

Introduction

Données et modèles dans le « gouvernement de l'épidémie »

Par Maurice RONAI
Ancien membre de la CNIL

et Aymeril HOANG
Ancien membre du Conseil scientifique Covid-19
et expert en numérique

Première grande pandémie du XXI^e siècle, la crise du Covid-19 a révélé l'impréparation des systèmes de santé et les limites de la « transformation numérique » des administrations. Cette « crise des données », inégale selon les pays, a donné lieu à une mobilisation sans équivalent : modélisations, foisonnement d'enquêtes, construction dans l'urgence de systèmes d'information, exploitation de nouvelles sources de données, accélération de la production statistique, enquêtes répétées de prévalence sur des échantillons de la population, déploiement de dispositifs pour la recherche et le suivi des contacts.

UNE CRISE MONDIALE DES DONNÉES

« Peu de pays disposaient des données en temps réel dont ils avaient besoin lorsque la pandémie a frappé, qui a rendu d'autant plus difficile le suivi de la situation de santé publique et le travail d'information de la population » observe l'OCDE. « La collecte et l'utilisation en temps réel des données médicales personnelles ont aussi mis au jour les faiblesses de longue date de la gouvernance des données – s'agissant notamment de la façon de partager les données nécessaires pour sauver des vies, tout en veillant aussi à protéger la vie privée et les libertés civiles. »

L'OCDE dresse ainsi le diagnostic d'une « crise des données », qui trouve sa source dans « les lacunes des infrastructures et des politiques en matière de données critiques » et dans « des approches fragmentées de la collecte, du partage et de l'utilisation des données non seulement d'un pays à l'autre, mais aussi à l'intérieur des pays, ce qui a accru la méfiance et réduit le respect des mesures de sécurité »¹.

Aux États-Unis, la gestion de la crise a buté sur « une infrastructure de données de santé publique fracturée et obsolète. La réponse de la nation à la crise du coronavirus a été affaiblie par une infrastructure de données de santé publique fracturée et obsolète »².

¹ Highlights from the OECD Webinar (2020) “The ties that bind: Government openness as key driver of trust”, September.

² “Preparing for and preventing the next public health emergency. Lessons learned from the Coronavirus crisis”, Final Report, December 2022.

L'Allemagne s'est débattue, tout au long de la crise, et encore aujourd'hui, avec un empilement de logiciels et de systèmes d'information, à l'origine de problèmes aigus d'interfaçage entre outils de gestion locaux, applications de *reporting* et systèmes de surveillance fédéraux. En juin 2022, le conseil des experts, chargé d'évaluer les mesures mises en œuvre dans la lutte contre la pandémie, conclut « qu'il n'existe pas de concept national [...] pour coordonner la recherche dans le domaine de l'épidémiologie et de la santé publique et pour prendre les décisions qui s'imposent, sur la base de meilleures données et d'analyses basées sur celles-ci »³.

En France, selon la mission indépendante nationale sur l'évaluation de la gestion de la crise⁴ (Mission Pittet), « la gestion de la crise a pâti d'un déficit dans l'architecture des systèmes d'information qui laissait de nombreux trous au regard des besoins de pilotage de la gestion de crise, comme la connaissance des cas de Covid ou celle en temps réel des capacités disponibles en lits de réanimation ».

UNE MOBILISATION INFORMATIONNELLE SANS PRÉCÉDENT

En France, la surveillance était initialement centrée sur les seuls cas graves et les morts, sur les hôpitaux et les urgences. Il a fallu adapter des systèmes d'information déjà existants, les faire converger et étendre la couverture de certains d'entre eux, mais aussi en créer de nouveaux, de toutes pièces, et en urgence. Brique par brique, un système de surveillance est ainsi mis en place en quelques mois qui va permettre de rendre visibles les tensions hospitalières, la mortalité, la circulation virale, l'émergence et la propagation des variants. « Aujourd'hui, on a quelque chose de quasi exhaustif. Personne en France n'aurait pensé qu'on puisse créer un système national comme celui-là en si peu de temps », reconnaissait un responsable de Santé publique France en septembre 2020⁵.

Un fait marquant de cette crise est la mobilisation tous azimuts de données non conventionnelles comme l'analyse des eaux usées ou l'exploitation des données de géolocalisation.

Cette mobilisation informationnelle s'est largement appuyée sur des dispositifs d'enquête faisant appel au public. Une floraison de questionnaires en ligne et d'enquêtes vit ainsi le jour, dès mars 2020, pour identifier les symptômes, cartographier la propagation et comprendre les mécanismes de transmission. À l'origine de ces dispositifs « d'épidémiologie participative », plus ou moins exigeants en termes d'engagement des répondants, figuraient des agences de santé publique, des hôpitaux, des universités, des ONG, des entreprises et des startups de santé numérique.

Au Royaume-Uni, le système de surveillance épidémiologique, pris en défaut dans un premier temps, va faire preuve de capacités inédites de réactivité et d'innovation. La création du National Core Studies va permettre de structurer six programmes d'études, et de financer des enquêtes de prévalence et des opérations de séquençage à un rythme et une échelle sans pareils au niveau mondial⁶.

³ "Evaluation der rechtsgrundlagen und maßnahmen der pandemienpolitik. Bericht des sachverständigenausschusses", juin 2022.

⁴ Rapport de la Mission indépendante nationale sur l'évaluation de la gestion de la crise de Covid-19 et sur l'anticipation des risques pandémiques.

⁵ « Urgence, cafouillages, "mille-feuille" d'indicateurs... Dans les coulisses des données très stratégiques sur le Covid-19 », FranceTVInfo, 10 septembre

⁶ En France, en réponse à l'exigence d'une meilleure coordination de la recherche dans le champ des maladies infectieuses et émergentes, est créée en 2021 l'ANRS-Maladies infectieuses émergente, issue de la fusion de l'ANRS avec REACTing afin de couvrir tous les domaines de la recherche : fondamentale, clinique, en santé publique et en sciences de l'homme et de la société.

L'ENJEU DE CONDUIRE DES ENQUÊTES DE PRÉVALENCE FRÉQUENTES SUR DES ÉCHANTILLONS DE LA POPULATION

Le suivi exhaustif de l'ensemble des patients testés a permis de construire et d'affiner toute une série d'indicateurs quotidiens, comme les taux de positivité, de dépistage et d'incidence de la maladie. Les personnes asymptomatiques étant moins susceptibles d'être testées, les indicateurs basés sur le dépistage de masse pouvaient, toutefois, sous-estimer la prévalence. Plusieurs contributeurs à ce numéro soulignent l'intérêt de procéder à un dépistage fréquent d'échantillons aléatoires de la population générale. Ils font référence à deux enquêtes britanniques, l'ONS Covid Infection Survey, pilotée par l'Office for National Statistics, et l'étude React, pilotée par l'Imperial College. Les conseils scientifiques et les autorités de santé du monde entier se sont largement appuyés sur ces deux enquêtes pour percer le brouillard épidémiologique.

LA RAISON MODÉLISATRICE FACE À LA DÉCISION POLITIQUE

La pandémie de Covid-19 a propulsé sur le devant de la scène politique la modélisation mathématique des maladies infectieuses. Elle permet, tout au long de la crise, d'interpréter les données, de donner un aperçu de ce à quoi pourraient ressembler les futures tendances épidémiques, et d'aider les décideurs à comprendre « l'espace de décision » dans lequel ils se trouvaient. La « raison modélisatrice »⁷ a ainsi occupé une place inédite dans la décision politique comme dans la conversation publique.

Alors que les gouvernements au Danemark, aux Pays-Bas ou en Australie disposaient de leur propre expertise en modélisation, au Royaume-Uni, les modélisations étaient menées à l'extérieur, en s'appuyant sur la communauté universitaire. Aux États-Unis, les épidémiologistes et les modélisateurs ont regretté l'absence d'une instance nationale pour confronter les prévisions et les modèles, et d'une interface avec les décideurs politiques et les autorités de santé pour discuter et évaluer les stratégies. Une attente satisfaite, en 2021, avec la création du Centre de prévision et d'analyse des épidémies que le Président Biden décrit comme « l'équivalent d'un service météorologique national pour les épidémies de maladies infectieuses ».

En France, la Mission Pittet appelait en mars 2021 l'attention sur « des capacités de modélisation mobilisables limitées en raison de la dissémination des compétences entre différentes équipes, toutes de dimension réduite »⁸. Elle recommandait de « renforcer les capacités françaises en matière de modélisation, par une politique volontariste d'orientation de la recherche »⁹. La procédure des appels d'offres compétitifs gérés par l'ANR, coûteux en temps de rédaction et d'évaluation, révèle ses limites. Des financements de base seront nécessaires pour permettre aux équipes d'épidémiologie de « dégager du

⁷ BOULLIER H., CORNILLEAU E., JOUZEL J. & JUVEN P. (2021), « Raison modélisatrice, incertitudes et pièges de l'abstraction », *Terrains & travaux*.

⁸ Mission indépendante nationale sur l'évaluation de la gestion de la crise Covid-19 et sur l'anticipation des risques pandémiques, mars 2021.

⁹ En France, la toute nouvelle ANRS-Maladies infectieuses et émergentes a depuis mis en place une action coordonnée « Modélisation des Maladies Infectieuses ».

temps pour des projets ambitieux, n'ayant pas peur de l'originalité et encourageant une vision à long terme »¹⁰.

UNE DIVERSITÉ DES APPROCHES POUR LA RECHERCHE ET LE SUIVI DES CONTACTS

La recherche des contacts est la pierre angulaire de la réponse aux épidémies. Cette « épidémiologie de terrain » a fonctionné, dans le passé, face à de nombreuses maladies infectieuses : qu'il s'agisse de freiner l'entrée du virus sur le territoire (phase 1), de limiter sa propagation (phase 2), ou d'atténuer les effets de l'épidémie (phase 3).

En 2020, les pays asiatiques sont parvenus à contrôler les transmissions communautaires par l'application de protocoles intrusifs en termes de traçage et de méthodes d'isolement souvent brutales.

En Europe, les capacités de détection et d'investigation des cas ont très vite été débordées : c'est dans un contexte de forte circulation du virus (phase 3) que les gouvernements ont déployé des dispositifs de recherche des contacts à grande échelle. L'ampleur des opérations de recherche des contacts était, là aussi, sans précédent : au Royaume-Uni, les enquêteurs sanitaires ont tracé 19 millions de cas et 37 millions de contacts. En France, ceux de l'Assurance maladie ont pris en charge 30 millions de cas positifs et 23 millions de cas contacts.

Un outillage numérique pour alléger et amplifier la recherche de contacts

La pandémie a renouvelé les modalités du traçage de contacts, avec la dématérialisation de certaines opérations. En France comme en Grande-Bretagne, lorsque les niveaux de contamination ne permettaient plus la réalisation d'un appel téléphonique auprès de tous les cas positifs, ceux-ci étaient invités par SMS à préparer la liste de leurs cas contacts et à les alerter eux-mêmes. Ces plateformes numériques ont constitué une « innovation importante en allégeant le travail des centres d'appels et en leur permettant de se concentrer sur les situations complexes ou les personnes qui n'avaient pas les équipements ou les compétences numériques », observe un retour d'expérience britannique : il estime qu'à l'avenir « l'auto-traçage numérique » (*self-tracing*) devrait être la norme : « les plans de préparation devront inclure le besoin de plateformes numériques à grande échelle »¹¹.

Traçage prospectif ou rétrospectif ?

Dans les pays occidentaux, on a essentiellement mis en œuvre une méthode de traçage dite « prospective » (*forward*) ou « amont » : on recherche et on alerte les personnes qui se sont trouvées à proximité d'un patient déclaré sur toute sa période de contagiosité pour qu'elles se testent ou s'isolent. En Europe, des voix se firent entendre, en 2020 et en 2021 pour mettre en œuvre le traçage dit « rétrospectif » (*backward*) ou « aval », déjà pratiqué au Japon : on remonte la chaîne de contamination jusqu'à la personne diagnostiquée afin de repérer l'apparition de *clusters*, identifier des événements supercontaminants

¹⁰ Samuel Alizon, « La recherche sur le Covid-19 est plombée par les appels à petits projets », *Le Monde*, 3 février 2023.

¹¹ Independent report, "Technical report on the Covid-19 pandemic in the UK", 2022.

ou repérer des individus superpropagateurs (“*superspreaders*”). Cette méthode avait été théorisée par le virologue japonais Hitoshi Oshitani, qui recommandait de se concentrer « sur la forêt et non pas sur chacun des arbres ». L’Assurance maladie a expérimenté le *retrotracing*, mais renoncé à le généraliser face à la remontée du nombre de cas fin juillet 2021.

Des dispositifs à maintenir et actionnables en cas de nouvelle épidémie

Tirant les enseignements des opérations de recherche des contacts, les experts de 39 pays européens¹² soulignent la diversité des stratégies de recherche des contacts, selon qu’ils pratiquaient ou non la recherche « à rebours », selon l’articulation ou non du traçage avec les stratégies de dépistage, selon le type de tests, selon les modes de communication avec les cas et les cas contacts (par téléphone, SMS, courriers ou courriels automatiques). Se projetant dans l’avenir, ces experts estiment que « les systèmes et compétences établis en matière de recherche des contacts doivent être maintenus et prêts à être réactivés rapidement si la situation épidémiologique l’exigeait ». Et qu’il faudra « mettre l’accent sur la recherche des contacts au sein des groupes vulnérables et dans les milieux à haut risque au cours de la période à venir ».

UN BILAN INCERTAIN POUR L’AUTOMATISATION DE LA DÉTECTION ET NOTIFICATION DES CONTACTS

Cette crise a été un laboratoire pour la mise en œuvre d’une nouvelle génération de solutions numériques dans la lutte contre les épidémies. La Corée, Taiwan, Hong Kong et Singapour firent en 2020 un usage intensif du numérique pour enrayer l’épidémie ; *contact-tracing* intrusif et obligatoire, exploitation des relevés bancaires et des historiques de géolocalisation (par Bluetooth, bornage téléphonique ou par GPS), surveillance à distance du respect des mesures d’isolement.

Si l’Europe écarta le recours aux solutions asiatiques, l’idée s’imposa, assez vite, en Europe et dans quelques pays, qu’il était possible d’automatiser, *via* le Bluetooth des téléphones mobiles, les opérations de traçage. Les choix fonctionnels et techniques de ces applications donnèrent lieu en Europe à d’après controverses (modèle décentralisé ou centralisé, recours ou pas à l’interface de programmation proposée par Google et Apple) qui débouchèrent sur des protocoles différents.

82 applications de ce type avaient été déployées en 2020 dans 31 pays. Selon AlgorithmWatch, « les résultats sont contradictoires. Les conclusions sur leur efficacité divergent considérablement d’un pays à l’autre et, parfois, d’une étude à l’autre. Les méthodes appliquées diffèrent tellement qu’il est, en fait, difficile – voire impossible – de comparer les résultats et de fournir une évaluation cohérente et complète de l’impact et de l’efficacité de ces applications »¹³.

¹² Bureau régional de l’OMS pour l’Europe & Centre européen de prévention et de contrôle des maladies (ECDC) : “Covid-19 contact-tracing: Country experiences and way forward”, 2022.

¹³ CHIUSI F. (2021), “Digital contact tracing apps: Do they actually work? A review of early evidence”, Algorithmwatch.

« Il est temps d'évaluer les applications de traçage des contacts Covid-19 », concluaient Valeria Colizza, Christophe Fraser et d'autres chercheurs influents en février 2021. Seule « une évaluation rigoureuse de leur efficacité permettra de mettre en balance les avantages pour la santé publique et les effets indésirables pour les individus et la société [...] Ce travail reste à faire »¹⁴.

Comment suivre, modéliser, anticiper la propagation du virus sans surveiller les personnes, sans collecter, exploiter et réutiliser des données personnelles ? Plusieurs retours d'expérience mentionnent les réutilisations de données que la France et l'Europe n'ont pas mises en œuvre, en raison du haut niveau de protection dont bénéficient les données de santé. On ne saurait se satisfaire de cette situation pour l'avenir. Quels sont les traitements de données qui doivent être établis de façon pérenne, et ceux qui doivent être créés et actionnés uniquement de façon transitoire au plus fort d'une épidémie ? Dans quels cas faut-il préférer des enquêtes auprès d'un échantillon plutôt qu'exploiter les données de populations entières ?

Faut-il, en cas d'urgence sanitaire, prévoir une extension des pouvoirs de réquisition de la puissance publique aux données détenues par des opérateurs privés¹⁵ ? Convient-il, comme le recommande un rapport sénatorial, d'imaginer la mise en place d'un "Crisis Data Hub"¹⁶ ?

Ces débats sont devant nous, si nous voulons faciliter la résolution de crises nouvelles tout en prévenant une bascule future vers des solutions franchement attentatoires aux libertés individuelles, ou qui, inventées ailleurs, ne protégeront pas les données, de santé ou autres, des Européens.

Ce dossier poursuit une réflexion engagée dans le numéro n°14 des *Annales des Mines - Enjeux numériques* (« Réponses numériques à la crise sanitaire », juin 2021) et dans le n°108 des *Annales des Mines - Responsabilité & Environnement* (« Premiers enseignements de la crise sanitaire », octobre 2022).

¹⁴ COLIZZA V., GRILL E. & MIKOLAJCZYK R. (2021), "Time to evaluate Covid-19 contact-tracing apps", *Nature Medicine*.

¹⁵ La loi du 23 mars 2020 instituant l'état d'urgence sanitaire prévoyait que le Premier ministre pouvait, par décret « ordonner la réquisition de tous biens et services nécessaires à la lutte contre la catastrophe sanitaire ». Cette disposition n'a jamais été actionnée pour accéder à des données privées pendant la crise sanitaire. Le projet de règlement européen sur les données ("Data Act"), encore en discussion, encadre strictement la possibilité pour les organismes du secteur public d'accéder et d'utiliser les données détenues par le secteur privé « afin de prévenir ou de répondre à une urgence publique ».

¹⁶ « Une plateforme sécurisée de collecte et d'échange de données dont l'unique fonction est de répondre aux situations de crise (sanitaire ou autre), lorsque des croisements de données massifs et dérogoires deviennent indispensables », Délégation sénatoriale à la prospective : Rapport d'information sur les crises sanitaires et outils numériques : répondre avec efficacité pour retrouver nos libertés, 2021.