

# Données de santé : le temps de la maturité et de l'ouverture ?



Avner Bar-Hen

Professeur de Statistiques, Université Paris Descartes, Pôle de recherche et d'enseignement supérieur Sorbonne Paris Cité

---

## La santé : un bien public

L'industrialisation et l'émergence des préoccupations sanitaires ont favorisé la mise en place d'une protection des travailleurs dès le milieu du 19<sup>ème</sup> siècle (protection des salariés les plus faibles avec la loi de 1841, inspirée du rapport Villermé, puis lois sur l'inspection du travail, les accidents du travail, les maladies). La protection n'a cessé de se renforcer tout au long du 20<sup>ème</sup> siècle. De par son étymologie, la statistique a été, dès son origine, un "outil de gouvernement" et il n'est donc pas surprenant que la santé soit l'un des domaines importants en termes de collecte des données.

Le lien entre politique publique et données fiables a été abondamment discuté (voir par exemple A. Desrosières : Histoire de la raison statistique<sup>1</sup>) et l'importance d'une information de qualité en santé est primordiale pour préserver un système de soins juste et efficace. Ceci est d'autant plus important dans un contexte (i) d'espérance de vie en constante augmentation, (ii) d'une spécialisation des traitements et des examens qui augmentent sensiblement les coûts et (iii) du rétrécissement du périmètre de l'Etat et de la puissance publique par des économies budgétaires.

La loi, notamment en France, encourage les citoyens à s'impliquer dans les données publiques. En effet, la loi du 17 juillet 1978 :

reconnaît à toute personne un droit très large d'obtenir communication des documents détenus par une administration, quels que soient leur forme ou leur support.

Ainsi, au nom de la citoyenneté et de la transparence des actions des gouvernements, de l'amélioration de la gestion du territoire, ou encore de la promotion de l'innovation, les données publiques ont donc une vocation naturelle à être partagées. Leur ouverture est d'autant plus facilitée qu'elle peut répondre à des objectifs politiques ce qui inclut évidemment la santé.

## Une mutation numérique qui bouleverse les codes et les habitudes

L'arrivée massive du numérique dans notre quotidien tend à créer des nouveaux modes d'implication dans l'espace public. Il en résulte un affaiblissement des corps intermédiaires dans les rapports sociaux et politiques. Les politiques publiques ne sont plus l'apanage des responsables politiques, marquant un changement de paradigme important dans nos démocraties. Beaucoup d'acteurs s'efforcent de construire des propositions sans passer par la "case" parti et revendiquent d'incarner l'intérêt général sans passer par la "case" Etat. La situation de quasi-monopole de la production de propositions politiques détenue par les

---

1. DESROSIERES, A., 2010, La politique des grands nombres. Histoire de la raison statistique, La Découverte/Poche, Paris, 3<sup>ème</sup> édition.

partis s'écorne. En parallèle l'élévation continue du niveau d'étude donne aux citoyens des outils (et crée des attentes) pour appréhender les questions de santé. La suite logique est donc une réappropriation de l'action publique et non plus une délégation aux responsables politiques, surtout lorsque leur capacité d'action est limitée et onéreuse. Il est donc naturel de voir se développer une demande de démocratie scientifique et des politiques volontaires sur l'ouverture des données publiques. A titre d'exemple, l'épisode de pollution de mars 2014 dans les grandes villes a montré la difficulté pour le citoyen de penser qu'il est une partie éventuelle de cette pollution par son comportement et il y a une forme de lâcheté à laisser aux politiques le soin de prendre des mesures très onéreuses pour un résultat limité d'autant plus que la volonté politique se heurte au poids de lobbies parfois puissants. La faiblesse de la mobilisation sur les effets délétères du diesel traduit sans doute la difficulté du positionnement des professionnels de santé et le faible poids de leur ministère de tutelle pour se faire entendre vis à vis d'acteurs économiques publics et privés. A contrario l'émergence de l'épidémie VIH a démontré la capacité de mobilisation importante du mouvement associatif concerné amenant à une forme originale de "partenariat" entre les associations de patients et les professionnels de santé, avec une réelle efficacité dans l'évaluation scientifique des traitements anti- rétro- viraux et le contrôle de la maladie, même si les conditions socio- culturelles du développement de l'épidémie ne sont pas directement extrapolables à d'autres situations morbides, notamment en terme d'égalité d'accès aux traitements.

## La santé, un monde à part dans cette évolution ?

L'un des domaines qui rencontre le plus de résistance à ce mouvement est la santé. Les arguments les plus fréquemment avancés sont la complexité des données, la garantie du secret médical et de la protection des données personnelles. Il y aurait donc une spécificité des données de santé.

La protection des données personnelles est un domaine très largement travaillé par la statistique publique et si la question est loin d'être anodine, il y a nombre d'experts et d'expériences au sein de l'administration française. Le but est d'avoir une information de qualité avec une protection stricte et rigoureuse des règles de confidentialité et de respect des libertés individuelles. Pourquoi ne pas s'inspirer des travaux qui permettent d'accéder à la production industrielle, aux revenus des ménages, aux votes ou à nombre d'informations à des niveaux de précision tout à faits pertinents pour l'ensemble des acteurs ? Différents niveaux de recueil existent, et les données ne sont publiées qu'avec un nombre minimal. Ce concept de degré d'agglomération est classique en statistique publique. Le suivi statistique sur une certaine durée soulève, lui, un tout autre problème, qui est celui de pouvoir suivre le parcours d'un certain nombre d'individus sur un certain laps de temps sans avoir à connaître leur identité véritable. Sous réserve d'une extension de ses compétences, la commission du secret statistique pourrait représenter un acteur central dans ce débat.

## L'OpenData n'est pas le Big Data

Il y a une confusion classique entre l'OpenData, qui implique de manière très claire l'absence de données personnelles et le respect du secret statistique, d'un côté, et, de l'autre, le "Big Data", ce secteur émergent de "l'économie numérique" qui se nourrit quasi-exclusivement de données à caractère personnel et qui alimente à juste titre d'importantes questions sur le respect par ses entreprises de la vie privée des citoyens. Si cette confusion est logiquement entretenue par les lobbyistes du "Big Data", elle l'est également par les opposants à la transparence démocratique qui abusent du faux prétexte de données personnelles pour refuser d'ouvrir des données absolument non sensibles. En réalité l'Open Data est à l'origine d'une prise de conscience sur un meilleur respect des données personnelles en permettant de détecter le non-respect des

règles en vigueur en termes d'anonymisation et de respect des données personnelles. Ceci a pu être vu dans deux cas récents :

1/ les données du système d'immatriculation des véhicules (cartes grises) ont été vendues en toute illégalité pendant des années et la légalisation de leur revente n'a pas réglé le problème de divulgation des données personnelles<sup>2</sup>

2/ à l'occasion des travaux préparatoires au passage en Open Data, des problèmes d'anonymisation ont pu être identifiés et finalement corrigés<sup>3</sup>.

Le secret médical n'est pas que du ressort des bases de données publiques. A côté de la collecte publique se développe un nombre important de nouvelles données soit personnelles<sup>4</sup>, soit indirectes (Google trends). Le risque donc est une dévaluation de la donnée publique au profit d'une information pas toujours vérifiée et pas toujours de qualité. Il serait certes inadmissible que l'on puisse identifier le porteur d'une maladie chronique à partir des données médicales mais dans le même temps on ne s'interroge qu'assez peu sur la détection de cette même maladie à partir d'un téléphone portable. Le magasin américain Target a pu identifier des femmes enceintes à partir de leur profil d'achat sans avoir aucunement besoin de données médicales. L'exemple du père se plaignant des publicités reçues et découvrant par ce biais la grossesse de sa fille a plus été vu comme un succès algorithmique que comme une intrusion dans la vie privée.

Ceci illustre une autre source de confusion : la protection de l'anonymat recouvre en réalité la protection de la vie privée des individus. Le consentement éclairé, la possibilité d'exclusion des bases de données devrait donc être tout autant en débat que l'anonymisation. Il est aussi pertinent de se demander qui doit être le garant de la vie privée. Le désengagement de l'Etat, sa versatilité<sup>5</sup> dans la gestion des données peuvent amener à se poser des questions sur son rôle naturel. Doit-on avoir plus confiance dans des organismes de recherche alors, qu'au contraire des services ministériels, la certification morale et technique est inexistante ? L'expertise des organismes de recherche ne se résume-t-elle pas à une auto-affirmation, qui mérite sans doute d'être questionnée régulièrement ?

## L'OpenData est utile à la démocratie

Les récentes affaires de santé publique (dont la plus emblématique est celle du Médiateur) montrent les limites de la gestion de l'information en santé dans notre pays. Rappelons que la plus grande partie des données recueillies par le système de santé est destinée, au départ, à sa seule gestion financière (en vue du remboursement des actes ou de la tarification de l'activité dans les hôpitaux).

Par ailleurs, ces données ne concernent que la part de la médecine prise en charge par les organismes sociaux, et elles ne sont disponibles que de manière très restreinte et parfois avec un retard de plusieurs années. Enfin, des procédures d'anonymisation apportent déjà un niveau très élevé de garantie de l'anonymat du traitement de l'information, au moins égal à celui mis en place dans nos pays voisins d'Europe du Nord.

Il est aujourd'hui possible de connaître les prescriptions mensuelles de chaque généraliste en Grande-Bretagne avec moins de six mois de délai, alors que les honoraires des professionnels de santé ne sont plus actualisés depuis 2010 en France. L'accès aux données de santé permettra

2. <http://www.senat.fr/questions/base/2012/qSEQ120700184.html>

3. [http://www.autorite-statistique-publique.fr/pdf/Actualites/ASP\\_PV\\_18\\_04\\_13.pdf](http://www.autorite-statistique-publique.fr/pdf/Actualites/ASP_PV_18_04_13.pdf)

4. Le quantified self: nouvelle forme de partage des données personnelles, nouveaux enjeux ? Sophie Vuilliet-Tavernier CNIL  
[http://www.cnil.fr/fileadmin/documents/La\\_CNIL/publications/DEIP/LettreIP\\_5.pdf](http://www.cnil.fr/fileadmin/documents/La_CNIL/publications/DEIP/LettreIP_5.pdf)

5. Pour l'exemple anglais voir : <http://www.theguardian.com/commentisfree/2014/jan/20/nhs-patient-care-data-policy-medical-information>

peut-être d'anticiper des problèmes, mais offrira aussi une information a posteriori : tel médicament est-il plus prescrit dans telle région ou telle ville ? Plutôt aux personnes âgées ? Aux hommes ? Il s'agit de conceptions socioculturelles différentes selon les pays. Tant que l'activité médicale sera tarifée à l'acte, il n'y aura pas de place pour une médecine sociale qui pourrait par exemple comprendre (comme en Angleterre) une rémunération pour la participation à la collecte des données de santé.

Revenons enfin sur la complexité des données : en quoi les données de santé sont-elles plus compliquées que celles de n'importe quel ministère, les données d'organismes internationaux, d'un moteur de recherche ou d'un réseau social (sans parler des questions de confidentialité dans ce dernier exemple) ?

Aujourd'hui, l'accréditation d'accès aux données de santé nécessite un décret ministériel, au même titre que l'accès à des terrains militaires, en raison des exigences de sécurité. L'accès aux données publiques de santé est restreint même une fois celles-ci totalement anonymisées, et cela doit changer ! Il n'est pas normal que seulement les "bien-pensants" et les "bien-sachants" triés par les organismes sociaux détenteurs de ces informations aient droit d'accéder aux données publiques de santé. Il faudrait que tout citoyen puisse réclamer l'accès permanent à ces données publiques anonymes. La collectivité accepte une restriction des libertés pour le bien commun : c'est le principe de l'obligation de la ceinture au volant ou de l'interdiction des armes dans les avions. A moins de considérer tout utilisateur de données comme un criminel en puissance, il semble plus raisonnable de prévoir des sanctions pour une utilisation abusive de ces données (sinon il faut aussi interdire la vente des fourchettes qui peuvent servir à crever les yeux des passants dans la rue). Les données de santé représentent un bien social qui n'est pas la propriété de d'agence, d'unités de recherche ou de gourous autoproclamés spécialistes universels.

## La dimension économique de la santé

La concomitance du débat sur l'ouverture des données de santé avec la diminution des budgets de recherche voire la mise en cause de programmes de prévention et de surveillance est troublante. Il est légitime de relier ce mouvement d'ouverture des données à une tendance profonde qui considère la santé comme un produit économique classique. Les données de santé représentent clairement un bien économique par leur potentielle réutilisation dans un contexte marchand : il y a donc une forme de symétrie entre les données de l'assurance maladie et les informations que détiennent les industriels à travers les données de vente. A titre d'exemple, relier les ventes aux profils de prescription est (entre autres) un enjeu économique. Au lieu de refuser toute évolution du système, ne vaut-il mieux pas se demander comment les gains attendus peuvent être partagés entre les différents acteurs, dont les patients et le secteur public (prévention, recherche, etc.) ?

## L'OpenData, une source de progrès pour la santé

Le suivi en temps direct du fonctionnement des moteurs d'avion permet une bien meilleure gestion des pannes. De la même manière, un suivi individualisé a de grandes chances d'être plus efficace qu'un profilage approximatif qui autorise des raccourcis plus ou moins douteux. Au lieu de proposer des dépistages pour tous comme le dépistage du cancer colorectal à 50 ans ou le dépistage de la trisomie 21 par amniocentèse à partir de 38 ans, il est peut-être possible d'arriver à une prévention mieux ciblée, plus efficace et moins onéreuse, évitant de stigmatiser telle ou telle sous-population.

Au-delà d'obscurantistes de la dernière heure accrochés à leurs privilèges, n'avons-nous pas un réel problème de compétence ? Lorsqu'une plate-forme comme Etalab ou datagouv ne possède même pas une interface de programmation (API) pour utiliser les données, pouvons-nous réellement parler de données ouvertes ? Vaut-il mieux financer de nouvelles études ou plutôt chercher à développer l'utilisation des données existantes en réunissant des spécialistes de l'utilisation de la donnée publique, de leur maniement et de leur analyse ? Une publication comme ce numéro du journal Statistique et Société de la SFdS est une opportunité pour faire avancer ce débat et la SFdS a une place naturelle sur cette question importante.

Le débat disparaîtra avec les derniers dinosaures des temps anciens. Cacher une difficulté à gérer une situation nouvelle en se réfugiant derrière un maintien de dogme dépassé peut prêter à sourire. Mais si les discours fondés sur ces données avaient pu circuler un peu plus facilement, un temps précieux au bénéfice de la santé publique aurait peut-être pu être gagné, tout en évitant nombre d'affaires qui sapent la confiance des Français en leur système de santé chaque fois que des lanceurs d'alerte, plus ou moins bien intentionnés et/ ou indépendants, expliquent que les données existaient, mais qu'elles n'ont pas été utilisées. Notons qu'ouvrir les données n'est pas suffisant. Dans le cas du Mediator, la Caisse régionale d'assurance-maladie de Bourgogne avait ainsi visiblement pu lancer une alerte mais l'impression d'un mélange constant des genres fait qu'il est difficile de démêler les volontés de suspendre des remboursements pour des questions de santé publique ou pour des questions budgétaires. Le mélange de l'approche médicale, comptable et parfois scientifique est encore une fois délétère.

Il est temps de faire confiance aux citoyens et qu'ils puissent se réapproprier leur propre système de santé, car c'est la condition de sa survie, dans un respect et une confiance mutuels regagnés entre tous ses acteurs. Si, au nom de principe douteux, quelques spécialistes autoproclamés veulent garder le monopole des données, il serait temps de leur demander des comptes !