VITALIS (André), « Big Data et autodétermination informationnelle des individus », *Études digitales*, n° 2, 2016 – 2, *Le gouvernement des données*, p. 41-49

DOI : 10.15122/isbn.978-2-406-07064-1.p.0041

VITALIS (André), « Big Data et autodétermination informationnelle des individus »

RÉSUMÉ – Les techniques numériques facilitent aujourd'hui l'expropriation par de grandes organisations privées et publiques, des données des individus. Ce phénomène participe à la violation des principes d'une autodétermination informationnelle reconnue par la loi. Il peut porter atteinte au fonctionnement démocratique. Seuls des projets de Big Data reposant sur le consentement pourraient satisfaire à une exigence démocratique minimale.

ABSTRACT – Digital technology now facilitates the expropriation of individuals' data by large private and public organizations. This phenomenon contributes to the violation of the principles of informational self-determination recognized by the law. It may undermine the functioning of democracy. Only consent-based Big Data projects could satisfy a minimum democratic requirement.

# BIG DATA ET AUTODÉTERMINATION INFORMATIONNELLE DES INDIVIDUS

Le phénomène du Big Data représente une nouvelle puissance de stockage et de traitement des informations et une nouvelle étape de l'informatisation de la société. L'accroissement continu de la quantité d'informations prélevées dans de nombreux domaines scientifiques, économiques ou sociétaux constitue le critère le plus évident pour caractériser cette étape. Aujourd'hui le volume de données double tous les 18 mois avec des milliards de données personnelles collectées chaque jour. En 2015, on dénombrait 145 milliards de mails, 4,5 milliards de recherches sur Google, 400 millions de tweets. Toutes les heures, 10 millions de nouvelles photos étaient téléchargées sur Facebook.

Ces données, soustraites à la maîtrise des individus, prennent une plus grande valeur avec le Big Data. Grâce aux réutilisations qu'il opère et qu'il peut toujours multiplier, il les convertit en ressources essentielles de l'économie numérique. Ce faisant, il ne respecte pas leur statut particulier et restreint la place et le rôle du décideur humain en confiant à des algorithmes, dans les situations les plus diverses, la recherche de la solution la plus efficace possible. À côté de l'homme augmenté des technologies numériques on n'aurait garde d'oublier l'homme amoindri par la dépossession de ses données et l'optimisation algorithmique qui en est faite.

## LE PERFECTIONNEMENT DU PROFILAGE DES POPULATIONS

### UNE SURVEILLANCE FONDÉE SUR LE RECUEIL ET L'EXPLOITATION DES INFORMATIONS SUR LES INDIVIDUS

Cette nouvelle modalité de surveillance indirecte est apparue avec l'installation au milieu du 19e siècle, des premières démocraties représentatives en Europe et aux États Unis. Une surveillance inconnue jusqu'alors que nous désignons avec Armand Mattelart sous le terme de « profilage des populations[1] », complète les disciplines théorisées par Michel Foucault[2], avant de les remplacer. À la différence de ces dernières basées sur l'intériorisation des normes et l'autocontrôle, le profilage laisse apparemment libres des individus surveillés en permanence. Il n'a pas cessé d'évoluer et de se moderniser au gré du progrès des techniques et des crises sociales et économiques. Son histoire peut être rapidement retracée à travers l'évocation des séquences les plus significatives.

Au début, seules les populations marginales sont concernées. Une attention particulière est portée aux prisonniers avec l'invention en 1833 par la préfecture de Paris, de la première fiche mobile qui permet de reprendre l'information contenue dans les registres puis de l'enrichir par la suite en la soumettant à diverses manipulations : classement, recherche ponctuelle à partir d'un nom, recoupements, etc. Les populations itinérantes font l'objet de mesures discriminatoires comme si l'absence de domicile fixe en faisait des classes dangereuses par nature. La liberté d'aller et venir, liberté démocratique pourtant essentielle, n'est pas la même pour tous. Après un livret ouvrier institué en 1781 et repris par Napoléon en 1807, visant à attacher l'ouvrier à son patron, une loi de 1912, oblige les populations nomades à posséder un carnet anthropométrique qui lors des déplacements doit être visé par les autorités. Quelques années plus tard, une carte d'identité sera exigée des étrangers.

Le développement du capitalisme entraîne dans l'entre-deux-guerres, une surveillance plus stricte du monde du travail. C'est à cette époque

1 Armand Mattelart et André Vitalis, *Le profilage des populations. Du livret ouvrier au cybercontrôle*, Paris, La Découverte, 2014.

2 Michel Foucault, *Surveiller et punir. Naissance de la prison*, Paris, Gallimard, 1975.

qu'apparaissent dans les ateliers, des appareils de pointage des allées et venues des travailleurs et le chronométrage de leurs gestes. Le taylorisme enferme le travail dans des normes de rendement et les spécialistes du marketing expérimentent les premières techniques de fichage du consommateur.

Après la seconde guerre mondiale, les États-Unis édifient un complexe militaro-industriel au sein duquel seront conçus les grands systèmes téléinformatiques qui vont servir de matrice à l'ensemble des futurs dispositifs de surveillance de masse. C'est également à cette époque que les États-providence européens constituent d'énormes fichiers en immatriculant tous les bénéficiaires de leurs prestations sociales.

Pendant la crise des années 70, le recours à l'informatique va apporter une puissance nouvelle en étendant le contrôle à toute la population. Les capacités de stockage et de traitement automatique des informations vont permettre la création de fichiers centralisés, la réalisation d'interconnexions et l'établissement de profils statistiques de mauvais payeurs ou d'enfants à risques. C'est le moment où on prend conscience des dangers que font courir aux libertés, des États en mesure de connaître grâce à cette technologie, les moindres faits et gestes de leurs populations. Des lois vont intervenir pour encadrer cette puissance informatique en posant des limites à l'indiscrétion et en donnant de nouveaux droits aux personnes fichées.

Les années 90 inaugurent une nouvelle séquence avec le développement des supports numériques et de l'internet grand public. La numérisation apporte une innovation technique de très grande portée : l'automatisation de la collecte des données. Tout support numérique laisse en effet des traces des opérations réalisées, des informations de retour qui vont être accaparées par les opérateurs souvent à l'insu des utilisateurs ignorants des utilisations ultérieures qui en seront faites. Cette automatisation rend invisible une opération souvent délicate où peut s'exprimer le consentement ou l'opposition des individus concernés. Ce sont les entreprises qui vont bénéficier de cette innovation et surtout les entreprises américaines qui vont pouvoir ainsi constituer une cartographie mondiale des entités individuelles. Cette automatisation de la collecte qui à côté des intérêts privés va servir après le 11 septembre des intérêts sécuritaires, complète la mise en données de l'individu et est le préalable essentiel du phénomène du Big Data.

### L'ACCROISSEMENT DES CAPACITÉS DE COLLECTE ET DE TRAITEMENT

Le rassemblement de masses énormes d'informations et la définition d'algorithmes pour les traiter, perfectionnent le profilage des populations. Désormais, on est en mesure d'établir des profils personnalisés très détaillés, en traitant en temps réel, les informations des internautes pour des fins publicitaires et pour influencer leurs achats comme le fait par exemple Amazon dans les recommandations à ses clients. Dans la lutte contre le terrorisme, on observe une volonté de multiplier les sources d'information pour observer et suivre pas à pas, un suspect ou intercepter toutes les communications de son environnement immédiat comme peut le faire par exemple l'« Imsi-catcher », un aspirateur de données mobile qui tient dans une valise.

On établit également des profils statistiques prédictifs sur la base de multiples corrélations. Par exemple, sur la base de plusieurs millions de modèles mathématiques utilisant des milliards de données, Google a pu suivre et prédire les épidémies de grippe dans le monde mieux que ne pouvaient le faire les statistiques gouvernementales et les organismes officiels de santé.

Les perfectionnements apportés par le Big Data dans l'élaboration des profils, permettent d'anticiper et de prévoir les comportements pour mieux les influencer ou les contraindre. Dans un tel contexte, le libre choix est menacé, les comportements étant de plus en plus conditionnés.

## UNE PROTECTION DES DONNÉES PERSONNELLES MISE À MAL

### UN STATUT PARTICULIER NON RESPECTÉ

Les lois votées dans plusieurs démocraties autour des années 70 pour se protéger des dangers de l'informatique, ont attribué un statut particulier aux données personnelles. La collecte et l'utilisation de ces données doivent obéir à un certain nombre de principes. On ne peut effectuer

un traitement que par rapport à une finalité et les données stockées ne doivent pas être excessives par rapport à cette finalité. Leur durée de conservation est limitée. Elles doivent être sécurisées. Certaines données sensibles ne doivent pas être collectées. Leur transfert vers un État étranger n'est autorisé que si cet État offre un niveau de protection équivalent. De nouveaux droits sont accordés aux personnes fichées : un droit à une information préalable au fichage, un droit d'accès à leurs données ainsi qu'un droit d'opposition. Par ailleurs, une autorité indépendante est créée pour veiller à l'application de ces règles. La Charte des droits fondamentaux de l'Union européenne après avoir reconnu un droit au respect de la vie privée, consacre un article particulier à la protection des données à caractères personnel, en insistant particulièrement sur la nécessité pour tout traitement, hormis les exceptions prévues par la loi, du consentement préalable de la personne concernée.

La numérisation des supports et surtout internet ont permis de contourner ces règles. Ainsi, devant des automatismes, on voit mal comment peut s'exprimer le consentement : l'utilisation de la technologie signifie l'acceptation implicite de sa logique. Surtout les grandes entreprises américaines qui dominent le réseau n'ont pas respecté ces règles dans la mesure où elles ne sont soumises dans leur pays qu'à une autoréglementation. Autrement dit, elles appliquent les règles qu'elles se fixent elles-mêmes et qui ne sont pas de nature à contrarier leurs intérêts commerciaux. Il est vrai qu'il existe deux interprétations différentes du droit à la vie privée aux États-Unis et en Europe. Alors que cette dernière en fait un droit de l'homme touchant à sa dignité, les États-Unis le considèrent comme un droit touchant à la liberté et se montrent plus sensibles aux accommodements favorables aux intérêts économiques.

Si on en croit le dicton, « on n'attrape pas les mouches avec du vinaigre ». Ces grandes entreprises ont attiré un nombre toujours croissant d'utilisateurs grâce à la qualité de leurs services et surtout à leur gratuité. Ces utilisateurs oublient que le prix à payer est l'expropriation de leurs données. L'intérêt immédiat des services offerts fait oublier les inconvénients et les dangers que peut comporter ce piratage des données. Il est vrai également que sur les réseaux sociaux, les individus aident ce piratage en livrant eux-mêmes des informations considérées jusqu'alors comme confidentielles, en semblant attacher peu d'importance à la préservation de leurs secrets.

Avec le Big Data, le statut particulier des données personnelles est frontalement remis en cause. Le principe de finalité est totalement bafoué par des réutilisations non prévues au départ. Il en est de même pour le principe de proportionnalité puisque l'on rassemble le plus de données possible pour les croiser entre elles. Même chose pour la limitation de la durée de conservation des données dans la mesure où leur conservation indéfinie peut toujours s'avérer profitable.

## UN VÉRITABLE FÉODALISME NUMÉRIQUE

Vue d'Europe, la situation actuelle est préoccupante. Les entreprises américaines dominent internet et le Big Data. Ces entreprises ont constitué d'énormes réservoirs de données dans des « data center » qui quadrillent le monde. Elles établissent une cartographie mondiale des identités en ayant violé délibérément le droit à la vie privée des internautes. Elles constituent bien souvent des monopoles de savoir et de pouvoir pour reprendre une expression d'Harold Innis[3]. Google connaît nos habitudes de navigation, Amazon nos préférences en matière d'achat, Facebook nos relations sociales, Twitter sait à quoi nous pensons, Apple suit nos déplacements, etc. Ces monopoles occupent une place privilégiée pour mettre en œuvre le Big Data avec divers partenaires qui sont obligés de faire appel à leurs services.

La puissance des GAFA (Google, Apple, Facebook, Amazon) confortée chaque jour par une économie de l'information qui les avantage, est menaçante pour les États qui sont dans l'incapacité de faire respecter des règles qu'ils ont eux-mêmes édictées. Comme l'ont montré les révélations d'Edward Snowden[4] l'État de surveillance global qu'est devenu l'État américain, sait cependant tirer profit de l'accumulation par des acteurs privés de masses d'informations personnelles dans lesquelles il pourra toujours puiser.

3 Harold Innis, *Empire and communications*, University of Toronto Press, 1950 ; *The bias of communication*, Toronto, University of Toronto Press, 1951.

4 Antoine Lefébure, *L'affaire Snowden. Comment les États-Unis espionnent le monde*, Paris, La Découverte, 2014.

## UNE AUTODÉTERMINATION INFORMATIONNELLE À RECONQUÉRIR

### LE CONSENTEMENT PLUS QUE LA PROTECTION

Le Big Data apporte d'incontestables avantages en termes de bénéfices économiques, de lutte contre la délinquance ou de progrès dans le domaine de la santé. Un ancien président du Conseil national du numérique a pu reprocher à la Commission nationale informatique et libertés, dans sa tâche de protection des données personnelles, d'être un frein à l'innovation et d'empêcher la France de prendre la route du futur pour être dans le peloton de tête des pays les plus prospères. Pour concilier le respect des libertés individuelles avec le Big Data, il n'y a pas de recette magique. Dans un État démocratique cette conciliation passe par une refondation basée sur une autodétermination informationnelle des individus.

L'autodétermination informationnelle peut se définir comme la capacité de l'individu de décider de la communication et de l'utilisation de ses données. Mise en avant et validée dans un récent rapport du Conseil d'État[5], cette notion est aujourd'hui fortement compromise. Les individus ont perdu la maîtrise de leurs informations et ignorent le plus souvent, les utilisations qui en sont faites. Le plus urgent est de reconquérir cette maîtrise. À la suite des révélations d'Edward Snowden, la prise de conscience de l'ampleur du fichage réalisé, devrait marquer le début de cette reconquête. L'Union européenne a adopté un règlement qui obligera en mai 2018 les grandes entreprises américaines à respecter le statut particulier qui s'attache en Europe aux données personnelles. La Cour de justice de l'Union a reconnu en 2014 un droit de déréférencement sur un moteur de recherche comme Google. Elle a également mis en cause une directive de 2006 qui imposait aux opérateurs de conserver les données de connexion de leurs utilisateurs de 6 mois à 2 ans, mesure qui selon elle était disproportionnée par rapport à la finalité poursuivie. Dans un arrêt du 6 octobre 2015, elle invalide un accord Safe Harbor qui depuis dix ans, a permis aux monopoles américains, de transférer dans leurs gigantesques « data centers », des

5 Conseil d'État, *Le numérique et les droits fondamentaux*, Étude annuelle, La Documentation française, Paris, 2014.

milliards de données personnelles sur les Européens sans leur donner les garanties et les protections exigées par leurs lois.

On observe dans les sondages une plus grande attention du grand public à la protection de ses données. Pour se protéger, une minorité dotée d'un solide capital culturel mais aussi technique, recourt au cryptage des messages ou à des solutions alternatives qui préservent l'anonymat sur les réseaux sociaux ou les moteurs de recherche[6].

### UNE REMISE EN CAUSE DE LA TRANSPARENCE

Les règles de protection des données traditionnelles ne sont plus adaptées au Big Data. Pour que l'individu puisse s'autodéterminer informationnellement, il faut recourir à des solutions plus radicales qui passent par la participation aux projets et surtout par une remise en cause de sa transparence.

« Code is law » note un juriste américain[7] pendant qu'un juriste français considère que de nos jours une normativité algorithmique a pris la place de la normativité juridique[8]. On constate en effet que des dispositifs techniques imposent des normes à leurs utilisateurs. Jusqu'à aujourd'hui, une préférence s'est manifestée pour le choix de technologies identifiantes ; il faut revenir sur ce choix et donner la préférence à des technologies qui respectent l'anonymat. Une approche appelée « Privacy by design » entend tenir compte des règles de protection de la vie privée, dans la conception même des technologies. La construction actuelle d'un internet des objets devrait permettre de mettre en œuvre cette approche.

Dans le présent, il conviendrait de recourir à des technologies propres qui ne comportent pas de mouchards ou de faille facilitant leurs prises de contrôle par des tiers. Il faudrait surtout revenir sur l'automatisation de la collecte moment stratégique qui permet l'accaparement et l'accumulation ininterrompue des données personnelles. Cette automatisation rendue invisible ne permet pas l'expression d'un consentement et d'une critique. Comme le préconise l'Electronic Frontier Foundation américaine : « Il

6 CECIL (Centre d'études sur la citoyenneté, l'informatisation et les libertés), *Guide de survie des aventuriers d'Internet*, Juillet 2015, URL : www.lececil.org

7 Lawrence Lessig, « Code is law ». On liberty in cyberspace, Harvard, *Harvard Magazine*, Janvier/février 2000.

8 Alain Supiot, *La gouvernance par les nombres*, Paris, Fayard, 2015. Benjamin Coriat (dir.), *Le retour des communs. La crise de l'idéologie propriétaire*, Paris, Les liens qui libèrent, 2015.

faut construire des systèmes qui ne collectent pas les données en premier. Il n'y a pas de meilleur substitut à la protection que le non-enregistrement des informations ». La pose de capteurs ou la communication des mesures qu'effectue une personne sur elle-même, peuvent être placées sous son contrôle. Il n'en est pas de même avec des supports numériques comme la carte bancaire, le téléphone portable ou internet qui collectent à son insu une information de retour pour le plus grand profit des entreprises. Il est urgent d'adopter des cartes bancaires qui ne délivrent pas cette information de retour ou des smartphones qui ne géolocalisent pas. Sur internet, la seule solution pour le moment est de privilégier les applications qui respectent l'anonymat.

Rendu moins transparent, l'individu serait mieux à même de négocier sa participation à des projets de Big Data. Des projets publics peuvent être facilités par l'obligation de donner des informations notamment en matière de santé mais naturellement en accordant des garanties. Dans le secteur privé, on peut penser que ce sont avec les entreprises qui ont la politique de confidentialité la plus stricte que des négociations auraient le plus de chance d'aboutir. La meilleure solution serait de renouer avec l'esprit communautaire et coopératif des débuts d'internet à la base de la création de l'encyclopédie Wikipédia ou du mouvement du logiciel libre. On peut envisager une mutualisation des données sous le contrôle de participants en mesure de discuter des finalités poursuivies, de la réutilisation possible de leurs informations, de la durée de conservation, de l'utilisation de données sensibles. C'est cette solution d'un commun[9] ni public ni privé, qui peut garantir la maîtrise des individus sur leurs informations ainsi que la transparence des décisions et des algorithmes. Ce commun serait la meilleure manière de poser des limites aux automatismes et aux invisibilités de notre présente condition post-orwellienne.

André Vitalis
MICA /
Université Bordeaux-Montaigne

9 Benjamin Coriat (dir.), *Le retour des communs. La crise de l'idéologie propriétaire*, Paris, Les liens qui libèrent, 2015.