de la personne ?

Xavier Guchet dans son excellent essai philosophique consacré à la médecine personnalisée⁴³ constate que la « *logique des sociétés de contrôle ne consiste plus à distinguer l'individu et la masse, à fixer l'individu au sein d'une masse par l'attribution d'un numéro de matricule, en développant des techniques politiques pour chacun des deux termes – respectivement les techniques de dressage des corps et celles de régulation des populations; la nouvelle logique est plutôt de fragmenter l'individu en « dividiels », c'est-à-dire ici en entités moléculaires susceptible vidées de toute référence à l'idée de totalité organique, connectées en réseaux complexes (les voies de signalisation moléculaire) , accessibles à un répertoire de stratégies d'actionnabilité susceptible d'être enrichi indéfiniment – les productions pharmaceutiques ainsi que les manipulations génétiques* », et démontre toute l'actualité de la pensée de Gilles Deleuze qui décrivait déjà le passage dans ce processus de transformation, *des sociétés disciplinaires en sociétés de contrôle.* « *On ne se trouve plus devant le couple masse-individu. Les individus sont devenus des « dividiels », et les masses, des échantillons, des données, des marchés ou des banques* »⁴⁴.

Notre capacité à décomposer en projections, traitements et analyses algorithmiques l'individu, toutes nos données, nous conduit à opérer, ce que nous appelons ici une

⁴¹ Propos du chercheur Sandy Pentland du MIT Lab publiés dans la revue TheEdge et traduits et cités par Hubert Guillaud sur son blogle 03/07/2013: <http://www.internetactu.net/2013/07/03/dautres-outils-et-regles-pour-mieux-controler-les-donnees/>

⁴² <https://www.cnil.fr/fr/reglement-europeen-protection-donnees>

⁴³ Guchet, Xavier, La médecine personnalisée, Un essai philosophique, collection « Médecine et Sciences Humaines », avril 2016, Les Belles Lettres.

⁴⁴ Gilles Deleuze, « Post-scriptum sur les sociétés de contrôle », in Pourparlers, 1972-1990, Paris, Minuit, 1990-2003, p. 240-247.

« *réduction datamétrique* » de la personne. La tentation d'une *réduction datamétrique* qu'elle soit biologique, physiologique, moléculaire, ne peut être que partielle et annihile l'individualité d'un individu, en le fragmentant et en le limitant à son *état individuel*. Cette réduction qui consiste à simplifier la complexité ontologique, psychologique, biologique de l'individualité, ne saurait tenir comme définition substantielle, ou essentielle de la personne humaine.

Le concept de personne en philosophie occidentale, depuis Saint-Augustin, en passant par Kant et jusqu'à la phénoménologie heideggerienne est étroitement lié sémantiquement aux concepts de liberté et d'autonomie de la volonté, tous deux irréductibles à un état numéraire, à des quantités numériques, à une « *réduction datamétrique* ». La liberté et l'autonomie d'une personne échappe aux enchaînements réglés de la nature et de ses mesures possibles. « *En d'autres termes, la personne est inconditionnée alors que la Nature est le domaine de ce qui est conditionné selon des lois universelles* »⁴⁵.

La cybernétique revendiquait déjà un statut et une définition scientifiques au concept de donnée et d'information; elle voulait ainsi réduire tout organisme vivant à *un système traitant de l'information*, traitant des données, transformées en information. La Cybernétique jetait les bases et les fondements d'une société du traitement de l'information naturel (homme) et artificiel (machine), où la personne humaine était définie et considérée comme un *système informationnel*. La cybernétique a enfanté les *sciences cognitives* (ou sciences de la cognition qui recoupent les neurosciences, l'intelligence artificielle, la psychologie, la philosophie...) et qui postulent que l'on peut réduire l'individu, une personne à ses divers traitements de l'information, fidèle à la métaphore (ou l'analogie) du cerveau et de l'ordinateur.

Cela signifie qu'une personne humaine est un système qui opère sur des représentations en fonction de leurs propriétés syntaxiques plutôt qu'en vertu de leur signification. Inspirés par les résultats d'Alan Turing, de nombreux chercheurs ont en outre admis que ces types de traitements pouvaient être réalisés par des machines artificielles complètement différentes du point de vue physique. La simulation et la modélisation informatiques pouvaient donc fournir de nouveaux moyens d'étudier le fonctionnement de l'esprit humain, c'est en cela que les *sciences cognitives* rejoignent ainsi le projet *cybernétique* dont elles sont issues: leur continuum épistémologique prend forme dans leur volonté d'intégrer dans un même cadre théorique l'étude des systèmes naturels et artificiels indissociablement.

A l'heure des mégadonnées et de l'accélération du développement de l'intelligence artificielle (machine learning, deep learning...) la tentative de *réduction datamétrique* de la personne n'a jamais été aussi proche et réaliste. Peut-être le signe de l'accomplissement prochain de l'œuvre cybernétique et cognitive...

⁴⁵ Guchet, Xavier, La médecine personnalisée, Un essai philosophique, collection « Médecine et Sciences Humaines », avril 2016, Les Belles Lettres.

3.3 Critique de la « société du calcul » et de la « gouvernance algorithmique ».

Alain Supiot dans son cours donné au cours au Collège de France sur la gouvernance par les nombres parfaitement pointé dès 2012, l'esprit de notre époque eu égard à notre fascination pour le calcul, la mesure, la quantification et comme le rappelle si justement Jacques Le Goff: « *Notre temps serait victime de « quantophrénie», de confiance immodérée et presque monomaniaque dans l'abstraction de chiffres et de nombres devenus le langage dominant de l'agir politique. C'est ce qu'Alain Supiot entend démontrer à travers une fresque qui nous fait remonter jusqu'à l'Antiquité, de souligner la permanence du rêve d'« harmonie par le calcul », d'un gouvernement « scientifique », d'une pure administration des choses, enfin émancipés des délices et poisons de la passion politique. Mais cette vaste recherche poursuit un autre but qui se révèle être son vrai mobile: faire ressortir les impasses et contradictions d'un mode de gestion qui, sous couvert d'impersonnalité réputée libératrice des individus, finit par ressusciter les « liens d'allégeance » dont on se pensait définitivement préservés* »⁴⁶.

Le Goff dans son ouvrage consacré à Supiot que « *tout a commencé dans le secteur privé par une extension constante de la calculabilité, des opérations de fabrication à l'ensemble de la gestion, y compris dans son volet social, avec prolifération des indicateurs sur des tableaux de bord de plus en plus sophistiqués. De proche en proche, le mouvement a gagné la société tout entière, sans épargner l'État: depuis l'application des calculs de probabilité à la vaccination, par exemple, ou le développement des assurances fondées sur le calcul actuariel...* »⁴⁷. Cette tendance s'est accélérée et généralisée à tous les domaines et secteurs de l'économie: prédictions et recommandations algorithmiques faites à des fins marketing et publicitaires, aide au diagnostic médical, pronostic médical, maintenance prédictive d'installations industrielle, prévisions de flux (marchandises, énergie...), prédictions du risque assurantiel (assurance comportementale, actuariat 2.0), prévision de décision de justice (aux USA)...

Nul n'échappe à l'omniprésence de l'analyse des données et à ce qu'il convient d'appeler désormais grâce aux travaux d'Antoinette Rouvroy, la « *gouvernementalité algorithmique* ».

« *La gouvernementalité algorithmique se caractérise notamment par le double mouvement suivant:*

a) *l'abandon de toute forme d'« échelle », d'« étalon », de hiérarchie, au profit d'une normativité immanente et évolutive en temps réel, dont émerge un « double statistique » du monde et qui semble faire table rase des anciennes hiérarchies dessinée par l'homme normal ou l'homme moyen ;*

b) *l'évitement de toute confrontation avec les individus dont les occasions de subjectivation se trouvent raréfiées.*

Ce double mouvement nous paraît le fruit de la focalisation de la statistique contemporaine sur les relations. Rouvroy s'intéresse à l'évaluation et la compréhension de la « gouvernementalité algorithmique » en construisant son analyse sur les fondements de la

⁴⁶Jacques Le Goff, « Alain Supiot. La gouvernance par les nombres. Cours au Collège de France (2012-2014) Fayard, 2015, Revue Projet 2015/6 (N° 349), p. 90-91.

⁴⁷ Ibidem.

théorie des processus d'individuation de la relation chez Simondon et sur l'émergence de formes de vie nouvelles sous la forme du dépassement du plan d'organisation par le plan d'immanence chez Deleuze-Guattari⁴⁸.

Rouvroy montre comment à partir de cette dialectique avec les principales philosophies contemporaines de la relation, la *gouvernementalité algorithmique* alimente et entretient un paradoxe fondamental, propre à la « *data-driven society* », qu'elle pointe parfaitement et qu'elle stigmatise: « *il apparaît alors qu'une pensée du devenir et des processus d'individuation par la relation réclame nécessairement du « disparate » - une hétérogénéité des ordres de grandeur, une multiplicité des régimes d'existence - que la gouvernementalité algorithmique ne cesse précisément d'étouffer en clôturant le réel (numérisé) sur lui-même. La gouvernementalité algorithmique tend plutôt à forclure de telles perspectives d'émancipation en reliant les processus d'individuation sur la monade subjective* »⁴⁹.

Conclusion

De ce maelstrom incroyable qui fut à l'origine de l'Internet puis du web: grâce à la Cybernétique, à la Contre-Culture, en passant par le Creative Ethos et la naissance d'une Creative Class, nous avons voulu montrer le rôle prépondérant et critique de la Data, ainsi que les conséquences de l'explosion des données numériques (Mégadonnées) et l'algorithmisation de la société qui pose question.

Le Dataclisme comme symptôme et la Data Science comme remède nous interroge sur les fondements et les risques d'un projet de société construite et gouvernée par les data et les algorithmes. De l'anonymisation des données, aux dangers d'une tentative de « réduction datamétrique » de la personne, jusqu'au contrôle d'une société computationnelle soumise aux règles et au fonctionnement d'une « gouvernance algorithmique », les dérives possibles sur les plans sociétal, économique et juridique semblent fondées et déjà engagées.

Pourtant des chercheurs comme Nicholas Diakopoulos et Sorelle Friedler⁵⁰ insistent sur le fait que tous les algorithmes et toutes les données sont conçus et créés au départ par des personnes bien humaines⁵¹. Même si les algorithmes génétiques par exemple peuvent évoluer eux-mêmes, tout comme les algorithmes d'auto-apprentissage automatique

⁴⁸ Antoinette Rouvroy, Thomas Berns, Gouvernementalité algorithmique et perspectives d'émancipation. *Le disparate comme condition d'individuation par la relation ?* Politique des algorithmes, Les métriques du web. Revue Réseaux, 2013/1 (n° 177).

⁴⁹ Ibidem.

⁵⁰ Nicholas Diakopoulos is an assistant professor at the University of Maryland, College Park. Sorelle Friedler is an assistant professor at Haverford College, and an affiliate at the Data & Society Research Institute.

⁵¹ Nicholas Diakopoulos & Sorelle Friedler, How to Hold Algorithms Accountable ? Algorithmic systems have a way of making mistakes or leading to undesired consequences. Here are five principles to help technologists deal with that. November 17th, 2016, MIT Technology Review: <https://www.technologyreview.com/s/602933/how-to-hold-algorithms-accountable/>

(machine learning) dont les modèles résultent d'une conception tout aussi humaine et dont les résultats sont façonnés par des décisions prises par l'homme, dans le choix des règles d'optimisation, dans la sélection des échantillons de données sélectionnées...

Ainsi, la responsabilisation implique l'obligation de rendre compte et de justifier la prise de décision algorithmique et d'atténuer les impacts sociaux négatifs ou les préjudices potentiels. Rendre les algorithmes responsables pourraient devenir possible et répond à cinq principes fondamentaux toujours selon Diakopoulos et Friedler: la responsabilité, l'explication, l'exactitude, la vérification et l'équité. Cette position doit nourrir les réflexions des data scientists actuels et futurs, notamment à propos du choix des algorithmes eux-mêmes et de leurs possibles effets et conséquences possibles sur le plan éthique.

Il est indispensable de promouvoir une société « *data responsable* » auprès des citoyens européens et donc limiter les *externalités négatives* des algorithmes si nous voulons éviter que ne s'installe en Europe une « *black box society* » (à l'américaine), comme la dénonce vigoureusement Franck Pasquale, professeur de droit à l'Université du Maryland et chercheur à Yale: « *les services de la boîte noire sont souvent beaux à voir, mais notre société de la boîte noire est devenue dangereusement instable, injuste et improductive. Ni les quants de New York ni les ingénieurs de Californie ne peuvent offrir une économie saine ou une société sûre. Ces tâches incombent à l'ensemble de la population, qui ne peut effectuer ce travail que si elle en comprend les enjeux* »⁵².

⁵²Pasquale, Franck, *Black Box Society*, Les algorithmes secrets qui contrôlent l'économie et l'information, septembre 2015, Editions FYP.

Bibliographie

Abbate Janet, *Inventing the Internet*, MIT Press, Cambridge 1999.

Berthier Thierry, Projections algorithmiques et cyberspace. R2IE - revue internationale d'intelligence économique - Vol 5-2 2013 pp. 179-195.

Bohn, R.E. (1994), « Measuring and Managing Technical Knowledge », Sloan Management Review, Fall 94 pp.61-73.

Dagnaud, Monique, Le Modèle Californien, Comment l'esprit collaboratif change le monde, Mai 2016, Odile Jacob.

Deleuze, Gilles, « Post-scriptum sur les sociétés de contrôle », in Pourparlers, 1972-1990, Paris, Minuit, 1990-2003.

De Montjoye Yves-Alexandre, « Unique in the shopping mall: On the reidentifiability of credit card metadata », *Science*, no 347, 30 janvier 2015.

De Montjoye Yves-Alexandre, « Unique in the shopping mall: On the reidentifiability of credit card metadata », *Science*, no 347, 30 janvier 2015.

De Montjoye Yves-Alexandre, Hidalgo César A., Verleysen Michel et Blondel Vincent D., « Unique in the Crowd: The privacy bounds of human mobility », *Nature SRep*, no 3, 25 mars 2013.

Delort Pierre, *Le Big data*, Kindle, 2016.

Diakopoulos Nicholas & Friedler Sorelle, How to Hold Algorithms Accountable ? Algorithmic systems have a way of making mistakes or leading to undesired consequences. Here are five principles to help technologists. November 17, 2016, MIT Technology Review.

Dictionnaire de l'information, Editions Armand Collin (3^{ème} édition), 2008.

Floridi, Luciano, « Semantic Conceptions of Information », *Stanford Encyclopaedia of Philosophy*, 2005. Article accessible en ligne à l'adresse suivante.

Goëta Samuel, « Fred Turner, Aux sources de l'utopie numérique. De la contre-culture à la cyberculture, Stewart Brand un homme d'influence », Questions de communication, 2013.

Guillaud Hubert, *D'autres outils et règles pour mieux contrôler les données*, billet de blog publié le 03/07/2013.

Hafner Katie et Matthew Lyon, *Where Wizards Stay Up Late*, 1996, N.Y., 1996.

Hauben Michael et Hauben Ronda, *Netizens: On the History and Impact of Usenet and the*

Internet, Los Alamitos, 1997.

Hauben Ronda, « À la recherche des pères fondateurs d'Internet. Pourquoi a-t-on besoin d'une histoire d'Internet? », *Multitudes* 2003/1 (no 11), p. 193-199.

Hull SC, Wilfond BS (2008), *What does it mean to be identifiable*. *Am J Bioeth.* 2008 Oct ; 8(10):W7-8.

Jezo-Vannier Steven, *Contre-culture(s) – Des Anonymous à Prométhée, Attitudes, Le mot et le reste*, 2013.

Le Goff, Jacques, « Alain Supiot. La gouvernance par les nombres. Cours au Collège de France (2012-2014) ? Fayard, 2015, *Revue Projet* 2015/6 (N° 349), p. 90-91.

Guchet, Xavier, *La médecine personnalisée, Un essai philosophique*, collection « Médecine et Sciences Humaines », avril 2016, Les Belles Lettres.

Leleu-Merviel Sylvie, Useille Philippe. Quelques révisions du concept d'information. *Hermes. Problématiques émergentes dans les sciences de l'information*, Lavoisier, pp.25-56, 2008, *Traité des sciences et techniques de l'information*.

Mayer-Schönberger Victor, Kenneth Cukier, *Big Data. A Revolution that Will Transform How*

We Live, Work, and Think, op. cit. [traduction de H. Dhifallah, Kindle, Empl. 133-137].

Mengal, Paul. « *Cybernétique, histoire d'un mot* ». *Res Publica* 18, 1998.

Minsky Marvin présente dans *Semantic Information Processing*, en 1969 la première cybernétique comme un tronc commun qui se serait divisé en trois branches: la « simulation cognitive » à la Allen Newell et Herbert Simon, l'« intelligence artificielle » et la « seconde cybernétique » ou théorie des systèmes auto-organiseurs.

Moulier Boutang Yann « *Le Capitalisme cognitif* » 2011 et « *L'économiste et l'abeille* » 2014.

Neamatullah, I., Douglass, M. M., Lehman, L.-W. H., Reisner, A., Villarroel, M., Long, W. J., Szolovits, P., Moody, G. B., Mark, R. G., and Clifford, G. D. *Automated de-identification of free-text medical records*. *BMC Medical Informatics and Decision Making*, 8, 2008.

Pasquale, Franck, *Black Box Society, Les algorithmes secrets qui contrôlent l'économie et l'information*, septembre 2015, Editions FYP.

Prax, J-Y., *Le manuel du knowledge management, une approche de deuxième génération*, Dunod, Paris, 2003.

Rheingold Howard, *Tools for Thought*, 1985.

Rosensweig Roy, « *Review Essay: Wizards, Bureaucrats, Warriors, and Hackers: Writing the*

History of the Internet », in *American Historical Review*, décembre 1998.

Rouvroy Antoinette, Berns Thomas, Gouvernamentalité algorithmique et perspectives d'émancipation. *Le disparate comme condition d'individuation par la relation ?* Politique des algorithmes, Les métriques du web. *Revue Réseaux*, 2013/1 (n° 177).

Salus Peter, *Casting the Net*, Reading, Mass, 1995.

Teboul Bruno et Berthier Thierry « *Valeur et Véracité de la donnée: Enjeux pour l'entreprise et défis pour le Data Scientist* » in « *La Donnée n'est pas donnée* », Actes du Colloque Ecole Militaire, publié chez Kawa en juillet.

Waldrop M. Mitchell, *The Dream Machine: J.C.R. Licklider, and the Revolution That Made Computing Personal* – août 2002.

Warden Pete “*Why you can't really anonymize your data; It's time to accept and work within the limits of data anonymization*”. O'Reilly Media, Inc. 17 mai 2011.

.....

