



Données massives, *big data* et santé publique : de quoi parle-t-on ?

Quelques définitions autour de la thématique de ce dossier avant d'aborder les enjeux que présente le *big data* pour les individus et pour la santé publique.

Santé publique et *big data* : concepts et définitions

Margo Bernelin

Sonia Desmoulin

Chargées

de recherche CNRS,
UMR 6297 CNRS,
université de Nantes

Thomas Lefèvre

Maître de

conférences,
praticien hospitalier,
Iris-UMR 8156-997

CNRS Inserm EHESS

université Sorbonne

Paris Nord

Santé publique et *big data* sont deux concepts aux définitions variables selon les contextes d'utilisation.

La santé publique

Nous parlons ici de santé publique au sens du domaine qui s'intéresse à la santé des populations, comme complément à la médecine, qui s'intéresse à la santé individuelle. La santé publique présente essentiellement une dimension organisationnelle et de prévention. Bien sûr, la santé des populations est liée à la santé des individus qui les composent, mais l'approche de santé publique est collective. On peut penser aux maladies infectieuses et à la vaccination, par exemple. Une maladie infectieuse de transmission interhumaine concerne l'individu et le groupe : il peut y avoir un intérêt à créer des institutions et des réseaux, à mobiliser des techniques (vaccins) et à adopter des politiques pour tenter de maîtriser, sinon d'éradiquer, une épidémie au niveau d'une population, sans pour autant que l'intérêt individuel direct soit évident pour chacun. Une personne qui pourrait être contaminée, puis véhiculer un virus sans en subir les effets, ou en ressentir des effets mineurs, pourra participer à sa diffusion, notamment vers des personnes qui pourront être symptomatiques, voire subir des conséquences plus graves. La santé

publique va ainsi recouvrir l'organisation du système de santé, des professionnels de santé et l'élaboration de politiques publiques de santé ainsi que les moyens de leur mise en œuvre. Elle implique aussi les personnes dans ces différentes dimensions, notamment selon un principe de démocratie sanitaire affirmant que les citoyens doivent pouvoir contribuer, donner un avis, sur les décisions de santé publique.

Les données

Le terme de données recèle quant à lui plusieurs sens. On peut parler de données de la science : autour de cette notion gravitent celles de données probantes, de preuves ou de faits scientifiques. On peut également parler de données au sens d'une mesure, d'une donnée « brute » – même si aucune donnée n'est jamais « brute », étant le résultat d'une construction sociotechnique. Tous ces types de données peuvent être mobilisés dans une prise de décision. Les données au sens de mesures, encore rares il y a peu, semblent se multiplier ces dernières années. La question de leur accès, de leur utilisabilité et de leur utilité se pose alors.

Big data et données massives

Le terme de *big data*, dont l'utilisation large montre qu'il ne peut être réduit à celui de données massives, ne possède pas de définition consensuelle, a

fortiori en santé et en santé publique. Son apparition dans la littérature scientifique médicale pourrait être rapportée à un dossier de la revue *Nature* publié en septembre 2008. Le terme est souvent attribué à un document technique d'un cabinet de conseil américain (META group/Gartner), apparu en 2001. En réalité, il n'y est pas fait mention du terme. En revanche, on y voit apparaître la base de définitions reprises assez fréquemment dans différents domaines et par la presse généraliste : les définitions en « V ». Par exemple, les 3 V : volume, variété et vélocité (en français), pour caractériser des données d'utilisation de plus en plus courante, ou que l'on souhaiterait exploiter. *Volume*, pour une grande quantité soit de caractéristiques d'un individu, soit pour un grand nombre d'individus (une cohorte de plusieurs centaines de milliers d'individus), ou bien les deux à la fois. *Variété*, pour souligner que l'on va pouvoir exploiter des données de natures diverses, que l'on peut répartir entre données structurées et données non structurées. Schématiquement, des données structurées seraient représentées par un grand tableau bien défini, où toutes les colonnes correspondent à des mesures standardisées (une tension artérielle en mmHg, une taille en cm), et les lignes, autant de personnes ou patients. Des données non structurées seraient toute autre source de données. Cela peut être de la vidéo, de la parole, du texte libre. *Vélocité*, pour le fait que l'on va traiter ces données très rapidement, en « temps réel ». Ce critère est actuellement moins pertinent dans le champ de la santé, même s'il prend de l'importance dans les situations de gestion de crise, mais il a été très important dans un domaine précurseur pour l'usage du *big data*, à savoir la finance et le *trading* « haute fréquence » (le traitement extrêmement rapide de millions et milliards de transactions financières).

Un autre domaine où le *big data* a été très vite présent est celui du marketing et de la publicité : l'idée est qu'en utilisant et recoupant des données diverses et variées de nombreuses personnes, en particulier des données dites comportementales (activité physique : les pas, le rythme auquel on marche ; consommation : ce que dit notre ticket de caisse ; utilisation de réseaux sociaux) ou des données géolocalisées (adresse personnelle, coordonnées GPS), il devient possible de cerner d'une part nos préférences et d'autre part nos caractéristiques associées à ces préférences : la publicité « ciblée » est née. Plutôt que nous soyons tous exposés sur une page Internet ou sur le bord de la route à la même publicité, des publicités pour un produit plutôt qu'un autre s'afficheront, basées sur notre historique de navigation Internet, le contenu de nos emails ou encore les caractéristiques que nous aurons nous-mêmes rentrés dans notre profil d'utilisation d'un réseau social.

Le concept a fait le voyage depuis ces domaines vers la santé. On trouvera ainsi dans la littérature scientifique médicale tant du *big data* que l'utilisation « ciblée »

de données, que l'on dénomme médecine 4 P (sur le modèle des 3 V et plus) : médecine préventive, prédictive, personnalisée et participative.

Les sources de données

Les sources de données en santé se multiplient : données recueillies, numérisées lors d'une hospitalisation ou d'une consultation médicale (diagnostic, traitement, examen médical ou biologique, d'imagerie...), données de consommation de soins (données enregistrées par l'assurance maladie, prescriptions), données d'objets connectés – qu'ils soient médicaux (tensiomètre, implant cardiaque) ou non médicaux à l'origine (pèse-personne, montre connectée, smartphone), et virtuellement toute autre source de données qui pourraient, dans un usage donné, être utiles : données sociodémographiques, géographiques... On comprend qu'en réalité, plus que le caractère massif des données, ce sont deux éléments essentiels qui vont potentiellement introduire une nouveauté : i) le fait de pouvoir accéder puis recouper des données qui d'habitude ne le sont pas, et ii) des moyens d'analyses et de traitement de ces données, comme par exemple « l'intelligence artificielle ». Cependant, ces données peuvent tout autant être analysées par des méthodes plus conventionnelles en médecine et en épidémiologie.

L'intelligence artificielle

L'intelligence artificielle trouve un nouveau souffle, un regain d'intérêt, en partie par le développement de nouveaux algorithmes, mais plus encore par l'accroissement des moyens de calculs (puissance et caractère répandu des ordinateurs) et surtout par l'accessibilité accrue de données couvrant de plus en plus de domaines de la vie quotidienne, y compris la santé. Un attrait et une efficacité renouvelée de l'intelligence artificielle – qui est aussi source de réticence de la part des professionnels de santé vis-à-vis de son utilisation – tiennent dans l'utilisation d'algorithmes qui vont « s'adapter » aux données, et établir des règles de décision non pas introduites *a priori* par l'homme, mais à partir de ce qui a été observé. Cela est bien sûr le fonctionnement théorique, et implique un certain nombre de limites, dont les biais d'apprentissage : un algorithme perpétuera, voire renforcera, les comportements observés et « appris » à partir des données. Ainsi, si les données ne concernaient que des patients de sexe masculin et d'origine caucasienne, rien ne dit que l'algorithme tiré de ces données saura correctement décider pour des patients de sexe féminin ou d'origine asiatique ou afro-caribéenne.

La santé publique de précision

Ces constats sont valables pour la médecine, centrée sur l'individu, mais peuvent l'être tout autant pour la santé publique, centrée sur les populations. On parle notamment depuis quelques années (2013) de santé publique de précision. On pourrait sans doute



tout autant parler de santé publique 4 P, celle-ci ayant dans ses attributions classiques la prévention, la participation (démocratie sanitaire, associations de patients) et la prédiction. Reste la personnalisation : en réalité, en santé publique, on peut chercher comme en marketing à « segmenter » les populations, c'est-à-dire à identifier certains ensembles de personnes qui présentent des caractéristiques similaires par rapport à un problème de santé. Le concept est voisin de celui de stratification du risque : tout le monde n'est pas exposé de la même façon aux mêmes risques de développer une pathologie donnée. À l'inverse, la personnalisation « vraie » – un diagnostic strictement unique, personnel, un traitement unique, personnel – n'existe pas et n'a probablement pas beaucoup de sens en général. De fait, la médecine 4 P est elle aussi basée sur une approche de groupe : le groupe des personnes partageant des caractéristiques similaires, impliquant un même traitement.

Pour l'heure, la santé publique de précision semble

se démarquer selon deux grandes dimensions de la médecine 4 P :

- la précision est une précision en termes d'échelles des mesures, par exemple l'échelle spatiale – en effet, les bons résultats d'un indicateur, comme la mortalité infantile selon les régions, peuvent masquer des hétérogénéités géographiques majeures si l'on y regarde à une plus petite échelle que la région ou le pays, et la précision révèle alors une aggravation de l'indicateur en certains lieux, et une amélioration en d'autres, donc une aggravation des inégalités géographiques de santé ;
- ce qui a toujours distingué l'approche sociale de l'approche individuelle, à savoir qu'il existe des caractéristiques propres au collectif, au social, déterminants de l'état de santé individuel et populationnel, qui ne sauraient se réduire à la somme des caractéristiques individuelles. Il ne s'agit au fond que d'une réactualisation, voire une exacerbation de la tension classique entre préférences et approches individuelles, et préférences et approches collectives. ●

Données massives en santé publique : quels enjeux pour les personnes ?

Margo Bernelin
Sonia Desmoulin
Chargées de
recherche CNRS,
UMR 6297 CNRS,
université de Nantes

Au cœur de la politique de santé publique, « l'observation épidémiologique et la surveillance de l'état de santé des populations » s'appuie désormais « sur les nouveaux outils d'exploitation des données », ainsi que le proclame Santé publique France sur son site Internet. La gestion de la crise sanitaire née de la propagation du virus SARS-Cov-2 (Covid 19) illustre bien l'intérêt de la collecte et du traitement des données massives pour la protection de la santé des populations, à des fins de diagnostics, de suivi de la progression pandémique et de recherche en santé publique. L'exemple montre aussi que sont traitées à la fois des données personnelles, comme le résultat des tests de dépistage PCR, et des données non personnelles, telles que les relevés de présence du virus dans les eaux usées de grandes villes. Pour être efficace, une telle approche suppose de collecter énormément de données, de natures diverses, et issues de plusieurs bases. Caractérisée notamment par les concepts de *volume* et de *variété*, cette démarche a logiquement des répercussions sur les droits des personnes. La protection de la vie privée et des données personnelles est difficile à tenir. La tension, bien connue en santé publique, entre protection des droits et libertés individuels et protection de la population est ainsi réactivée. En effet, s'appuyant en partie sur le traitement de données personnelles, l'usage des données massives en santé publique nécessite le concours des individus au risque de leur vie privée.

L'utilisation des données personnelles de santé

L'usage des données massives en santé publique s'appuie en partie sur la collecte et le traitement de données personnelles, c'est-à-dire de données qui permettent d'identifier directement ou indirectement des individus. Ainsi, noms, adresses, identifiants et autres numéros spécifiquement attachés à une personne sont autant de données personnelles, qui peuvent être utilisées en santé publique. Au sein de ces données, les données de santé tiennent une place particulière en ce qu'elles informent sur l'état de santé passé, présent ou futur d'un individu, qu'il s'agisse de sa santé mentale ou physique. Par exemple, les résultats sanguins ou de radiographie ou les données de remboursement des soins révèlent l'état de santé d'un individu et sont, par conséquent, des données personnelles de santé. Des données plus anodines peuvent également fournir des informations sur l'état de santé, comme une adresse qui mettrait en évidence l'exposition à un environnement pollué et au risque de développer des pathologies spécifiques. Le traitement de ces données variées en masse pourrait, par exemple, mettre en lumière des liens entre la prise d'un médicament, une zone géographique de domiciliation et l'expression d'effets secondaires.

Une double difficulté naît de cette démarche. La première tient à l'impératif de protection de la vie privée des personnes. Étant susceptibles de révéler des informations sur le comportement, l'état de santé et les conditions de

vie des individus, les données personnelles font l'objet d'un encadrement juridique spécifique (loi Informatique et libertés en France depuis 1978 ; Règlement européen sur la protection des données personnelles (RGPD) adopté en 2018). S'agissant des données de santé, les risques d'usages aboutissant à une discrimination néfaste pour la personne concernée sont importants. En conséquence, ces données sont considérées comme sensibles et ne doivent pas, en principe, être recueillies ni utilisées, sauf pour des motifs dérogatoires tels que les soins vitaux, les services de prestations ou de contrôle par les organismes d'assurance maladie, le traitement par les agences régionales de santé ou les traitements nécessaires à la recherche, aux études et évaluations dans le domaine de la santé. Très souvent recherché, le consentement exprès de la personne permet également le traitement de telles données. Cependant, la protection conférée par le consentement peut être discutée dès lors que la personne est en situation de vulnérabilité (un patient en attente de prise en charge ou de soins par exemple) et/ou que le consentement ne correspond en réalité qu'à un acte irréflecti (comme un clic mille fois répété sur un ordinateur ou un téléphone portable). Ainsi, en matière sanitaire, un équilibre délicat doit être recherché entre protection des droits des personnes et objectif de protection de la vie et des populations.

La seconde difficulté vient de la possibilité d'explorer différentes bases de données et, en croisant les éléments ainsi collectés, de remonter de manière indirecte à des informations potentiellement sensibles pour les personnes. Dès lors, la distinction entre les données personnelles et les autres données est mise à l'épreuve. Les recherches en informatique montrent que les techniques d'anonymisation ne sont pas toujours efficaces (par exemple pour les données génétiques) ou qu'elles trouvent rapidement leurs limites lorsque plusieurs bases de données peuvent être croisées.

La tension entre protection des droits individuels et protection de la population

Si l'impératif de protection de la population semble ici prendre le pas sur le souci de préserver l'intimité des individus et les usages discriminatoires, cette préoccupation ne disparaît pas. Elle se traduit dans la mise en place de statuts spéciaux (hébergeurs de données de santé) et d'outils spécifiques (système centralisé sous la responsabilité d'autorités publiques). Cette tension entre protection individuelle et protection collective n'est pas nouvelle. En effet, dès le xx^e siècle des règlements nationaux imposaient aux médecins de déclarer les cas de rougeole ou de scarlatine. Les règlements sanitaires municipaux pouvaient également imposer aux hôteliers et aubergistes les mêmes obligations de déclaration. Ainsi, des données personnelles étaient déjà collectées afin d'endiguer une épidémie par l'organisation de quarantaines.

Aujourd'hui, la protection des données personnelles de santé devient d'autant plus difficile à assurer que les

données des patients ne sont plus seulement traitées à un échelon local (celui du CHU par exemple), mais à un niveau national. Une grande partie des données est concentrée au sein de la plateforme nationale des données de santé Health Data Hub (données sur les soins pris en charge par l'assurance maladie, données de remboursement des soins...). À ces données pourraient s'ajouter d'autres, plus hétérogènes : adresses, géolocalisation, contacts, photos, etc. Certains réseaux sociaux proposent, par exemple, de collecter des données de géolocalisation afin de suivre l'évolution de la pandémie de Covid-19. Ainsi, il ne s'agit plus simplement de déclarer aux institutions nationales une pathologie avérée, mais de collecter par différents biais un nombre important de données très diverses auprès de patients mais aussi, et surtout, de personnes en bonne santé.

Bien que les objectifs portés par la santé publique soient louables, le flot des données collectées, notamment sur les déplacements des malades ou sur les contacts avec des personnes à risque, interroge sur notre capacité à maintenir un cadre protecteur pour la vie privée et à éviter que les données collectées ne soient ensuite réutilisées à des fins discriminatoires ou commerciales. De graves failles de sécurité sont régulièrement signalées, comme dans l'exemple donné par l'association Greenbone Networks sur les données d'IRM et de scanner¹. Or, les données peuvent intéresser de nombreux acteurs privés, comme les assureurs afin d'analyser le risque « santé » de chaque assuré, mais aussi les banques ou certains employeurs.

Par conséquent, le concours des individus à la santé publique à l'ère des données massives pose des questions particulières que le seul impératif de solidarité dans le partage des données ne saurait balayer. L'emploi des données massives a toutefois d'indéniables effets positifs sur les droits individuels, notamment en matière de droit à l'information, ainsi que le montre la création du « Tableau de bord de suivi de l'épidémie de Coronavirus en France », centralisant les données de Santé publique France, de la Direction générale de la santé et les données sur les transferts de patient. Un même outil peut servir des finalités collectives et de protection des personnes.

Conclusion

L'usage des données massives pour la santé publique remplace plus que jamais l'individu au cœur de la promotion et de la protection de la santé de la population. Cependant, cette mobilisation n'est pas sans risques, notamment concernant le dévoilement d'informations sur le comportement, l'état de santé et la vie privée des personnes. Le thème de ce dossier invite donc à réinterroger le cadre et les objectifs de la santé publique, ainsi qu'à tester la robustesse des équilibres établis à la lumière du recours croissant aux données massives. ●

1. <https://www.lemondeinformatique.fr/actualites/lire-plus-d-un-milliard-d-images-medicales-exposees-77688.html>



Enjeux pour la santé publique du traitement des données massives diverses

Margo Bernelin
Sonia Desmoulin

Chargées
de recherche CNRS,
UMR 6297 CNRS,
université de Nantes

« **L**a médecine de demain ne sera pas la même que celle d'aujourd'hui : prédictive, personnalisée, numérique, elle devra sans cesse s'adapter aux nouveaux enjeux et aux nouvelles technologies¹. » C'est par cette phrase que la Stratégie nationale de santé 2018-2022 indique vouloir faire de l'introduction du numérique en santé, et notamment des *données massives*², l'avenir des politiques publiques en santé et positionner la France dans un mouvement global. Ainsi, tant la médecine que la santé publique doivent bénéficier des avancées du numérique. Dans ce cadre, le développement d'outils algorithmiques exploitant des masses de données diverses au service de la santé des populations semble constituer, avec les nouvelles capacités d'archivage, une innovation phare de ce début de XXI^e siècle. Cela se combine avec la possibilité d'exploiter l'ensemble des bases de données publiques et privées constituées notamment par la collecte de données issues d'objets connectés. Les données exploitées peuvent être très variées : signaux biologiques, dépenses de santé, observance des traitements, facteurs environnementaux, habitudes de vie, etc. Aussi, le volume des données traitées et leur diversité sans précédent ouvrent la voie à de possibles transformations majeures en santé publique.

De nombreuses perspectives

Les pouvoirs publics attendent des études menées sur les données collectées l'élaboration de nouveaux plans de prévention en santé, des analyses inédites en matière de risques épidémiques, des traitements mieux ciblés pour les patients et la découverte de voies de recherche insoupçonnées en santé publique, par exemple, de nouveaux facteurs de risques associés à des pathologies pour développer des mesures préventives. Le contexte pandémique né de la propagation du virus SARS-Cov-2 (Covid 19) a montré l'intérêt porté par les pouvoirs publics à l'analyse de données, afin d'anticiper l'évolution de la situation. En France, Epidemap, outil de modélisation issu de trois modèles traitant respectivement des données sur la répartition des habitants (bâtiment par bâtiment), des simulations de déplacements et de

données d'épidémiologie a été conçu pour établir des scénarios de propagation du virus en fonction d'options de confinement à domicile ou de déconfinement. Il a participé à légitimer les mesures de gestion de crise³. Pour lutter contre la transmission virale, les autorités publiques ont également soutenu le développement d'applications mobiles ou de plateformes telles que StopCovid ou ContactCovid. Ces outils visent à collecter des données pour informer les personnes concernées d'un contact possible, durant les jours précédents, avec une personne dépistée ultérieurement, mais aussi pour informer les autorités publiques et les institutions sanitaires sur les parcours géographiques des personnes (dépistées ou non) et sur la détermination des cas contacts. Le relatif échec de l'application StopCovid (beaucoup moins utilisée en France qu'en Allemagne, par exemple) n'a pas remis en question la volonté du gouvernement français, puisqu'une nouvelle version de l'application mobile est prévue. Rebaptisée TousAntiCovid, elle inclut une dimension informative supplémentaire sur les risques sanitaires (symptômes de la maladie, statistiques...) et sur les lieux de dépistages, en plus de l'alerte sur les « cas contacts ».

Toutefois, ces évolutions s'accompagnent de nouvelles interrogations. En sus des enjeux de protection des droits et libertés individuels, en lien avec la protection de la vie privée et des données personnelles, des questions relatives aux frontières et aux acteurs de la santé publique sont ainsi posées. La délimitation des champs de la santé publique était déjà sujette à discussion, en ce qu'ils pouvaient être restreints à l'approche populationnelle et aux dimensions institutionnelle et organisationnelle, ou inclure plutôt, de manière large, tous les domaines couverts par le Code de la santé publique (y compris la médecine). L'approche par le croisement de données diverses en grand volume semble élargir encore les perspectives et donner une importance particulière au concept d'exposome récemment introduit dans le Code de la santé publique et voulant que « l'intégration sur la vie entière de l'ensemble des expositions qui peuvent influencer la santé humaine⁴ » constitue un objectif primordial des politiques de santé publique. Si toutes sortes de données, constamment collectées, peuvent être compilées et croisées pour concevoir de nouvelles politiques et de nouvelles mesures de prévention, la détermination de ce qui relève spécifiquement de la santé publique devient ardue. Tout relève alors possiblement de la santé publique.

1. Stratégie nationale de santé 2018-2022. Ministère des Solidarités et de la Santé, Paris : 2017, p. 63.

2. Il n'existe pas de terme légal ou consensuel en France pour la traduction de l'expression *big data*. Deux expressions sont couramment employées : mégadonnées et données massives. Toutes deux ont l'inconvénient de ne pas suffisamment insister sur la caractéristique de diversité des données traitées qui, en sus de la rapidité de traitement grâce aux nouveaux outils informatiques et techniques algorithmiques, constitue la spécificité de l'approche par le *big data*. Sur le concept, voir le chapitre initial de ce dossier.

3. <https://www.data.gouv.fr/fr/organizations/epidemap/>

4. Art. L. 1411-1 CSP

Une intervention grandissante du secteur privé

Quant à la dimension organisationnelle, elle est également bousculée par l'idée qu'il ne serait peut-être plus nécessaire de passer par des institutions et des réseaux créés par l'État ou par les autorités publiques pour traiter les alertes et transmettre les messages de prévention. La place du secteur privé dans les activités de collecte, de stockage et de traitement des données de santé (applications en ligne, plateformes nationales) interroge. Alors que par le passé la collecte et l'analyse des données à des fins de santé publique étaient essentiellement opérées par des services de l'État à des fins d'intérêt général, elles s'appuient aujourd'hui plus souvent sur le concours d'entreprises privées non spécialisées en santé. Les plateformes développées par des réseaux sociaux pour suivre la pandémie de Covid-19 en sont un exemple⁵. Désormais, les entreprises du numérique ou les assureurs s'impliquent au-delà de la simple collecte de données, par la transmission de messages à caractère sanitaire auprès de leurs abonnés-clients-assurés. Un tel constat n'est pas anodin. D'une part, une collecte si volumineuse de données personnelles par des sociétés privées enrichit encore un peu plus leur catalogue de données, faisant peser un risque accru sur la vie privée des citoyens dont les vies sont documentées. D'autre part, la qualité et la fiabilité des informations sanitaires offertes par ces entreprises peuvent être questionnées. En effet, comment s'assurer de la fiabilité du contenu d'une application qui partage des informations sur les moyens de lutte contre certaines pathologies ou virus ? Le passage de la donnée/mesure à l'information et à la donnée probante est ici crucial. Par exemple, les informations offertes par l'agence Santé publique France s'appuient sur le concours avéré de chercheurs dont les moyens d'investigation peuvent être retracés.

La place grandissante du secteur privé se fait également sentir dans les activités de conservation des données. Des entreprises privées telles que Google, Amazon ou Microsoft proposent leurs services. C'est dans ce contexte que le Health Data Hub, plateforme nationale des données de santé, appelée à jouer un rôle central pour la recherche en santé publique, a annoncé que la société étatsunienne Microsoft serait en charge de l'hébergement des données de santé, bien qu'elles soient souvent considérées comme un « trésor national ». Ce choix a été principalement justifié par les capacités importantes d'hébergement de cette société. Sous réserve que les conditions d'archivage respectent le droit européen en matière de protection des données personnelles, le droit français ne s'oppose pas à cette imbrication « public/privé », mais elle a pour conséquence de rendre l'État dépendant du secteur privé en matière de santé et notamment de santé publique. De nombreuses critiques ont été formulées à cet égard, pointant l'inopportunité de choisir une

entreprise étrangère qui dispose déjà d'une puissance financière et commerciale considérable.

Au final, l'introduction des données massives diverses en santé publique paraît renouveler profondément cette discipline centenaire. En effet, les acteurs de la santé publique se diversifient notamment au profit du secteur privé et d'entreprises dont les activités étaient jusqu'alors fort éloignées du champ de la santé. La discipline semble également se renouveler dans ses pratiques, les contributions à ce dossier montrant que les recherches en santé publique évoluent sous l'attraction de la disponibilité nouvelle de données nombreuses et variées. Les objectifs de prévention des risques et de promotion de la santé de la population semblent néanmoins perdurer. Il reste toutefois à vérifier si, dans l'avenir, l'évolution vers toujours plus de stratification (en sous-groupes de populations en fonction de caractéristiques sociales, environnementales, physiologiques ou génétiques notamment) et vers une médecine « de précision » (adressant des messages préventifs et des propositions thérapeutiques individualisés) ne remettra pas en cause l'idée même d'une approche populationnelle globale et de choix collectifs. ●

5. Covid-19 Interactive Map & Dashboard, plateforme proposée aux abonnés de Facebook. https://covid-survey.dataforgood.fb.com/survey_and_map_data.html