



La position des institutions publiques françaises dans la promotion et l'utilisation des données en santé publique

La promotion et l'utilisation des données de santé sont organisées pour un service public, par des institutions qui peuvent être en concurrence.

Thomas Lefèvre

Maître de conférences, praticien hospitalier, Iris-UMR 8156-997 CNRS Inserm EHESS université Sorbonne Paris Nord

Sabine Guez

Post-doctorante, Iris-UMR 8156-997 CNRS Inserm EHESS université Sorbonne Paris Nord

Cette tribune n'a pas pour vocation de retracer une sociohistoire précise du rôle des institutions publiques françaises dans la promotion et l'utilisation des données en santé publique, mais d'en donner quelques éléments de repères et de questionnements.

Porter l'attention du privé vers le public en matière de données de santé

S'agissant du *big data*, de l'intelligence artificielle, plus largement de l'irruption du numérique dans notre quotidien, l'attention a surtout été focalisée sur les acteurs privés, comme Google, Amazon, Facebook, Apple ou Microsoft (Gafam), en particulier pour des raisons de souveraineté nationale et de protections juridiques variables et non garanties partout dans le monde de la même façon. Ces acteurs ont proposé, proposent encore régulièrement, de se positionner comme intermédiaires voire comme substitués à ce qui peut être considéré comme des fonctions régaliennes, au minimum des missions de service public. On peut penser à la fonction *safety check* de Facebook, en cas de catastrophe naturelle ou d'événement imprévu de portée collective, menaçant la vie des personnes, ou aux applications smartphone proposées par Google et Apple aux gouvernements dans le cas du *case tracking* dans le contexte de l'épidémie de Covid-19. En comparaison, la position des institutions publiques dans la promotion et l'utilisation des données n'a fait l'objet que de peu d'attention. Or en France l'organisation et l'utilisation des données en santé tiennent essentiellement à des initiatives des institutions publiques,

aux niveaux gouvernemental et ministériel, depuis une quarantaine d'années.

Le pilotage médico-économique des établissements de santé basé sur la donnée

Une façon d'aborder le sujet est de partir de la création du PMSI (Programme de médicalisation du système d'information [des établissements de santé]) au début des années 1980, dont l'utilisation a été renforcée en 1996, puis en 2005. L'idée derrière le PMSI est la quantification et la standardisation d'un certain nombre d'informations ayant trait aux hospitalisations : les diagnostics, les durées de séjour. Conjointement, un travail de classification et de nomenclature se met en place, s'institutionnalise, notamment avec la création de l'ATIH (Agence technique de l'information sur l'hospitalisation) en 2000. Il s'agit d'une part d'identifier et de mettre à jour l'ensemble des GHM (les groupes homogènes de malades) à partir des données collectées dans les hôpitaux et, d'autre part, de faire coïncider une nomenclature médico-économique, sur laquelle la tarification à l'activité (la T2A) se base dès 2005. Des données concernant un grand nombre de malades et d'hospitalisations sont collectées; par des approches algorithmiques, on identifie des groupes de malades qui se « ressemblent » tant en termes médicaux (diagnostics...) que de coûts du séjour en hôpital. Ces groupes permettent de « segmenter » la population hospitalière comme on le fait en marketing pour identifier des sous-populations de clients. La technique n'est en rien novatrice et est importée des pratiques du privé. Cela sert alors à donner

une assiette au budget d'un établissement de santé. Au sein de ces établissements, les départements d'information médicale (DIM) ont, dans beaucoup de cas, la tâche désormais principale de gérer le PMSI local, et de justifier le budget de leur établissement. Au niveau national, les données des PMSI locaux sont fédérées, consolidées.

Les données au-delà du PMSI et du Sniiram : la fédération des données, leur gestion et leur accessibilité

Le PMSI est donc une source de données historique en France. Une autre source de données elles aussi médico-administratives, gérée par une institution publique, est le Sniiram (Système national d'information inter-régimes de l'Assurance maladie). Elle est gérée par la Caisse nationale d'assurance maladie et fait partie de la base du Système national de données de santé (SNDS), regroupant de fait avant tout les données du PMSI et du Sniiram.

Simultanément, d'autres acteurs publics se sont manifestés de façon croissante quant à l'utilisation des données de santé, au-delà des données médico-administratives. On a ainsi assisté à l'émergence des entrepôts de données de santé des hôpitaux, tendant à rassembler l'ensemble des données dérivées de tous les examens et observations effectués lors d'un séjour hospitalier. Le panorama des organismes publics pouvant participer à l'écosystème de la production et de l'utilisation des données de santé en France se complète par les universités (facultés de médecine), les unités Inserm et les DIM déjà cités, ainsi que plusieurs organismes ministériels : Drees, Acof, CNSA...

Nous avons donc ici le panorama classique d'une multitude de sources de données à réconcilier, néanmoins doublé d'une multitude d'acteurs, tous publics, aux intérêts concurrents. Cette concurrence est fréquemment réduite à l'argument d'une supposée propriété des données : l'hôpital propriétaire des données captées dans son établissement, l'Inserm propriétaire des données générées dans un cadre de recherche, la faculté propriétaire des données recueillies par son personnel hospitalo-universitaire – lui-même fréquemment affilié à une unité Inserm et exerçant en établissement hospitalier. Une solution organisationnelle et technique proposée pour une forme de convergence de ces sources de données, solution dont la teneur réelle reste néanmoins à préciser à l'épreuve de la pratique et du temps, s'incarne

depuis le 30 novembre 2019 dans la plateforme de données de santé, le Health Data Hub, constitué en groupement d'intérêt public (GIP). Conçu comme un guichet unique facilitant l'accès et l'utilisation des données pour la santé pour différents acteurs, le Health Data Hub cristallise cependant plusieurs critiques d'autant plus aisément qu'il apparaît comme une personne identifiable plutôt que la multitude des acteurs. Ces critiques sont liées aux aspects de la nouvelle gestion publique (*new public management*). Un exemple est le recours à l'externalisation pour des tâches ou des missions importantes, comme l'hébergement des données attribué à Microsoft. Un autre exemple est celui de la conformité avec le référentiel de sécurité Système national des données de santé, qui par ricochet réglementaire, est applicable ou demandé désormais à des acteurs qui n'en ont pas les moyens, et les privés de données auxquelles ils avaient jusque-là accès, nécessaires à la réalisation de leurs missions.

Parallèlement, ce problème de convergence des sources de données, et de « propriété », aurait pu être au moins techniquement résolu par la création et l'utilisation du dossier médical partagé (DMP), annoncé par la loi du 13 août 2004 relative à l'assurance maladie : il s'agirait simplement d'un carnet de santé électronique, étendu à toute la vie et non plus aux seules premières années. Le projet ne prend pas, certainement pour des raisons de visions divergentes et de concurrences interinstitutionnelles. L'Agence des systèmes d'information partagée en santé (Asip, devenue Agence du numérique en santé en décembre 2019) est créée en 2009, dont une des missions est spécifiquement de relancer le projet du DMP. La Caisse nationale d'assurance maladie reprend le projet à son compte en 2015. Finalement, la dernière itération de ce concept devrait se réaliser dans l'Espace numérique de santé, défini par la loi du 24 juillet 2019, à partir du 1^{er} janvier 2022.

La concurrence n'est pas limitée à une concurrence inter-institutions. Elle existe également au sein même des institutions. Un de ses aspects est une confrontation culturelle et liée à un changement dans les formations des élites françaises, où les écoles de management et de commerce sont depuis plusieurs années plus prisées que les écoles d'ingénieurs. Traditionnellement, au sein de l'institution publique, les personnes en charge des études et des statistiques – la statistique d'État, et ses extensions, *via* le

captage et l'automatisation, la fédération des données produites et numérisées – sont majoritairement issues de l'École nationale de la statistique et de l'administration économique (Ensaie). On observe, depuis quelques années, le recrutement de personnes plus souvent issues d'écoles privées de management et de commerce (HEC, Essec, Edhec...), voire la constitution interne de départements dédiés et distincts des départements des études et statistiques autour de ces recrutements. Ces nouveaux départements ou directions sont en général en charge de l'innovation ou de la transformation « digitale », et se retrouvent face aux départements historiques des systèmes d'information et des études statistiques. Il existe enfin probablement une concurrence, ou un frein à la convergence, sur le versant sanitaire : les experts en santé publique sont, en France, issus d'au moins trois horizons différents, parfois cumulatifs, mais dont les logiques institutionnelles sous-jacentes sont plutôt concurrentielles là aussi : les médecins de santé publique, les chercheurs en épidémiologie, et les diplômés de l'École des hautes études en santé publique (EHESP).

Globalement, on assiste donc à des efforts publics de concentration des données personnelles, captées à partir de différentes institutions publiques mais aussi depuis le secteur du privé *via* le secteur du travail, définie et imposée par la voie législative. Les efforts d'ouverture de ces données semblent réels, mais loin de reposer ou de favoriser une déconcentration, comme cela aurait pu, ou pourra être, en se basant sur la participation et l'accord de chaque citoyen, *via* un DMP ou tout autre dispositif semblable et acceptable par la société. À ce jour, tout sondage effectué pour mesurer l'acceptation des Français d'un tel dispositif a toujours pointé vers une adhésion massive au principe : quiconque passe de son généraliste aux urgences, puis à l'hospitalisation, puis d'un hôpital à un autre, d'un spécialiste à un autre, sait combien il serait utile de disposer d'un dossier unique.

La donnée pour décider, le fait scientifique et la santé publique

Le protectionnisme de l'État dès qu'il est question de données de santé, envers les autres nations, mais aussi envers une partie du secteur privé, parfois entre institutions publiques et enfin même envers les citoyens, est souvent justifié par le caractère sensible de ces données – on met en avant le caractère individuel, mais d'un point de vue national, elles sont aussi sensibles en cela qu'elles



renseignent sur la population –, le secret professionnel et la valeur des données pour l'industrie et la recherche. Il existe également la dimension que la récente épidémie de Covid-19 a mise en exergue : le besoin de disposer de connaissances scientifiques d'une part, mais également de données pour décider, pour gouverner. Plusieurs commentateurs auront relevé la porosité des champs lexicaux quant à la gestion de l'épidémie en cours, relevant tous des prérogatives régaliennes, à commencer par le langage martial : nous sommes en guerre contre le virus ; un état d'urgence sanitaire peut être déclaré, délivrant des pouvoirs spéciaux aux représentants de l'État. Dans le cadre de l'épidémie, les deux sources évoquées, supports de la décision éclairée, ont été prises en défaut, au sens de leur caractère utile, adapté et précis à un moment donné de l'épidémie : les connaissances scientifiques d'une part, la production de données utiles d'autre part. Ce qui est aux racines historiques de la statistique, dénombrer les morts et les naissances, a été défaillant, et les logiques concurrentielles entre institutions, un problème majeur : le CépiDc, unité Inserm unique dévolue au recueil et à l'étude des causes médicales de décès, en position d'asphyxie depuis plusieurs mois sinon années, a sans doute été un des seuls acteurs à avoir été privé de la manne financière ouverte à destination de la recherche et de l'industrie face à la Covid-19. L'écosystème en place n'a pas pu être mobilisé, renforcé, pour fonctionner et produire les données nécessaires. D'autres institutions publiques ont occupé le terrain de façon peu convaincante ou insuffisante pour documenter efficacement et rigoureusement les connaissances sur la mortalité, comme

l'Insee, l'Ined ou Santé publique France. Et pour cause : aucune de ces institutions ne dispose des causes médicales de décès.

Symétriquement, le recueil de données de santé peut être, au-delà de son caractère informatif sur l'état de santé d'une population, le biais par lequel introduire une forme de gouvernement des individus, voire d'une police sanitaire, et la possibilité, parfois inédite, de rapprocher des institutions publiques, et même privées, dans cette optique de contrôle et de régulation des flux de population. Le cas de la santé publique de précision dans le cadre de la gestion de l'épidémie de Covid-19 est éclairant. Prenons l'exemple des applications téléphoniques de *tracking* (traçage des cas). La majorité des pays à hauts revenus s'est dotée d'une application de *tracking*. Aucune n'a été en position d'être sanitaires utiles – la couverture nécessaire, c'est-à-dire la proportion de personnes installant et utilisant réellement l'application, est estimée à plus de 60 % pour être utile. Il semble que, dans le meilleur des cas, la proportion des personnes ayant téléchargé une telle application a été de 30 à 40 %. Le cas français (StopCovid) est une illustration classique de nos institutions : un outil propriétaire, développé par les acteurs habituels publics et industriels privilégiés, sans aucune évaluation extérieure. Dans d'autres pays, ce qu'il est important de souligner est alors l'articulation entre institutions derrière le déploiement de l'application, et son insertion dans une politique sanitaire plus large. Ainsi, en Chine, l'application délivrait ni plus ni moins des autorisations – un passeport – de circulation physique selon son niveau individuel de risque. Les algorithmes et les

caractéristiques pris en compte pour élaborer le niveau de risque sont inconnus. En Corée du Sud, le système repose sur la circulation d'informations entre les acteurs du système de santé, la police ou encore les aéroports et le gouvernement.

Dans l'épidémie de Covid-19, jusqu'à présent, l'*evidence-based medicine* (EBM, médecine basée sur les preuves) a été mise à l'épreuve face à d'autres conceptions, dont l'école des pragmatiques, et semble ne pas avoir été en position de contribuer significativement à l'adaptation des politiques de santé. Le passage discret à une santé publique de précision, soutenant une *evidence-based policy* (EBP, politique basée sur les preuves), est, d'un point de vue sanitaire, encore moins convaincant.

Les problématiques autour des données de santé révèlent plus que jamais différentes tensions connues, que l'on peine à dénouer : les tensions au cœur de la santé publique, devant concilier intérêt individuel et populationnel ; les tensions inhérentes au *new public management*, et la porosité sélective, inconfortable, entre public et privé dans la répartition et la réalisation de missions de service public ; les tensions, enfin, entre acteurs participant à la production et à l'utilisation des données. En effet, la promotion et l'utilisation des données sont avant tout organisées par et pour un service public, dont les acteurs sont en concurrence. Parmi eux, les producteurs de données ne veulent pas être dépossédés d'une partie de leur travail, qu'ils n'ont pas les moyens de valoriser souvent seuls, la donnée étant présentée comme une valeur en soi, source d'avantage compétitif (entre hôpitaux par exemple). ●