

Big data, Big Problems ?

Enjeux éthiques et sociaux du *big data*

Sami Coll

Sommaire

- A. Introduction
- B. Qu'est-ce que le *big data* ?
- C. Enjeux éthiques et sociaux du *big data*
 - I. Le *big data* n'est pas le savoir
 - II. Le *big data* n'est pas démocratique
 - III. Le *big data* produit de la discrimination
 - IV. La vie privée n'est pas la solution miracle
- D. Conclusion

A. Introduction

Le déploiement du *big data* est d'une brûlante actualité et mobilise de nombreux acteurs, publics et privés, comme des chercheurs désireux d'en définir les contours. Toutefois, malgré la multiplication des études à son sujet, les approches critiques qui s'inscrivent en obéissant à la rigueur scientifique des sciences sociales demeurent rares. Cet article propose de répondre à cette carence en proposant d'approcher le *big data* comme une forme de connaissance qui n'est en fait pas si inédite que cela, mais dont les finalités viseraient dorénavant en priorité la performance économique plutôt que des idéaux démocratiques comme la liberté et l'équité.

Sous le couvert de l'innovation numérique et de ses bienfaits pour la croissance et le bien-être de l'humanité, les pratiques de détournement d'usage des données, autrefois dissimulées,¹ sont désormais assumées et promues, au même titre que le croisement et la circulation des données entre les bases de données.² Aussi, le *big data*, notamment en facilitant techniquement la collecte de données sans finalité prédéfinie et le croisement des données, représente indéniablement un défi de taille pour la protection de la vie privée. En effet, face aux attentes des milieux économiques et politiques, la loi de protection des données, telle qu'elle

¹ David Lyon, *Surveillance Studies. An Overview*, 2007.

² Danah Boyd/Kate Crawford, *Six provocations à propos des big data*, in : Mounier (éd.), *Read/Write Book 2. Une introduction aux humanités numériques*, 2012, 197 ss.

existe en Suisse (LPD), en Europe (95/46/CE) et dans d'autres pays,³ est mise en difficulté. Aux Etats-Unis, le pays où sont situés les géants du numérique et par lequel des données personnelles transitent en masse, la régulation du trafic de données est très limitée.⁴ En outre, bien qu'absolument nécessaire, le débat sur la protection de la vie privée ne suffit pas. Paradoxalement, il tend même à dissimuler les enjeux éthiques et sociaux qu'il veut pourtant traiter. Par-là, les formes de surveillance exercées par le *big data* et le développement de la protection de la vie privée semblent à certains égards se renforcer mutuellement, en constituant un outil politique de gouvernance des individus.⁵

Après avoir donné une définition du *big data*, cet article s'attèle à décrypter ses enjeux éthiques et sociaux. Puis, il souligne l'impossibilité d'y répondre par la seule notion de vie privée et va jusqu'à suggérer une complicité involontaire entre la vie privée et la surveillance. Pour ce faire, il se base sur une littérature diversifiée et sur des travaux empiriques, dont une recherche sur les cartes de fidélité en Suisse⁶ que l'on peut considérer comme un système précurseur et annonciateur du *big data* et une recherche en cours sur le *big data*.

B. Qu'est-ce que le *big data* ?

Historiquement, le terme *big data* aura désigné une technologie qui permet un stockage massif et non-structuré d'informations dans des bases de données, ainsi que leur traitement exhaustif par des algorithmes informatiques.⁷ Techniquement, il marque en effet une rupture de paradigme par rapport aux bases de données relationnelles et structurées ayant fait leur apparition dans les années 1970.⁸ Son principe général est d'autoriser un stockage des données en masse sans plus avoir à se préoccuper de leur nature au moment de leur récolte. Les liens entre les données et le sens qui peut leur être donné se construit dans un deuxième temps à l'aide d'algorithmes de traitement des données. Selon ses promoteurs, l'intérêt principal du *big data* réside dans le fait qu'il peut faire des prédictions, même parfois en temps réel, dans une diversité de domaines : santé, biologie, sexualité, météorologie, achats, assurances, culture, gestion des réseaux électriques, etc.

En revanche, sur le plan de sa mise en œuvre concrète, le terme *big data* finit par désigner aujourd'hui plus généralement un ensemble de pratiques en voie de banalisation, indépendamment de leur mode de stockage (structuré ou non-structuré) : la récolte et le traitement d'informations à très grande échelle. Le *big data* en tant que technologie facilite la multiplication de nouvelles applications dont les finalités peuvent être réinventées, détournées, d'autant plus lorsque que les bases de données peuvent s'interconnecter et se croiser avec une facilité dé-

³ EPIC, Privacy & Human Rights. An International Survey of Privacy Laws and Developments, 2006.

⁴ Colin J. Bennett, The Privacy Advocates. Resisting the Spread of Surveillance, 2008.

⁵ Sami Coll, La vie privée comme outil de gouvernance. Surveiller et fidéliser le lien marchand, Les Cahiers du numérique 2014, 45 ss.

⁶ Sami Coll, Surveiller et récompenser. Les cartes de fidélité qui nous gouvernent, 2015.

⁷ Arvind Sathi, Big Data Analytics. Disruptive Technologies for Changing the Game, 2012.

⁸ Patrice Flichy/Sylvain Parasio, Présentation, Réseaux 2/2013, 9 ss.

concertante.⁹ Un tel potentiel d'innovation ne manque pas de susciter de grandes attentes de la part des milieux économiques qui perçoivent les données comme étant « le nouvel or noir ». ¹⁰ Par exemple, les bases de données des programmes de fidélité commencent à intéresser grandement les sociétés d'assurances, car elles permettraient, selon elles, d'établir un profil de risque personnalisé fiable. Ce scénario, qui semblait encore il y a peu inimaginable, fait maintenant partie des projets prospectifs de *Swiss Re*, la deuxième plus grande société de réassurance au monde.¹¹ Enfin, à l'heure actuelle, la technologie du *big data* est très majoritairement développée, exploitée et maîtrisée par le secteur privé, le plus souvent pour la mise en œuvre de « systèmes automatiques de recommandation »¹² dont le principe est de proposer de façon personnalisée des produits censés correspondre au profil de consommation de chaque client.¹³

C. Enjeux éthiques et sociaux du *big data*

I. Le *big data* n'est pas le savoir

Les algorithmes informatiques qui assurent le fonctionnement des systèmes de recommandation, qui représentent la mise en œuvre la plus répandue du *big data*, ont été disséqués par des chercheurs en sciences de l'information.¹⁴ Ils ont également été abordés théoriquement en tant que dispositif de contrôle, une « boîte noire », qui contrôle la circulation de l'argent et de l'information.¹⁵ Malgré que le *big data* soit monopolisé par le secteur privé, des scientifiques, des sciences naturelles comme des sciences sociales, y voient une potentielle révolution épistémologique du fait qu'il permettrait de faire de la recherche avec des données exhaustives, un fait sans précédent dans l'histoire des sciences.¹⁶ A ce propos, une littérature critique, bien qu'elle ne rejette pas radicalement cette opportunité, discute la validité épistémologique du mode de production de connaissance impliqué,¹⁷ notamment le fait qu'il serait basé essentiellement sur la recherche de corrélations, qu'il néglige la recherche de causalités, et qu'il vise prioritairement une

⁹ John Gilliom/Torin Monahan, *SuperVision. An Introduction to the Surveillance Society*, 2012.

¹⁰ Françoise Dupont, *BigData : entre régulation et architecture*. Introduction, *Statistique et Société* 4/2014, 9 ss.

¹¹ Mehdi Atmani, *Le Big Data révolutionne l'industrie de l'assurance*, *Le Temps* du 11.11.2014.

¹² Marc Ménard, *Systèmes de recommandation de biens culturels. Vers une production de conformité ?*, *Les Cahiers du numérique* 2014, 69 ss.

¹³ *Ibid.*

¹⁴ Erik Brynjolfsson/Yu Jeffrey Hu/Michael D. Smith, *Long Tails vs. Superstars: The Effect of Information Technology on Product Variety and Sales Concentration Patterns*, *Information Systems Research* 2010, 736 ss.

¹⁵ John Cheney-Lippold, *A New Algorithmic Identity. Soft Biopolitics and the Modulation of Control, Theory, Culture & Society* 6/2011, 164 ss ; Frank Pasquale, *The Black Box Society. The Secret Algorithms That Control Money and Information*, 2015.

¹⁶ Boyd/Crawford (note 2).

¹⁷ Jose van Dijck, *Datafication, dataism and dataveillance. Big Data between scientific paradigm and ideology*, *Surveillance & Society* 2014, 197 ss.

efficacité immédiate.¹⁸ Cette critique s'adresse tout particulièrement à deux croyances : une première envers la possibilité d'une « donnée brute » prétendument neutre, sortie de son contexte de production et censée représenter la quintessence de la vérité ;¹⁹ et une deuxième envers un principe selon lequel l'exhaustivité des données pallierait les biais liés à la récolte des données.²⁰

Ces croyances font en effet fi de toute une histoire épistémologique de la connaissance.²¹ Cette « nouvelle » façon de produire de la connaissance prenant corps avec les usages du *big data* ressemble en fait beaucoup à celles qui ont précédé l'âge préclassique. La connaissance y était guidée, en effet, par une utopie heuristique divinatoire où « tout serait immédiat et évident ». ²² Loin d'être révolutionnaire, le *big data* semble, sous cet angle, renouer avec les principes de prédiction et de recherche de similitudes entre les objets, les pensées et les représentations qui dominaient la connaissance avant les transformations traversées pendant l'âge classique, puis à la modernité, pour être porté par ce que nous appelons aujourd'hui la science, la médecine ou les sciences humaines. Enrichi par de nombreuses rénovations épistémologiques, le savoir moderne est censé notamment décrypter les relations de causalité, une tâche potentiellement négligée par le *big data* dès lors qu'elle n'est plus essentielle pour remplir des objectifs d'efficacité économique.²³

Faut-il dès lors voir dans le projet de connaissance numérique portée par le *big data* une perte d'autonomie du savoir par rapport aux objectifs préclassiques de prédiction ? Plus généralement, le *big data* n'est-il pas le support utopique d'une heuristique qui se met plus ou moins consciemment dans une position de recul épistémologique pour servir des objectifs immédiats, surtout pécuniaires ? Pour *Michel Foucault*, les modalités de production du savoir sont indissociables des enjeux de pouvoir qui traversent une société.²⁴ En suivant son raisonnement, *François Ewald* examine en prenant en considération les objectifs politiques et économiques du secteur privé et des Etats la façon dont le *big data* est développé dans le secteur des assurances.²⁵ Le monde scientifique, lui aussi, est chargé d'intérêts individuels et collectifs dont l'analyse force à relativiser l'objectivité des chiffres et des résultats.²⁶

II. Le *big data* n'est pas démocratique

Le *big data*, de par la grande opacité des secteurs qui développent et exploitent en premier lieu les algorithmes qui le composent,²⁷ se trouve au centre d'une réparti-

¹⁸ *Viktor Mayer-Schönberger/Kenneth Cukier*, *Big data. La révolution des données est en marche*, 2014.

¹⁹ *Daniel Rosenberg*, *Data before the Fact*, 2013.

²⁰ *van Dijck* (note 17).

²¹ *Michel Foucault*, *Les mots et les choses*, 1966.

²² *Idem*, 45.

²³ *Mayer-Schönberger/Cukier* (note 18).

²⁴ *Foucault* (note 21).

²⁵ *François Ewald*, *Assurance, prévention, prédiction... dans l'univers du Big Data*, 2012.

²⁶ *Theodore M Porter*, *Trust in Numbers. The Pursuit of Objectivity in Science and Public Life*, 1995.

²⁷ *Pasquale* (note 15).

tion des pouvoirs décisionnels entre la sphère gouvernementale et le secteur privé. Une forme de **pouvoir technocratique** entre en compétition avec le pouvoir politique démocratique et sa bureaucratie reposant sur la **légitimité rationnelle-légale**,²⁸ le premier n'étant pas assujéti au même devoir de transparence ni de responsabilité que le second. Celles et ceux qui constituent et bénéficient de ce pouvoir technocratique ne sont pas élus démocratiquement. Par exemple, lorsque *Eric Schmidt*, alors encore le président-directeur général de *Google*, énonce une ânerie d'une immense violence normative telle que « *If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place* », ²⁹ quels sont les moyens mis à disposition des usagers des services de *Google*, qui sont tout autant de citoyens, sinon pour le destituer, au moins pour former un contre-pouvoir ? D'ailleurs, en manifestent-ils seulement le souhait, face aux merveilles proposées et promises par le géant de l'Internet ? De la même manière, les moyens mis en œuvre par les sites de réseaux sociaux pour encourager la production de données entre les utilisateurs à des fins d'exploitation commerciale³⁰ est révélateur de l'emprise de ce pouvoir technocratique sur le marché économique global.

Le pouvoir et la légitimité de la recherche publique est également remise en question, en particulier celle des sciences sociales. Leur méthodologie procède généralement par échantillonnage, conformément à des règles épistémologiques passablement complexes. Face à une recherche privée qui a le privilège d'avoir accès à une masse de données récoltées par des entreprises privées, la recherche académique publique pourrait se trouver en difficulté face à l'argument et à la croyance que l'exhaustivité produirait par définition des résultats fiables. Par exemple, des chercheurs de l'Université privée américaine *Cornell* ont obtenu dans le cadre d'une collaboration avec *Facebook* l'accès à leurs données pour étudier la propagation des émotions négatives sur le réseau social,³¹ ou encore pour étudier à quel moment une personne se rétracte pour ne finalement ne pas valider une publication sur son mur.³² Des conclusions en sont volontiers tirées sur la nature psychologique et sociale de l'être humain, sans beaucoup se préoccuper des biais méthodologiques.

III. Le *big data* produit de la discrimination

L'accroissement du pouvoir technocratique que le *big data* autorise n'a pas les mêmes missions que le pouvoir démocratique basé sur sa légitimité rationnelle et légale.³³ Les algorithmes informatiques qui proposent des horizons de consommation conformes aux profils des usagers renforcent l'homogénéité sociale. Le rôle

²⁸ *Max Weber*, *Economie et société* I. Les catégories de la sociologie, 1971.

²⁹ Lors d'une interview à la CNBC du 9.12.2009, <<https://www.youtube.com/watch?v=A6e7wFDHzew>> (consulté 12.10.2015).

³⁰ *Olivier Glassey/Sami Coll*, Le « gisement de données » des réseaux sociaux. Le modèle d'affaire de Facebook, in : *Badillo/Roux* (éd.), *Le futur est-il « E-Media » ?*, 2014, 158 ss.

³¹ *Adam D. I. Kramer/Jamie E. Guillory/Jeffrey T. Hancock*, Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks, *PNAS* 2014, 8788 ss.

³² *Sauvik Das/Adam Kramer*, Self-Censorship on Facebook, *ICWSM* 2013.

³³ *Weber* (note 28).

d'Amazon, par exemple, n'est pas de remplir une mission culturelle en démocratisant l'accès à des biens culturels autrement réservés à des élites. Qui cherche le « 4-action pack de Chuck Norris » ne se voit pas proposer une compilation des œuvres de Claude Debussy. Les algorithmes commerciaux du *big data* ont une toute autre motivation politique. Ils sont développés par des acteurs convaincus qu'en proposant aux clients de manière personnalisée des produits qui correspondent à leurs goûts, voire même prédire ce qu'ils vont aimer à l'avenir, ils assureront un accroissement de leurs ventes et de leur profits.

Les problèmes posés par les pratiques de collecte et de traitement des données ont été abordés par les études sur la surveillance bien avant l'émergence du *big data*.³⁴ Ces travaux ont notamment mis en avant la nécessité d'étudier les processus d'exclusion et de discrimination que les pratiques de profilage et de tri automatique peuvent renforcer.³⁵ En s'inscrivant dans une personnalisation de masse, bien que les systèmes de recommandations permettent certes à un individu de faire des découvertes, ils tendent à l'entretenir dans un horizon culturel délimité.³⁶ On peut craindre alors un renforcement des inégalités sociales entre individus, dans la mesure où les pratiques culturelles sont une source importante de distinctions sociales, notamment quand une mobilité sociale ascendante est recherchée.³⁷ Les dynamiques de distinction individuelles et sociales qui ont été observées dans les usages de la mode,³⁸ dans les pratiques culturelles,³⁹ dans les espaces géographiques de vente⁴⁰ ou dans les relations marchandes personnalisées,⁴¹ sont renforcées par l'intermédiaire des systèmes de recommandation. Et elles se manifestent tant dans le cadre d'une relation entre une entreprise et un client qu'entre les utilisateurs d'un réseau social tel que Facebook.⁴² Sur ce dernier, des banques privées envisageant précisément d'observer la structure sociale en ligne de ses clients prospectifs pour décider de l'attribution ou du refus d'un crédit. Plus délicat encore, si les assurances en venaient à s'inspirer concrètement des projets de SwissRe en se mettant à développer des partenariats avec la grande distribution afin de pouvoir par exemple calculer le montant d'une prime dans le domaine de la santé, le principe de solidarité serait balayé pour donner place à des pratiques discriminatoires qui feront de l'Etat social une histoire ancienne.

³⁴ Gilliom/Monahan (note 9).

³⁵ Roger Burrows/Nicholas Gane, Geodemographics, Software and Class, *Sociology* 2006, 793 ss ; Cheney-Lippold (note 15) ; Oscar H. Gandy, *Coming to Terms with Chance. Engaging Rational Discrimination and Cumulative Disadvantage*, 2009.

³⁶ Eli Pariser, *The Filter Bubble. What the Internet Is Hiding from You*, 2011.

³⁷ Pierre Bourdieu, Les trois états du capital culturel, *Actes de la recherche en sciences sociales* 1979, 3 ss.

³⁸ Quentin Bell, *Mode et société: essai sur la sociologie du vêtement*, 1976 ; Frédéric Godart, *Sociologie de la mode*, 2010 ; Georg Simmel, *La tragédie de la culture et autres essais*, 1988.

³⁹ Pierre Bourdieu, *La distinction. Critique sociale du jugement*, 1979 ; Bernard Lahire, *La culture des individus. Dissonances culturelles et distinction de soi*, *Communication et langages* 141/2004, 97 ss.

⁴⁰ Sharon Zukin, *Point of Purchase. How Shopping Changed American Culture*, 2004.

⁴¹ Coll (note 6).

⁴² Sami Coll/Olivier Glassey/Claire Balleys, Building social networks ethics beyond « privacy »: a sociological perspective, *IRIE* 16/2011, 47 ss.

IV. La vie privée n'est pas la solution miracle

Par son projet d'aspiration généralisée de données sans objectif préétabli obligatoirement et d'interconnexion de bases de données, le *big data* se heurte frontalement à l'esprit des lois sur la protection des données, par exemple à l'art. 4 LPD qui exige qu'une personne doive être consentante et éclairée sur la finalité de la collecte d'informations personnelles. L'immense majorité des juristes spécialisés en droit de la protection des données reconnaissent, en effet, les limites du dispositif actuel de protection des données et de la vie privée qui n'arrive plus à faire face à la réalité d'un développement rapide des systèmes de production et d'analyse de données.

Malgré cela, la notion de vie privée continue d'occuper une place centrale dans la littérature, bien que quelques chercheurs considèrent, depuis une vingtaine d'années pour les plus précurseurs, qu'elle ne permet pas de saisir l'ampleur des enjeux de la surveillance.⁴³ D'aucuns lui reprochent notamment de rester prisonnière d'une conception excessivement individualiste,⁴⁴ problème auquel des auteurs ont cherché à répondre en l'enrichissant d'une dimension collective.⁴⁵ Une démarche que d'autres jugent vaine en arguant que la notion de vie privée n'est définitivement plus adaptée, ni pour protéger les citoyens, ni pour penser la surveillance.⁴⁶ Même ses plus ardents défenseurs conviennent en effet que le problème essentiel de la vie privée est d'être abordée le plus souvent comme ayant une forme universelle, dans sa perception et dans ses pratiques.

Notre recherche sur les cartes de fidélité⁴⁷ a démontré empiriquement l'incompatibilité entre la vision juridique de la vie privée (nous l'avons appelée la **vie privée normative et légale**), en particulier de la vie privée informationnelle telle qu'elle est développée dans les textes sur la protection des données, et la vision subjective des personnes interviewées (la **vie privée subjective**). Lorsqu'il a été demandé aux consommateurs interviewés de définir la vie privée en leurs propres termes, ils expriment une approche largement plus diversifiée en allant au-delà de l'enjeu de la perte ou de la reprise de contrôle sur ses données personnelles. Les individus définissent la vie privée principalement en termes relationnels, en faisant mention de leurs sentiments, de leur sexualité, de leurs libertés de choix, de leur corps, de leur spiritualité, ou encore en délimitant des espaces de temps. Si une très fine minorité d'entre eux relie plus clairement la notion de vie privée à l'information, ce n'est qu'encore plus rarement sous une forme numérique.

⁴³ Amitai Etzioni, *The limits of privacy*, 1999 ; Bénédicte Rey, *La vie privée à l'ère du numérique*, 2012.

⁴⁴ Felix Stalder, *Autonomy beyond privacy? A rejoinder to Bennett*, *Surveillance & Society* 2011, 508 ss.

⁴⁵ Priscilla M. Regan, *Response to Bennett: Also in defence of privacy*, *Surveillance & Society* 2011, 497 ss ; Paola Tubaro/Antonio A Casilli/Yasaman Sarabi, *Against the Hypothesis of the End of Privacy An Agent-Based Modelling Approach to Social Media*, 2014 ; Alan F. Westin, *Social and Political Dimensions of Privacy*, *Journal of social issues* 2003, 431 ss.

⁴⁶ Stalder (note 44) ; John Gilliom, *A response to Bennett's « In defence of privacy »*, *Surveillance & Society* 2011, 500 ss.

⁴⁷ Coll (note 6).

La complexité de la notion de vie privée s'est révélée encore plus franchement lors de l'observation des consommateurs dans leurs usages. Lorsqu'ils sont par exemple directement confrontés aux systèmes marchands de collecte et de conservation de leurs données personnelles, ils mettent en œuvre dans leur vie quotidienne une vie privée (nous l'avons appelée la **vie privée vécue**) ayant une forme qui ne ressemble ni à la vision de la **vie privée normative et légale**, mais plus étonnamment encore non plus à celle de la vision de la **vie privée subjective** provenant pourtant des consommateurs eux-mêmes.⁴⁸ Dans la pratique, l'idée de la collecte des données telle qu'elle est effectuée par les systèmes d'information reste une chose très abstraite pour la plupart des consommateurs. Les situations qui déclenchent des sentiments d'atteinte sont beaucoup plus concrètes, par exemple la gêne manifeste ressentie par une femme d'âge moyen quand un jeune homme séduisant lui demande sa date de naissance pour établir une carte de fidélité.

Paradoxalement, ce sont même les mesures destinées à protéger la vie privée des clients qui produisent parfois des sentiments d'atteinte. Par exemple, dans les centres d'appels que nous avons observés, pour protéger leurs données contre des accès frauduleux, les opérateurs sont tenus d'identifier les appelants en leur demandant leur numéro de carte, leur domicile, leur téléphone, ainsi que de répondre à deux questions ; combien de points ils pensent avoir sur leur compte, le dernier magasin où ils sont allés, ou la dernière récompense qu'ils ont commandé. C'est précisément ces questions, bien qu'honnêtement destinées à protéger leurs données, qui produisent un sentiment de malaise. En un sens, les entreprises observées semblent parfois vouloir **protéger la vie privée des clients contre leur gré**.⁴⁹ Ces derniers sont, dans les faits, prêts à dévoiler leurs informations personnelles plutôt volontiers,⁵⁰ même s'ils produisent un discours, que les études rassurantes ne manquent pas de capturer, où ils disent se préoccuper de leur vie privée.

Eu égard à ces résultats, le principe d'autodétermination informationnelle qui guide les lois de protection des données,⁵¹ à savoir que chaque personne est censée être un acteur responsable de sa propre vie privée (notamment en utilisant le droit d'accès aux données, leur modification ou leur suppression), montre de sérieuses limites. En outre, il devient de plus en plus difficile d'exercer un prétendu droit d'accès aux données, car le principe de transparence sur lequel il repose n'est aujourd'hui qu'une fiction. Au-delà de la réticence des entreprises privées à les livrer, en particulier les agences de renseignement économique, les données qui sont par exemple récoltées par les systèmes de fidélité n'en restent pas à leur apparent statut de banalité. Elles sont traitées par des algorithmes de « forage des données » (*data mining*) qui permettent d'établir des profils psychologiques ou

⁴⁸ Coll (note 5).

⁴⁹ Jan Holvast, *History of Privacy*, in : de Leeuw/Bergstra (éd.), *The History of Information Security. A Comprehensive Handbook*, 2007, 737 ss.

⁵⁰ Caroline Lancelot Miltgen, *Vie privée et marketing. Étude de la décision de fournir des données personnelles dans un cadre commercial*, *Réseaux* 3/2011, 131 ss ; Alain Rallet/Fabrice Rochelandet, *La régulation des données personnelles face au web relationnel : une voie sans issue ?*, *Réseaux* 3/2011, 17 ss.

⁵¹ Bennett (note 4).

sociologiques.⁵² Bien que les données brutes apparaissent parfaitement anodines pour celles et ceux qui les fournissent, il est possible d'en déduire des données sensibles telles que l'orientation sexuelle, les opinions politiques, l'état de santé ou l'appartenance religieuse.⁵³ Il suffit par exemple de constater qu'un client achète souvent des épices pour la viande, sans jamais acheter de viande ni d'alcool, pour émettre l'hypothèse qu'il est musulman, dans le but de lui adresser une publicité pour une future gamme de produits *hallal*. Si la majorité des consommateurs que nous avons interviewés n'ignorent pas que leurs données sont conservées par les entreprises (c'est sans aucun doute un progrès), ils ignorent en revanche le fait qu'elles puissent se transformer en données d'une autre nature (un prochain progrès à espérer).

Malgré toutes ces limites, dans les projets technologiques, dont le *big data*, la vie privée continue à être utilisée systématiquement comme caution éthique.⁵⁴ Le principe d'autodétermination informationnelle qui l'accompagne étroitement, qui donne le droit aux individus mais aussi *de facto* le **devoir** de veiller à leur vie privée, est solidement ancré dans une vision très individualiste de la société. Plus encore, la vie privée est devenue la condition du développement de l'économie numérique. Contrairement à la vision dominante, les individus ne sont pas en négociation constante avec différentes plateformes et supports numériques pour échanger leur vie privée contre des services. Le rapport de pouvoir quant à la définition de la vie privée – où elle commence et où elle s'arrête – se joue entre les sphères politiques, économiques et juridiques. Pour l'instant, comme résultat, les individus n'ont guère d'autre choix que de suivre la danse et d'adopter une « pratique de la vie privée » qui soit compatible avec les objectifs du *big data*.

D. Conclusion

Le *big data* est abordé de nombreuses manières, mais encore très peu depuis la perspective critique et sociologique développée dans cette contribution. Elle amène un éclairage sur les modalités de production d'une nouvelle forme de connaissance et sur son utilisation pour remplir des objectifs économiques et de performance, aux dépens de valeurs démocratiques comme la liberté et l'équité. Dans ce cadre, elle met en doute la compétence de la notion de vie privée à limiter la faim pantagruélique et croissante en données personnelles, et la conception d'un individu qui serait capable de s'en protéger par lui-même.

Le *big data* est en contradiction manifeste avec les principes des lois de protection des données. Mais les gouvernements ne peuvent, ni ne veulent (ni devraient, peut-être) entraver son développement. Comment envisager, dans un tel contexte, le maintien du rôle de l'Etat dans le développement de l'économie nu-

⁵² Jason Millar, Core Privacy. A problem for Predictive Data Mining, in : Kerr/Lucock/Steeves (éd.), Lessons from the Identity Trail. Anonymity, Privacy and Identity in a Networked Society, 2009, 103 ss.

⁵³ Sami Coll, Consommation sous surveillance : Le biopouvoir des programmes de fidélisation, Revue suisse de sociologie 2014, 475 ss.

⁵⁴ Randy K. Lippert/Kevin Walby, Governing Through Privacy : Authoritarian Liberalism, Law, and Privacy Knowledge, Law, Culture and the Humanities online 2013.

mérique, et dans la préservation de l'équité et de la vie privée des citoyens ? L'article a amené plusieurs éléments à considérer.

Premièrement, la démarche proposée par le *big data* n'est ni nouvelle ni révolutionnaire. Le prétendre, c'est aussi faire du *marketing* ... pour vendre les technologies du *big data*. L'idée que l'on pourrait fabriquer de la découverte « pure » et « objective » à partir de données « brutes » sans s'inscrire dans un cadre théorique connu et validé, sans formuler des hypothèses ou sans assurer une supervision spécifique, est une démarche épistémologiquement naïve. La théorie de la connaissance et l'expérience de la recherche nous apprend que la récolte et l'interprétation de données reposent inévitablement sur des préconceptions théoriques, sur une vision du monde, et sur des objectifs politiques, même inconsciemment.

Deuxièmement, les données ne sont pas neutres, mais contextuelles. Elles sont produites pour répondre à des objectifs emprunts de contingence politique, économique et historique. Par exemple, *Google* récolte l'historique des requêtes de ses utilisateurs pour leur proposer de la publicité ciblée et de meilleurs résultats de recherche. Cela se produit dans un contexte où une entreprise privée domine l'Internet, fait rendu possible dans une société mondialisée et libérale. Un grand magasin qui a réussi à faire porter une carte de fidélité à la majorité de ses clients ne devrait pas attribuer un risque plus élevé pour une maladie cardiovasculaire à une personne qui n'achète que des frites et du saucisson. Au-delà de la remise en question du principe de solidarité qui serait lourdement remis en question, peut-être que la personne en question va aussi s'empiffrer de légumes achetés au marché du coin parce qu'ils ont encore un goût. Certes, recycler des données à d'autres fins, par exemple scientifiques, est envisageable et intéressant, mais cela implique un biais qu'il faut identifier et contrôler. Sinon, on peut en inférer n'importe quoi. Un peu comme le fit *Joe Pyne*, présentateur américain de télévision et vétéran du Vietnam, en adressant à *Frank Zappa* : « Vous avez des cheveux long, alors vous êtes une fille ? ». Lequel lui répondra : « Vous avez une jambe en bois. Alors vous êtes une table ? ».

« Savoir, c'est pouvoir » suggérait *Francis Bacon* en s'enthousiasmant de l'emprise du scientifique sur la nature. Un savoir qui, dans la société de l'information, donne au détenteur de données personnelles une emprise sur une personne ou l'ensemble de la population. Un pouvoir d'autant plus fort si ces données, et les algorithmes qui les triturent, sont dans des « boîtes noires » protégées par le secret industriel. Un secret largement renforcé par l'ignorance de l'existence d'un tel pouvoir par la plupart des citoyens et des consommateurs. « Pouvoir, c'est savoir », comme l'aura également argué *Michel Foucault* dans l'ensemble de son œuvre. Le philosophe analyse ce qui relie les formes de production de la connaissance aux structures de pouvoir qui régulent une société. Quelles institutions ont le pouvoir de définir ce qu'est le savoir légitime ? Qui fait le savoir ? Le savoir brut, la vérité, la connaissance désintéressée, appelons-la comme on veut, n'existe pas. La connaissance est le résultat de procédures, de méthodes, de croyances et d'intérêts à défendre. Elle est le reflet d'une époque et répond à des impératifs politiques, économiques et historiques. Aussi, le *big data* est une forme de « vérité », sûrement utile et efficace, mais qui sert les valeurs de notre capitalisme : individualisme, performance, efficacité et profit. Un savoir qui ne mérite peut-être pas l'enthousiasme utopique qu'on lui porte aujourd'hui.

Enfin, la protection de la vie privée est la moindre des choses que l'on puisse chercher à garantir aux citoyens et aux consommateurs, mais elle ne suffit pas. Peut-on vraiment attendre et exiger des individus qu'ils apprennent à « gérer » leur propre vie privée, comme les politiques de protection des données le suggèrent ? On a pu voir dans cette contribution la complexité de la notion de vie privée. On peut regretter sa résistance à la compréhension, mais les divergences sur la façon de l'aborder ne sont-elles pas aussi les garantes d'une liberté, celle de pouvoir la vivre comme chacun l'entend ? En effet, si les individus avaient une perception de la vie privée au diapason de celle qui est prescrite par les autorités, serait-elle autre chose que la cible d'un pouvoir, politique ou économique, visant à forger les individus pour devenir les sujets d'une surveillance généralisée ?

Mais surtout, la discussion sur l'éthique et les conséquences sociales du *big data* doit dépasser les questions techniques sur la vie privée. La constitution d'un savoir biaisé et basé sur des données personnelles, utilisé par des acteurs qui n'ont pas été forcément choisis ou démocratiquement élus, pose un grand nombre de problèmes. Comment peut se défendre un groupe social face à un algorithme désincarné et non identifiable qui le discrimine automatiquement ? Qui gouverne le *big data* et ses algorithmes ? Peut-on élire ou destituer un algorithme ? Difficile d'apporter une réponse définitive. Mais une véritable transparence des données, pour commencer, qui assurerait un accès non seulement aux données brutes mais aux algorithmes de traitement des données, représenterait déjà un progrès considérable. Car, de par leur capacité de fabriquer des profils psychologiques et sociologiques depuis des données personnelles à l'apparence anodines, ces algorithmes donnent un pouvoir considérable aux institutions qui les contrôlent, dans la plus grande ignorance des citoyens. En réponse, en se concentrant surtout sur la vie privée et en cherchant à en donner une définition aussi opérationnelle que possible, les gouvernements, déchirés entre l'enthousiasme de l'avenir économique radieux promis par les entrepreneurs du *big data*, d'un côté, et la protection des individus, de l'équité et des libertés, de l'autre, pourraient même involontairement en devenir les complices.

Au-delà de ces critiques qui ont pour but premier de faire avancer le débat sans désigner de coupables, il est clair que le fondement des bases d'une « éthique digitale » ne peut se faire sans la collaboration des entrepreneurs du *big data*. Ces derniers réclament continuellement des éclairages sur l'implication sociale de leurs différents projets plus ou moins innovants. Loin de vouloir mettre en place un *big brother* numérique, ils se sentent concernés par les implications sociales de leurs inventions et réclament des éclairages sociologiques pertinents, qui sont hélas insuffisants. « *Don't be evil* », affirme le slogan de *Google*...