Données et santé
valeurs, acteurs et enjeux

Arthur Charpentier et Raphaël Suire

Maîtres de conférences, Université de Rennes 1

Les données numériques nous concernent tous. En tant qu'usagers d'objets connectés ou de services, nous laissons des traces à mesure que nous utilisons, consultons, notifions, commentons des contenus ou des services. D'ailleurs, même quand l'usager ne fait rien, l'objet ou la plateforme qui proposent du contenu sont capables de remonter cette inactivité, qui constitue bel et bien une information. En soi, chaque trace recueillie isolément donne peu d'information sur qui nous sommes. Ces données numériques, à condition de savoir les stocker et les croiser, équivalent à de l'or noir et à un carburant aux performances éprouvées pour de nombreuses organisations. Cependant, cette exploitation représente à bien des égards une source de tensions entre des usagers sensibles à la protection de leurs données personnelles (notion de privacy concerns*) et ces mêmes acteurs. Elle interroge également en profondeur le régulateur, car les réels gagnants seront peu nombreux. Il faut entendre ici ceux qui au final vont posséder et capturer la valeur des données agrégées. Enfin, c'est sur un nouveau terrain de jeu que pourraient s'exercer ces tensions. Et les promesses associées à la santé connectée sont stratosphériques.*

Données en silo vs données transversales

La question de la collecte des données, et de leur valorisation, est antérieure à la dématérialisation massive du quotidien des usagers et des individus. Au premier rang, les banques, les assurances, et ensuite les commerces de détail, tous ont mobilisé des compétences afin de mieux comprendre les comportements de leurs usagers, de fidéliser ces derniers ou de limiter les phénomènes d'antisélection (correspondant notamment au cas où l'assuré détient davantage d'informations que l'assureur). D'une certaine manière, les données collectées sont circonscrites aux métiers de l'entreprise et aux usages ou aux services offerts par ces entreprises. Par ailleurs, elles sont souvent explicitement produites par les usagers. En dehors de la question spécifique de l'asymétrie d'information concernant les caractéristiques des assurés, le consommateur qui passe à la

caisse du supermarché ou qui consomme des produits bancaires laisse explicitement la trace de ce qu'il a fait. L'analyse de ces traces fera le reste afin de définir un *cluster* d'appartenance et formuler des offres plus adaptées à ses pratiques. La dématérialisation massive du quotidien de l'usager, en particulier à travers un Smartphone, ne change pas tant son comportement d'achat au supermarché ou à l'égard de sa banque (quoique), mais offre surtout de nouvelles opportunités de mieux comprendre qui il est, dès lors qu'il va laisser des traces à bien des endroits. Le contact client-usager peut alors s'envisager non plus physiquement, à un moment précis et dans un lieu défini. Collecter ces données potentiellement massives, et en retirer de la valeur, nécessite de nouvelles compétences, et les organisations traditionnelles ne sont pas toujours les mieux placées pour cela.

Que peut-on faire de ces données massives ?

Comme le rappelle Rochelandet [2010], l'exploitation des données personnelles se fait le plus souvent au détriment d'une certaine forme de respect de la vie privée. Aussi, plus on utilise intensément le numérique, et ce dans des dimensions variées d'usage de services et d'objets (téléphones mobile et fixe, télévision, tablette, objets connectés, voiture, maison, etc.), et plus ceux qui offrent ces services en savent à propos des usages et donc des usagers. L'antisélection, historiquement en faveur de l'assuré, menace d'être inversée avec des assureurs qui pourraient être plus informés que les assurés eux-mêmes. On pourra penser à la chaîne de magasins Target, dans la banlieue de Minneapolis, qui offrit à une adolescente de 17 ans plusieurs bons de réduction nominatifs pour des produits destinés aux futures mamans (berceaux, vêtements pour bébé, etc.), après avoir prédit sa grossesse [Hill, 2012]. Target avait réussi à anticiper (à partir de produits achetés davantage – significativement – par des femmes enceintes), à quelques jours près, à quel stade de grossesse l'adolescente se trouvait. La maladresse a été d'envoyer les coupons de réduction par la poste, le courrier ayant été ouvert par les parents de l'adolescente. Cela a ainsi posé de manière frontale la question fondamentale de l'utilisation des données et de l'information qui peut en être extraite.

L'assurance est un transfert de risques (de l'assuré vers l'assureur), pour une période donnée, moyennant le versement d'une prime ou cotisation versée en début de période. Récolter de l'information pendant la période de couverture ne devrait pas avoir d'impact sur la prime. Pour autant, le numérique peut changer la donne et faire évoluer la tarification vers un modèle de *pay-as-you-do*. Ainsi le principe de *pay-as-you-drive*, qui propose de moduler la prime en fonction de la conduite de l'assuré, utilisé en assurance auto. Mais à partir de là, les déclinaisons du *pay-as-you* peuvent être multiples. Manger, dormir, s'activer, interagir, courir, marcher… autant d'attitudes que l'on peut désormais observer et mesurer à l'aide d'objets connectés les plus variés et qui impliquent éventuellement une contrepartie monétaire. À la baisse, mais également à la hausse. On pourrait alors voir ces principes de tarification comme une remise en cause du mécanisme fondamental de l'assurance, supposant un calcul de prime ex ante. Mais il est aussi possible de noter que ces données permettent de réduire le risque d'aléa moral, car les efforts de prévention sont alors observables.

En dehors du champ de l'assurance, la contrepartie de cette divulgation d'informations est une amélioration de l'offre de services ou des versions – ou déclinaisons – de services en cohérence forte avec les préférences révélées par les usagers [Varian, 1997]. Cela vaut également pour des stratégies de discrimination tarifaire, comme le notent Acquisti et Varian [2005]. Une partie du succès d'Amazon, ou encore de Netflix, repose sur cette maîtrise et l'analyse des traces laissées sur leurs services respectifs. Des algorithmes fondés sur du *machine learning* (ou « apprentissage machine ») permettent de suggérer et d'anticiper des choix et des contenus. De la même manière, en assurance, on imagine aisément que suivre les traces laissées par un assuré peut en dire long sur son aversion au risque,

sur son niveau de prudence, etc. Alors, qui propose ces interfaces et ces censeurs ?

Les Gafa (Google, Apple, Facebook, Amazon) et Microsoft comptent parmi les grandes plateformes qui sont aujourd'hui en pointe dans cette maîtrise du prédictif et de la qualification fine des audiences croisées. D'ailleurs, Google travaille désormais avec les réseaux de surveillance épidémiologique puisque, mieux que les médecins, il sait à partir des requêtes ce que sont les signaux très faibles d'une épidémie de grippe par région, par exemple (Google Flu). Ainsi, ce sont ces acteurs qui aujourd'hui recrutent massivement les *data scientists*, ce sont également eux qui cherchent à agréger toujours plus de services afin de proposer aux usagers des écosystèmes souvent verrouillants pour qui ne les comprend pas [Suire, 2015]. Cet oligopole est la conséquence d'un mécanisme de marché très puissant. En effet, une grande partie de la création de valeur repose sur la capacité à capitaliser et à développer des externalités de réseaux pour les usagers. Ainsi, l'utilité d'un service ou d'un usage va croître avec le nombre d'usagers du service (site d'échange de biens par exemple) ou avec celui des usagers d'un autre type de service (sites de rencontres par exemple), de telle sorte que l'on dit souvent que si les barrières à l'entrée sont relativement faibles sur des marchés inexistants, ces mêmes marchés deviennent très peu contestables une fois des positions acquises. Les effets de réseaux se nourrissant des effets de réseaux, la configuration à l'équilibre est souvent du type *winner takes all*, ou *winner takes the most* plus fréquemment. Comme dans un modèle d'urne de Polya, celui qui part le premier peut durablement conserver son avantage comparatif [Arthur, 1989]. Il en va ainsi du marché de la recherche en ligne (Google), du marché de la messagerie Web (Gmail), du marché des biens culturels (Amazon), du marché des interactions sociales (Facebook), du marché du streaming vidéo (YouTube), etc. Beaucoup de prétendants mais finalement peu de gagnants. Le principe de l'agrégation des services sur une même plateforme ou de la centralisation des accès à une offre large et diversifiée de services est la conséquence et la cause de cette structure de marché. Ainsi, celui qui utiliserait une Google Car (qui se conduit seul, rappelons-le), puis qui commanderait un dîner à partir de son compte Gmail et qui paierait avec son compte Google Wallet dirait en l'espace de très peu de temps à Google où il est, quelle est la vitesse du véhicule, ce qu'il va manger, combien de personnes seront là, etc. Un profilage à la valeur conséquente.

L'exploitation de ces données personnelles se fait parfois à l'insu des usagers mais aussi avec leur assentiment, ceux-ci n'étant pas prêts à supporter les coûts d'une solution dite « *pro-privacy* ». Au fond, l'on fait avec le plus simple, car c'est toujours le principe de moindre effort qui domine ; chercher une alternative et comprendre comment se rendre anonyme reste coûteux [Acquisti, 2004]. Il n'en reste pas moins que les entreprises du numérique (ou *pure players*) sont très souvent « *data driven* » et savent, presque par construction, que l'exploitation de ces données constitue une grande partie de leur valorisation. Ce n'est souvent pas le cas des entreprises traditionnelles qui s'engagent avec plus ou moins d'entrain sur la voie de la transformation digitale. Le marché de la santé connectée n'échappe d'ailleurs pas à cette observation. Il est actuellement en pleine ébullition, et beaucoup d'acteurs prennent position ou renforcent des positions déjà acquises.

Le marché de la santé connectée

L'idée qu'avec une vie numérisée il y aurait une opportunité pour rendre nul un risque d'aléa moral est à relativiser. En effet, s'il est théoriquement possible d'agréger l'ensemble de l'information et le comportement de l'usager à partir des traces qu'il laisse, alors le contrat qui lui est proposé et la prime associée pourraient être juste en proportion du risque assuré. On en est loin. D'abord parce que ce marché est aujourd'hui particulièrement balkanisé – et très loin d'un point d'équilibre – et ensuite parce que les données de santé sont hautement sensibles.

La première source de données numériques individuelles et de santé est liée aux objets connectés, et en particulier à tous ceux qui permettent de mesurer une performance ou une action individuelle (*quantified self*). C'est un marché dynamique avec des objets populaires, bracelets (Fitbit), montres (Apple, Samsung, Nike, etc.), mobiles (Runkeeper, Runtastic, etc.), qui capturent les pas, le poids, les courses à pied, etc. Ce sont des assistants individuels qui fonctionnent en silo et qui peuvent aider, tels des « concierges », à améliorer sa qualité et son hygiène de vie. Évidemment, multiplier les objets et des systèmes non interopérables est peu pratique pour l'usager et pousse à une standardisation ou à un accès unifié à ses objets. C'est le sens du HealthKit d'Apple, qui permet d'alléger le travail des développeurs d'applications santé en utilisant des briques logicielles simplifiées. Samsung avance d'ailleurs dans le même sens avec le Simband. Nous pourrions également évoquer IBM, qui, pour asseoir sa position dans le domaine de l'e-santé avec son supercalculateur Watson, vient de faire l'acquisition de la société Truven Health Analytics et de récupérer, contre 2,6 milliards de dollars, les profils de 215 millions de patients ainsi que des informations sur de nombreux cliniciens et épidémiologistes et autres personnels de santé.

Fin avril, le *New Scientist* (1) révélait que les dossiers médicaux de 1,6 million de patients londoniens (tous patients de trois hôpitaux gérés par le National Health Service) avaient été transférés à une entreprise appartenant au groupe Google. Le but annoncé était de tester des algorithmes d'« apprentissage machine » sur des patients souffrant d'insuffisance rénale aiguë. Des données importantes – sur un historique de cinq ans – pour aider à détecter et à guérir de telles maladies ont ainsi été transférées, mais aussi des données sensibles (tous les rapports quotidiens de l'hôpital, les résultats d'examens mais aussi les mentions d'overdoses, avortements, VIH, etc.). Ces données sont bien évidemment à très haute valeur car non seulement elles permettent, à l'échelle d'un pays, d'orienter une politique publique de prévention de risques sanitaires mais aussi, à l'échelle des particuliers, de limiter les risques d'accidents et/ou de traitements individuels coûteux. Ainsi en va-t-il, une nouvelle fois, de Google, qui sait, nous l'avons évoqué, quel conducteur nous allons être, l'énergie que nous consommons (thermostat et maison connectés) mais qui sait aussi le nombre de pas que nous faisons, notre rythme cardiaque, etc. « *Don't be evil* » est la devise de Google. On aimerait le croire, mais les possibilités des capteurs n'ont pas de limites tant que la promesse de mieux-être suscite l'adhésion et que le coach artificiellement intelligent convainc. À moins que les assureurs, encore attentifs, n'entrent dans la partie.

Ce sont en effet les premiers intéressés par cette maîtrise de la quantification du soi. Non seulement, comme beaucoup, ils amorcent leur transformation digitale avec retard mais stratégiquement une question reste ouverte. Comment, dans un contexte de jeu à somme nulle, encourager seulement les bons comportements sans jamais sanctionner les mauvais élèves ?

Les différents acteurs de la santé connectée

À l'évidence, cette maîtrise à 360 degrés du comportement est hors de portée des assureurs. Il est par conséquent fort probable que des alliances avec les industries numériques soient à attendre. Cela peut être au gré des concepteurs d'objets connectés (voiture, bracelets, etc.) ou alors avec les Gafa, qui sont les mieux placés concernant cette maîtrise omnicanale du comportement. Ils peuvent également s'engager dans des consortiums de plateformes plus ou moins ouvertes de données de santé (Human API, healthCare.gov, etc.) avec un objectif d'interopérabilité des acteurs et des formats de données. Mais les assureurs pourraient également faire le choix d'accompagner tout simplement les usagers (ou Internet) dans l'utilisation des « traqueurs » de santé en devenant un partenaire au long de la vie et en développant un lien de proximité avec les assurés [Best, 2015].

Et puis, comme souvent sur les marchés numériques, il y a les disrupteurs de la chaîne de valeur. Ceux que l'on n'attend pas et qui peuvent modifier en profondeur le paysage en redistribuant la valeur des données numériques. C'est le cas de Lenddo, qui cible les individus des pays en développement, envers lesquels les méthodes de scoring traditionnelles s'avèrent très discriminantes (peu de revenus stables, pas de logement, etc.). Dès lors, lenddo.com propose d'estimer, à partir des liens sociaux dérivés des profils numériques (Facebook, Twitter, LinkedIn, etc.), qui est l'usager et d'établir un indicateur de confiance qui est revendu aux prêteurs ou aux assureurs. Au fond, la nature et le type de nos interactions sociales numériques se substituent au scoring traditionnel. C'est d'ailleurs une intermédiation qui pourrait se généraliser avec l'apparition récente des *data brokers* [Tanner, 2016]. Ce sont des intermédiaires qui achètent des données de santé aux usagers et les revendent à qui est susceptible de les convoiter. La réglementation autour de ces acteurs est encore très flottante, et les usagers peuvent ne pas savoir ce que l'on saura d'eux après agrégation de différentes données, cela laissant potentiellement le champ libre à des stratégies particulièrement discriminantes à leur endroit.

Quelle place pour la vie privée ?

Comme le notaient Acquisti et Varian [2015], les firmes collectent des données non seulement pour discriminer par les prix mais également pour personnaliser leurs offres en les associant à des services complémentaires, eux aussi davantage personnalisés. Cette stratégie devrait augmenter le coût d'opportunité de l'anonymat. Dans ce contexte, prétendre vouloir avoir une vie privée devient forcément suspicieux. Posner [1981] fait ainsi le parallèle entre la notion de *privacy* et celle de « vice caché » en droit commercial. La *privacy* serait une forme de « protection légale illégitime des pratiques fondées sur la tromperie », pour reprendre la terminologie de Rochelandet [2010], en instaurant une rente de situation à l'avantage des individus bénéficiant d'une asymétrie informationnelle juridiquement protégée. Hermalin et Katz [2006] notent que, dans le contexte de l'assurance santé, en cas d'information parfaite, les personnes en mauvaise santé (ou susceptibles de l'être) se verraient infliger des prix plus élevés, et donc certaines d'entre elles seraient probablement exclues du marché. En l'absence de *privacy*, lorsque les individus sont obligés de divulguer leur état de santé, il y a une forte incitation à ne pas effectuer d'examen de santé, ce qui pourrait constituer un des travers de cette approche. Marisol Touraine vient de rappeler que les assureurs n'avaient pas la possibilité de collecter les données individuelles de santé. Une réglementation « en silo » qui n'augure rien de bon, puisque les Gafa collectent comme ils l'entendent, et souvent de manière très partielle et incomplète, des données sur les comportements individuels, y compris ceux ayant trait à la santé. « *Ignorance is bliss* », comme on dit, en espérant que l'ignorance soit autant du côté de l'assuré que de celui de l'assureur. Elle l'est sans doute beaucoup moins du côté des géants du numérique.

Note

1. Revealed: Google AI Has Access to Huge Haul of NHS Patient Data", newscientist.com, 29 avril 2016. https://www.newscientist.com/article/2086454-revealed-google-ai-has-access-to-huge-haul-of-nhs-patient-data/

Bibliographie

ACQUISTI A., "Privacy in Electronic Commerce and the Economics of Immediate Gratification", in *EC'04: Proceedings of the 5th ACM Conference on Electronic Commerce*, New York, ACM Press, 2004, p. 21-29.

ACQUISTI A. ; VARIAN H. R., "Conditioning Prices on Purchase History", *Marketing Science*, vol. 24, n° 3, 2005, p. 367-381.

ARTHUR B., "Competing Technologies, Increasing Returns and Lock-In by Historical Events", *Economic Journal*, vol. 99, n° 394, 1989, p. 116-131.

BEST J., "Yes, Insurers Want Your Health Data But Not for the Reason You Think", zdnet.com, 3 novembre 2015. *http://zd.net/1XM5GKZ*

Grand View Research, "E-Health Market Analysis", 2015. *http://bit.ly/1xCw298*

HERMALIN B. ; KATZ M., "Privacy, Property Rights And Efficiency: The Economics of Privacy as Secrecy", *Quantitative Marketing and Economics*, vol. 4, n° 3, 2006, p. 209-239.

HILL K., "How Target Figured Out a Teen Girl Was Pregnant Before Her Father Did", forbes.com, 16 février 2012. *http://onforb.es/1UN3QFB*

POSNER R. A., "The Economics of Privacy", *The American Economic Review*, vol. 71, n° 2, 1981, p. 405-409.

ROCHELANDET F., *Économie des données personnelles et de la vie privée*, coll. « Repères », La Découverte, 2010.

SUIRE R., « La déconnexion volontaire : nouvelle fracture numérique », inaglobal.fr, 1er juin 2015.

TANNER A., "How Data Brokers Make Money Off Your Medical Records", scientificamerican.com, 1er février 2016.

VARIAN H., "Versioning Information Goods", Working Paper, University of Berkeley, 1997.