
La pandémie de Covid-19 : un moment charnière pour l'ouverture des données en France



Samuel GOËTA¹

Co-fondateur de Dataactivist ; Maître de conférences associé en sociologie, Sciences Po Aix (Mesopolhis UMR 704)

TITLE

The Covid-19 pandemic: a turning point for open data in France

RÉSUMÉ

Cet article revient sur l'ouverture des données liées à la pandémie de Covid-19 en France pour montrer en quoi elle peut être considérée comme un moment charnière pour le mouvement récent de l'*open data*. S'appuyant principalement sur des sources ouvertes et des observations personnelles, il revient d'abord sur l'émergence des premiers outils internationaux de suivi de la pandémie de Covid-19, pour ensuite étudier la mise en place des outils de surveillance de la pandémie en France. Il montre aussi comment les premières ouvertures ont été accompagnées par des initiatives citoyennes de collecte et de visualisation des données. Ces initiatives citoyennes ont abouti à l'émergence de médiateurs de données qui ont joué un rôle important dans la diffusion et la compréhension des données sur la pandémie.

Mots-clés : Covid-19, pandémie, médiation, open data, covid tracker, tableaux de bord.

ABSTRACT

This article looks back at the release of data related to the Covid-19 pandemic in France to explain how it can be considered as a turning point for the recent open data movement. Relying mainly on open sources through a documentary research and personal observations, it first looks back at the emergence of the first international tools for monitoring the Covid-19 pandemic, and then studies the implementation of tools for monitoring the pandemic in France. It also shows how the first openings were accompanied by citizen initiatives of data collection and visualization that led to the emergence of data mediators who played an essential role in the dissemination and understanding of data on the pandemic.

Keywords: Covid-19, pandemic, mediation, open data, covid tracker, dashboard.

1. samuel@dataactivist.coop

« *On peut débattre de tout, sauf des chiffres.* » C'est ainsi que se concluait une des campagnes de publicité du gouvernement en faveur de la vaccination mettant en avant que 8 personnes sur 10 hospitalisées à cause de la Covid-19 n'étaient pas vaccinées². Ce message ne fait pas qu'user des chiffres comme argument d'autorité, il témoigne de la place centrale qu'ont occupée les données dans les débats relatifs à la pandémie de Covid-19. Selon l'évolution de ces chiffres, nous pouvions envisager de revoir des proches, de nous rendre sur notre lieu de travail, de partir en voyage, d'aller dîner au restaurant ou de sortir de chez soi passée une certaine heure. Cette situation inédite a été résumée par Emmanuel Didier³ dans AOC le 16 avril 2020 : « *Toutes les mesures politiques, tous nos discours, toutes nos réflexions sont orientés et conformés à un degré rarement atteint, par des outils quantitatifs.* » Ce constat s'est étendu bien au-delà de la première vague et de son confinement drastique. Très vite après l'arrivée des premiers cas en France, l'ouverture des données permettant de suivre l'évolution de l'épidémie a constitué une demande essentielle pour une multitude d'acteurs.

Nous nous intéressons ici aux indicateurs locaux et nationaux qui ont servi à suivre l'évolution de la pandémie et qui ont été scrutés quotidiennement par une grande part de la population (contaminations, entrées à l'hôpital, passages en réanimation, décès, tests, injections de vaccins, état des stocks, part des variants, en passant par les taux d'incidence et de positivité, le niveau du R0...) De vastes systèmes d'information ont permis de collecter toutes ces données sur les individus et la circulation du virus. Conformément à la législation, les données mises à disposition du public ont été expurgées de données à caractère personnel, agrégées ou anonymisées pour éviter toute réidentification d'individus. Depuis 2016, la loi pour une République numérique prévoit que les bases de données et les données qui présentent un intérêt économique, social, sanitaire ou environnemental doivent être diffusées en ligne, mises à jour et « dans un standard ouvert, aisément réutilisable et exploitable par un système de traitement automatisé. » Cette exigence de lisibilité des données par les machines a été essentielle pour de nombreux acteurs qui ne souhaitaient pas seulement accéder aux chiffres. Obtenir des données exploitables par les machines avec un fort niveau de détail a permis de refaire certains calculs, de proposer de nouvelles analyses, de nouvelles formes de visualisations, de renforcer ou de contester une décision.

Cet article est ainsi centré sur l'ouverture des données liées à la pandémie de Covid-19 en France et vise à documenter ce moment charnière pour le mouvement récent de l'*open data*. Pour décrire les frictions dans l'ouverture des données, je m'intéresserai aussi aux infrastructures qui ont été mises en place pour faire remonter les données collectées localement (Ronai, 2021). L'article s'appuie principalement sur des sources ouvertes à travers une recherche documentaire d'articles de presse et de documents officiels. Il s'appuie aussi sur des observations issues de mes activités associatives d'administrateur bénévole du portail d'accès aux documents administratifs « madada.fr » et professionnelles ayant participé à l'organisation et l'animation du Hackathon Covid, un événement rassemblant producteurs et réutilisateurs de données en avril 2021.

L'article décrit d'abord les premiers outils internationaux de suivi de la pandémie de Covid-19 qui ont posé une exigence internationale d'ouverture et d'actualisation des données de veille sanitaire. Il nous conduit ensuite à étudier la mise en place des outils de surveillance de la pandémie en France et à comprendre comment les premières ouvertures ont été accompagnées par des initiatives citoyennes de collecte et de visualisation des données. Nous verrons ensuite que l'utilisation massive des données sur la pandémie a conduit le gouvernement à relancer la politique d'*open data* en France qui semblait s'essouffler. Les médiateurs de données ont

2. Gouvernement (2021), « #COVID19 | Aujourd'hui en France, 8 personnes sur 10 hospitalisées à cause de la Covid-19 ne sont pas vaccinées. On peut débattre de tout, sauf des chiffres », *Twitter*. Adresse : <https://twitter.com/gouvernementfr/status/1429145022427893760>

3. Didier E. (2020), « Politique du nombre de morts », *AOC media - Analyse Opinion Critique*. Adresse : <https://aoc.media/opinion/2020/04/15/politique-du-nombre-de-morts/>

joué un rôle essentiel dans l'ouverture de nouvelles données sur la pandémie, en particulier CovidTracker, un site web lancé par le *data scientist* Guillaume Rozier, sur lequel je me concentrerai ici. L'article se conclut en évoquant le Hackathon Covid, un événement qui s'est tenu en avril 2021 à l'initiative de la Direction interministérielle de la transformation publique (DITP) afin de répondre à la pandémie en renforçant les liens entre les acteurs de la société civile et les administrations.

1. Une pandémie instantanément mise en données

Il n'est pas excessif d'affirmer que la pandémie a débuté par une ouverture de données. Alors que le 31 décembre 2019 les autorités de la ville de Wuhan en Chine annonçaient publiquement les signes d'une épidémie de pneumonie, le département des zoonoses du Centre chinois de contrôle et de prévention des maladies soumettait le 5 janvier 2020 le génome du virus identifié dans le marché de fruits de mer de Wuhan⁴. Le 11 janvier, ces données étaient publiées sur Genbank⁵, une des plus anciennes et plus grandes bases de données en génétique au monde (Strasser, 2011). La publication du génome a donné le départ de la course internationale aux vaccins, plusieurs équipes déjà préparées à l'éventualité de l'émergence d'un nouveau virus se constituant dès le 11 janvier. Dans un documentaire⁶ de la BBC, la professeure Teresa Lambe de l'université d'Oxford a raconté qu'elle a reçu une notification de la publication du génome dans la nuit du 10 au 11 janvier et s'est lancée immédiatement dans la conception d'un nouveau vaccin : « *Au petit matin, toujours en pyjama dans ma chambre, je me suis collée devant mon ordinateur pour développer un vaccin.* » Le 16 janvier, sans avoir eu accès au virus, une équipe de l'hôpital universitaire de la Charité de Berlin menée par le professeur Christian Drosten publiait le protocole du premier test RT-PCR⁷.

Les données de veille épidémiologique ont circulé plus difficilement du fait de la difficulté à détecter les cas, mais aussi de l'opacité qui a régné en Chine sur l'émergence de la pandémie. Une initiative a toutefois donné une grande visibilité aux chiffres sur la circulation de l'épidémie. Le 21 janvier, Ensheng Dong, doctorant à l'école d'ingénieur de l'université Johns Hopkins, sur les conseils de Lauren Gardner, sa directrice de thèse, a développé un tableau de bord de suivi de la pandémie⁸. Composé d'une carte du monde avec des points rouges regroupant les cas, le tableau de bord a attiré en quelques semaines plusieurs millions de visiteurs.

4. « Timeline of the COVID-19 pandemic in January 2020 », 2021, *Wikipedia*. Adresse : https://en.wikipedia.org/w/index.php?title=Timeline_of_the_COVID-19_pandemic_in_January_2020

5. « Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1, complete genome », 2020. Adresse : <http://www.ncbi.nlm.nih.gov/nucleotide/MN908947.1>

6. Gale C. (2021), « Covid-19, la course aux vaccins », *ARTE*. Adresse : <https://www.arte.tv/fr/videos/104055-000-A/covid-19-la-course-aux-vaccins/>

7. Corman V. M. et al. (2020), « Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR », *Eurosurveillance*, vol. 25, n° 3. Adresse : <https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.3.2000045>

8. « How the Johns Hopkins coronavirus tracker was built and what it tells us », *The Washington Post*. Adresse : https://www.washingtonpost.com/local/johns-hopkins-tracker/2020/06/29/daea7eea-a03f-11ea-9590-1858a893bd59_story.html

« Meet The Team Behind The Coronavirus Tracker Watched By Millions », 2020, *NPR*. Adresse : <https://text.npr.org/833073670>

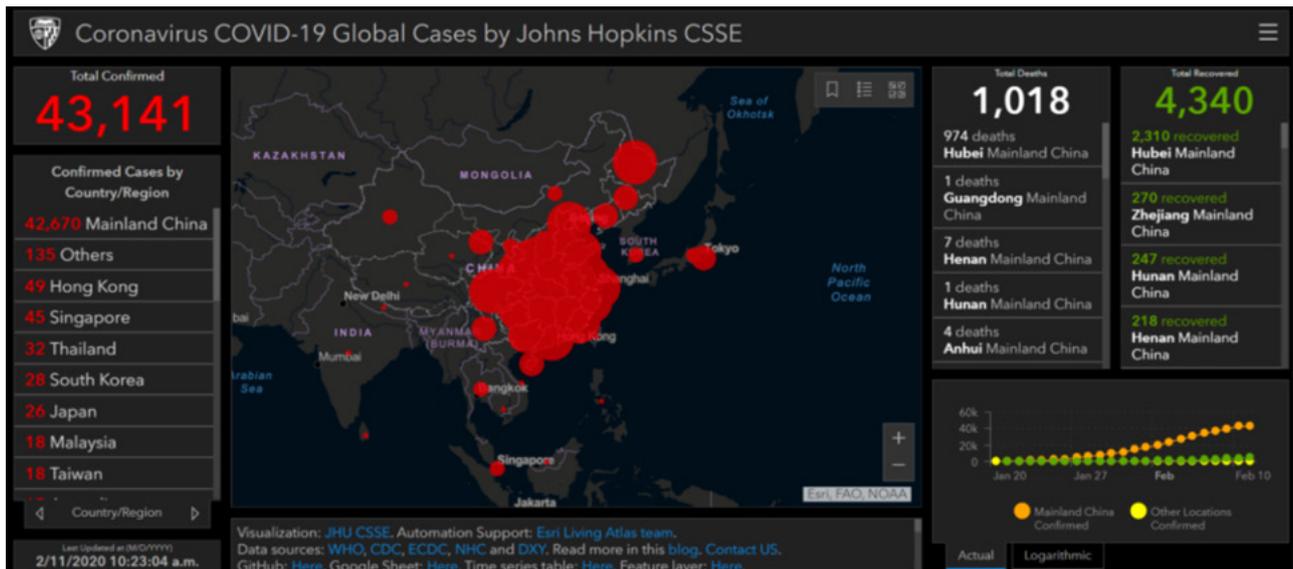


Figure 1 – Tableau de bord de l'université Johns Hopkins (février 2020)

Dans un premier temps, jusqu'au 31 janvier, toutes les données étaient collectées manuellement, puis le processus a été semi-automatisé (Dong, 2020). Pour la Chine, le tableau de bord mettait à jour toutes les 15 minutes DXY, une plateforme alimentée par des professionnels de santé qui agrégeait les chiffres transmis par les autorités à la presse locale dans chacune des provinces. Dans les autres pays, les données étaient actualisées principalement de manière manuelle à partir des sites des autorités de santé locales et nationales. Ces données sont devenues la source internationale de référence, au moins pendant la première vague de l'épidémie, pour connaître l'évolution des contaminations partout dans le monde. Le tableau de bord de l'université Johns Hopkins est resté une référence au-delà de la première vague même si d'autres acteurs internationaux ont pris le pas, comme Our World in Data qui agrège aujourd'hui les données sur les contaminations, les tests, les hospitalisations, les décès ou les vaccinations dans une centaine de pays dans le monde.

Pourquoi revenir sur cet outil dans un article centré sur la France ? Le tableau de bord de l'université Johns Hopkins a posé une exigence d'ouverture et de mise à jour quotidienne des données, devenues la norme internationale en matière de veille sanitaire sur la Covid-19. Cet outil a aussi montré l'importance de la médiation, au-delà des données brutes, pour aider le public à suivre l'évolution de l'épidémie.

2. Un dispositif de surveillance en France mis en place en urgence

En France, un dispositif de surveillance de la propagation de la Covid-19 a été mis en place à partir du 10 janvier 2020⁹. Conçu à la suite de la pandémie de grippe A (H1N1) en 2009, le plan national de prévention et de lutte « pandémie grippale » de 2011 prévoyait des outils de pilotage adaptés aux quatre stades de l'épidémie¹⁰. Le premier stade vise à freiner l'introduction du virus sur le territoire. Le 24 janvier 2020, les trois premiers cas de Covid-19 sont déclarés en France¹¹. Le deuxième stade qui vise à freiner la propagation du virus est déclenché le 29 février avec une centaine de cas déclarés sur le territoire. Les autorités ont tenté de retracer

9. Santé Publique France (2020), Livret de présentation « Le suivi de l'épidémie de COVID-19 », Adresse : https://solidarites-sante.gouv.fr/IMG/pdf/livret_suivi_epidemie_covid-19_-_270320.pdf

10. Secrétariat général de la défense et de la sécurité nationale (2011), Plan national de Prévention et de lutte « Pandémie grippale ». Adresse : https://solidarites-sante.gouv.fr/IMG/pdf/Plan_Pandemie_Grippale_2011.pdf

11. « Chronologie de la pandémie de Covid-19 en France », 2021, Wikipédia. Adresse : https://fr.wikipedia.org/w/index.php?title=Chronologie_de_la_pand%C3%A9mie_de_Covid-19_en_France&oldid=184735959

et d'isoler l'ensemble des cas contacts, mais très vite, l'épidémie est devenue hors de contrôle. Au troisième stade qui est déclaré par le gouvernement le 14 mars 2020, le plan national de prévention prévoit l'arrêt de la surveillance individuelle des cas avec l'augmentation rapide du nombre de cas. Le ministère des Solidarités et de la Santé et Santé Publique France étaient alors passés dans une stratégie de surveillance dite « syndromique » qui consiste à identifier la part de la population qui présente des symptômes de l'infection à la Covid-19 même si ces signes se retrouvent dans d'autres maladies¹². Pour assurer cette surveillance, Santé Publique France a comptabilisé les passages aux urgences hospitalières pour suspicion de Covid-19 en s'appuyant sur un réseau de 194 services de réanimation et le dispositif SI-VIC (Système d'information pour le suivi des victimes) conçu en 2016 suite aux attentats de Paris pour des événements et des catastrophes impliquant un grand nombre de victimes. Concrètement, ces systèmes d'information ont permis de collecter des indicateurs essentiels pour suivre l'engorgement des hôpitaux comme le nombre de patients hospitalisés pour Covid-19, placés en réanimation, guéris ou décédés. La médecine dite « de ville » a aussi fait l'objet d'une surveillance avec une remontée du nombre quotidien de patients présentant les signes d'une infection respiratoire aiguë. Ces données sont collectées parmi les praticiens membres du réseau Sentinelles ou d'une des associations SOS Médecins puis extrapolées à l'ensemble des généralistes de la même zone¹³. La prévalence des pathologies saisonnières comme la grippe est estimée avec une méthodologie similaire.

Avec l'accélération de la propagation du virus, les autorités ont continué à communiquer quotidiennement le nombre de cas confirmés malgré des avertissements sur son site : « *Du fait de la difficulté de l'identification et de la confirmation biologique de l'ensemble des cas de Covid-19, les données présentées sous-estiment le nombre réel de cas*¹⁴. » La France n'a pas procédé à un dépistage de grande envergure de la Covid-19, les tests ayant généralement été réservés à des publics particuliers : malades à l'hôpital, personnel soignant, personnes à risques... Malgré ces limites évidentes, l'indicateur du nombre de cas confirmés a pourtant été très scruté, à la fois en valeur absolue et en comparaison avec d'autres pays, y compris certains ayant adopté des politiques de tests massifs de leur population. Le nombre de tests pratiqués en France était lui-même à prendre avec précaution. Les tests remontaient des laboratoires hospitaliers et d'un système de surveillance appelé « 3labos » reposant sur trois laboratoires centralisateurs de prélèvements : Eurofins, Biomnis et Cerba. Ces réseaux fédèrent 1300 laboratoires représentant environ 30 % du total des 4 000 sites de prélèvements en France. Un article de Libération expliquait cette situation par l'urgence qui a conduit les autorités à s'appuyer sur les systèmes d'information déjà en place pour la surveillance épidémique avec ces laboratoires sous-traitants traditionnels des hôpitaux français¹⁵. Ces trous dans l'appareil statistique se sont retrouvés aussi sur le nombre de décès qui était probablement l'indicateur le plus scruté pour évaluer l'impact de l'épidémie. On le sait aujourd'hui, environ 44 % des décès liés à la Covid-19 ont eu lieu dans des établissements d'hébergement pour personnes âgées dépendantes (EHPAD)¹⁶ et d'autres établissements médicaux sociaux. À partir du mois de mars, plusieurs clusters sont apparus dans ces établissements, mais leur signalement ne faisait pas l'objet d'une collecte systématique, faute d'outil adapté. Le 28 mars, le ministère des Solidarités et de la Santé a déployé une application pour permettre à ces établissements de signaler les

12. Santé Publique France (2020), *op.cit.*

13. Réseau Sentinelles > France > Surveillance continue.
Adresse : <https://www.sentiweb.fr/france/fr/?page=methodes>

14. « Cas confirmés, hospitalisés, guéris... Comment les malades du coronavirus sont comptabilisés en France », 2020, *FranceInfo*. Adresse : https://www.francetvinfo.fr/sante/maladie/coronavirus/cas-confirmes-hospitalises-gueris-comment-les-malades-du-coronavirus-sont-comptabilises-en-france_3884155.html

15. Mathiot C., « Covid-19 : est-il vrai que la France est incapable de compter le vrai nombre de tests pratiqués chaque jour ? », *Libération*. Adresse : https://www.liberation.fr/checknews/2020/04/15/covid-19-est-il-vrai-que-la-france-est-incapable-de-compter-le-vrai-nombre-de-tests-pratiques-chaque_1785211/

16. « Les résidents d'Ehpad représentent 44 % des morts du Covid-19 », 2020, *Le Monde.fr*. Adresse : https://www.lemonde.fr/les-decodeurs/article/2020/12/03/les-residents-d-ehpad-representent-44-des-morts-du-covid-19_6062084_4355770.html

cas possibles (ceux présentant le tableau de la maladie) ou confirmés par un test PCR parmi les pensionnaires ou le personnel ainsi que les décès dans la structure ou à l'hôpital de ces patients¹⁷. Les établissements devaient réaliser un signalement dès le premier cas détecté afin de suivre les cas et les décès et de mettre en place des mesures de gestion des cas groupés. Ces indicateurs ont été intégrés dès le 2 avril dans les points épidémiologiques de Santé Publique France¹⁸.

3. Des premières ouvertures de données incitées par des initiatives citoyennes

Dès l'apparition des premiers cas, Santé Publique France et le ministère de la Santé ont communiqué ces indicateurs aux médias sans pour autant ouvrir les données. Ces indicateurs étaient principalement communiqués sous la forme de communiqués de presse au format PDF ou à l'oral lors de points presse : impossible de télécharger les données pour réaliser des analyses ou proposer de nouvelles visualisations. Au moment du confinement, Santé Publique France n'avait pas de compte sur data.gouv.fr (Chignard, 2021). Son outil d'observation cartographique Geodes n'a inclus certaines données sur la Covid-19 qu'à partir du 18 mars 2020¹⁹. Ce manque de données publiques actualisées et dans un format exploitable était déploré à de nombreuses reprises sur les réseaux sociaux à la fois par des chercheurs, mais aussi des acteurs variés, sans expertise sanitaire spécifique, qui ont voulu proposer leur aide face à la crise. Philippe Lemoine, chercheur en philosophie à l'université Cornell, se faisait leur porte-parole le 16 mars 2020 sur Twitter²⁰ déplorant la non-accessibilité des données et la comparait avec des pays comme la Corée du Sud ou l'Italie qui ont ouvert les données détaillées utilisées pour la gestion de la pandémie : « *Pour que les experts puissent donner des conseils de bonne qualité aux décideurs politiques, il faut que des données aussi complètes que possible soient publiées régulièrement et, à cet égard, la situation est absolument scandaleuse. Le site de Santé Publique France ne publie qu'un malheureux communiqué sur le nombre de cas et de morts en France [...] Les agences régionales de santé publient aussi des chiffres sur ce qui se passe sur leur territoire, mais ces derniers sont aussi très pauvres et les informations publiées varient beaucoup selon les régions.* »

Pour y remédier, des collectifs citoyens ont produit et actualisé des données dès le début du mois de mars 2020. Le 6 mars, Lior Perez, ingénieur en science des données chez Météo France, lançait un appel sur LinkedIn pour répondre à cette situation²¹ : « *Amis data scientists, avez-vous quelques minutes pour m'aider à monter un projet open data ? L'objectif est de consolider les données Covid-19 françaises dans un fichier csv, comme l'ont fait des data scientists coréens dans leur pays. Nous ne savons pas nous servir d'un stéthoscope, mais nous pouvons peut-être aider ceux qui savent en leur apportant de la data facilement exploitable et de la dataviz. Les contributions attendues sont simples : lire quelques communiqués de presse et reporter les infos dans un csv. Avec un contributeur par région, ça ne prendrait que quelques minutes.* »

17. Santé Publique France (2020), « Protocole de la surveillance active des cas de COVID-19 en EHPAD et EMS ». Adresse : https://www.preventioninfection.fr/wp-content/uploads/2020/03/2020_mars-28_COVID-19_protocole-SV-SGLT_ems_SPF.pdf

18. Santé Publique France (2020), « Point épidémiologique hebdomadaire du 2 avril ».

19. Santé Publique France (2020), « Covid-19 : Point épidémiologique - Situation au 15 mars 2020 à minuit ». Adresse : <https://www.santepubliquefrance.fr/content/download/237536/2535424>

20. Philippe Lemoine sur Twitter, 2020. Adresse : <https://twitter.com/phl43/status/1239590836947517442>

21. Perez L. (2020), « Amis data scientists, avez-vous quelques minutes pour m'aider à monter un projet open data ? », *LinkedIn*. Adresse : https://www.linkedin.com/posts/liorperez_opendata-covid19-dataset-activity-6642013973241704448-55L2/



Figure 2 – Message de Lior Perez sur LinkedIn

En publiant ce message, il s’est inspiré d’un travail similaire réalisé en Corée du Sud et a créé un dépôt sur GitHub contenant un premier fichier pour six régions. Les données étaient collectées²² à partir des communiqués de presse des Agences régionales de santé (Ars). Le 14 mars 2020²³, ce projet a fusionné avec celui d’une autre équipe initiée à titre personnel par Jérôme Desboeufs et Bastien Guerry, deux agents d’Etalab, la mission interministérielle en charge de l’ouverture des données (Goeta, 2018). L’initiative Data Against Covid²⁴ se présentait comme « une organisation informelle issue de la société civile avec des développeurs, des data scientists, des ingénieurs, des chefs de projets et talents de tous corps de métiers qui s’organisent ensemble pour fournir des données consolidées sur l’épidémie de covid 19 et proposer des outils de visualisation » afin de « favoriser une information claire, précise et accessible pour tous sur la propagation du virus ». Une communauté en ligne intitulée opencovid19-fr s’est structurée autour de canaux de discussions sur Slack. 1 200 membres y ont échangé près de 90 000 messages (Chignard, 2021). Dans la continuité du projet de Lior Perez, une cinquantaine de personnes se sont affairées à collecter les données à partir des sites de Santé Publique France, des agences régionales de santé, le ministère des Solidarités et de la Santé et certaines préfectures. Un guide de la contribution a permis d’établir une méthodologie qui comprenait l’archivage systématique des sources et la production de métadonnées précises sur chacun des chiffres²⁵. Les contributeurs reportaient même les chiffres annoncés lors des points presse quotidiens du directeur général de la santé et du ministre. À partir du 13 mars 2020, cette communauté a produit un tableau de bord de suivi de l’évolution de l’épidémie en France intitulé veille-coronavirus.fr alimenté par

22. Perez L. (2020), « Update README.md · lperez31/coronavirus-france-dataset@5e87cb8 », *GitHub*. Adresse : https://github.com/lperez31/coronavirus-france-dataset/commit/5e87cb88abc46458639bd218254d3136e0ecc97d?short_path=b335630#diff-b335630551682c19a781afebcf4d07bf978fb1f8ac04c6bf87428ed5106870f5

23. « Ces citoyens luttent contre le coronavirus grâce à une arme : la data », *Les Echos Start*. Adresse : <https://start.lesechos.fr/societe/engagement-societal/ces-citoyens-luttent-contre-le-coronavirus-grace-a-une-arme-la-data-1191396>

24. « Data Against COVID-19 / FR ». Adresse : <https://opencovid19-fr.github.io/>

25. GitHub – opencovid19-fr/data : Consolidation des données de sources officielles concernant l’épidémie de COVID19. Adresse : <https://github.com/opencovid19-fr/data>

les données collectées par les contributeurs d'opencovid19-fr²⁶.

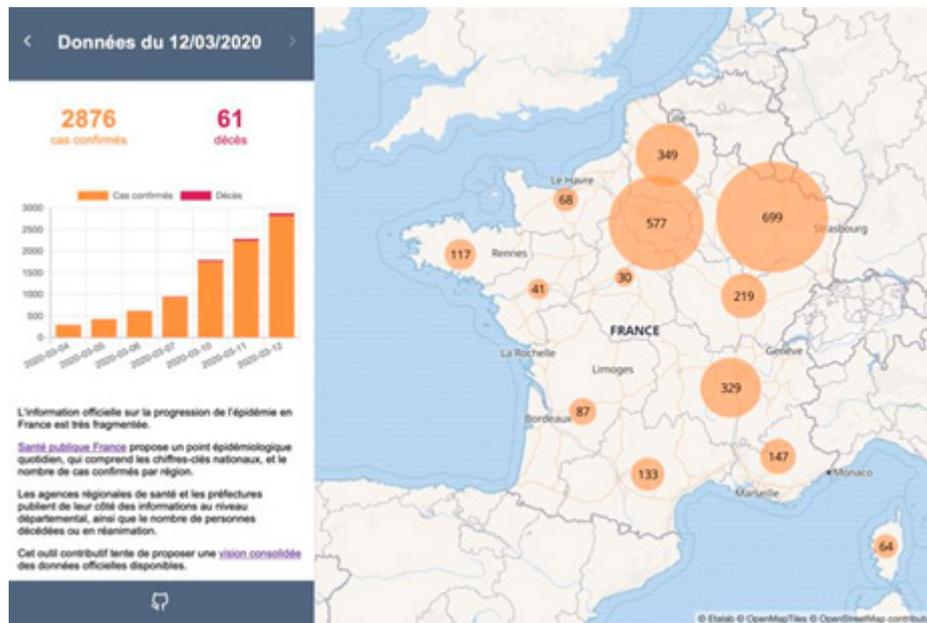


Figure 3 – Tableau de bord veille-coronavirus.fr

Les autorités sanitaires ont diffusé certaines données officielles en open data à partir du 19 mars 2020 quand Santé Publique France a publié sur data.gouv.fr les données sur la pression hospitalière (nombre de patients hospitalisés, en réanimation, retournés à domicile ou décédés), les urgences (nombre de passages pour suspicion de Covid-19) et la médecine de ville (actes médicaux au sein du réseau SOS Médecins pour suspicion de Covid-19)²⁷. Le 27 mars, Santé Publique France publiait en complément un jeu de données sur le nombre de tests de dépistage de la Covid-19 réalisés par département au sein du réseau « 3labos » évoqué précédemment. Ce système de surveillance a été remplacé en mai 2020 par un système d'information dédié au dépistage de la Covid-19, nommé SI-DEP. Opéré par l'Assistance publique hôpitaux de Paris (AP-HP), SI-DEP collecte à l'échelle nationale tous les résultats de tests de dépistage de la Covid-19²⁸. En complément de ces données, l'Insee a publié le 27 mars 2020 « à titre exceptionnel » des données quotidiennes sur l'évolution de la mortalité en France, toutes causes de décès réunies. Chaque vendredi, l'institut statistique publiait des données sur le nombre de décès constatés 11 jours plus tôt (7 jours pour les communes ayant dématérialisé la déclaration des décès). Des fichiers détaillés permettent de connaître la date et la commune du décès, l'âge et le sexe du défunt, sa commune de résidence et le lieu de décès (hôpital ou clinique, domicile, maison de retraite, autres)²⁹. Dans un billet accompagnant leur publication³⁰, les statisticiens de l'Insee ont averti que l'évaluation de la surmortalité due à la Covid-19 exige « beaucoup d'humilité ». La situation inédite du confinement a provoqué de nombreux effets qui perturbent l'interprétation des données : la transmission ralentie des déclarations par voie postale, la sous-mortalité pour d'autres causes que la Covid-19 (comme les accidents de la route) ou une surmortalité pour les accidents cardio-vasculaires ou domestiques. S'interrogeant sur la

26. Tableau de bord COVID-19, 2020.

Adresse : <https://web.archive.org/web/20200329222434/https://dashboard.covid19.data.gouv.fr>

27. Données relatives à l'épidémie du covid-19 – data.gouv.fr, 2020.

Adresse : <https://web.archive.org/web/20200323212238/https://www.data.gouv.fr/fr/datasets/donnees-relatives-a-lepidemie-du-covid-19/>

28. « Le Système d'Information de Dépistage Populationnel (SI-DEP) opéré par l'AP-HP évolue pour fournir des résultats de tests certifiés », APHP. Adresse : <https://www.aphp.fr/actualite/le-systeme-dinformation-de-depistage-populationnel-si-dep-opere-par-lap-hp-evolue-pour>

29. « Téléchargement des fichiers des décès quotidiens – Nombre de décès quotidiens », Insee. Adresse : <https://www.insee.fr/fr/statistiques/4487988?sommaire=4487854#documentation-sommaire>

30. « Mourir de la grippe ou du coronavirus : faire parler les chiffres de décès publiés par l'Insee... avec discernement », 2020, Le blog de l'Insee. Adresse : <https://blog.insee.fr/mourir-de-la-grippe-ou-du-coronavirus-faire-parler-les-chiffres-de-deces-publies-par-linsee-avec-discernement/>

très forte hausse des décès à domicile dans certaines communes, un journaliste de Médiapart a montré que le lieu du décès devait être pris avec grande précaution³¹ : beaucoup de mairies cochaient la case « à domicile » pour le lieu du décès dès lors que le défunt était domicilié dans un Ehpad.

Après ces premières ouvertures de données, fin mars 2020, le tableau de bord veille-coronavirus.fr a rapidement été repris au sein de l'État afin de devenir l'outil officiel du gouvernement pour le suivi de la pandémie. Cette internalisation a été facilitée par la présence de certains agents d'Etalab au sein du projet opencovid19-fr et d'une collaboration entre Etalab, le ministère des Solidarités et de la Santé, le Service d'information du gouvernement (SIG) et Santé Publique France³². Le 28 mars 2020, lors d'un point presse quotidien avec le Premier ministre, le directeur général de la santé, Jérôme Salomon, le 28 mars, annonçait l'ouverture des données liées à la pandémie et la publication du tableau de bord.

Son discours vantait les bénéfices de l'open data et soulignait l'importance de la collaboration avec les acteurs de la société civile, désignant vraisemblablement les bénévoles d'opencovid19-fr : « *Le libre accès aux données relatives à l'épidémie, ce qu'on appelle l'open data, assure la confiance des citoyens dans les éléments qui leur sont communiqués, favorise les actions de prévention contre la propagation du virus et facilite la prise de décision. La diffusion et l'exploitation de ces données ont été rendues possibles grâce à une collaboration étroite entre plusieurs administrations et la société civile.* » Ces données ont encore gagné en importance au moment du déconfinement. Le 28 avril 2020, le Premier ministre Edouard Philippe présentait la stratégie de déconfinement à l'Assemblée nationale, comprenant une différenciation selon la circulation du virus dans les territoires. La carte de vigilance distinguait pour chaque département un niveau rouge ou vert basé sur trois indicateurs : la proportion de passages aux urgences pour une suspicion de Covid-19, le taux d'occupation des lits de réanimation par des patients atteints de Covid-19 et le taux de couverture des besoins en tests estimés au 11 mai³³. À partir du 28 mai, ces critères ont été remplacés par quatre indicateurs : le taux d'incidence (tests positifs pour 100 000 habitants), le taux de positivité des tests, l'évolution du R0 (nombre de personnes contaminées par chaque malade) et le taux d'occupation des lits en réanimation. Ces indicateurs très scrutés ont été publiés le 30 avril sur data.gouv.fr et visualisés dans le tableau de bord du gouvernement. Fin juin 2021, des données au niveau infradépartemental (EPCI) ont été publiées par Santé Publique France pour permettre un suivi plus localisé de l'épidémie³⁴.

4. Des données réutilisées massivement, une relance de la politique d'ouverture des données

Les données publiées sur data.gouv.fr liées à la Covid-19 ont répondu à une très forte demande de la part d'utilisateurs très variés. De mars à octobre 2020, plus du tiers des visites (37 %) sur data.gouv.fr concernaient des jeux de données publiés par Santé Publique France, l'Insee (fichier des personnes décédées) ou opencovid19-fr³⁵. Si l'on consulte les statistiques disponibles en ligne sur stats.data.gouv.fr, la fréquentation du portail national data.gouv.fr a explosé en particulier pendant les périodes de confinement.

31. « Suivi de l'épidémie : le cafouillage de l'Insee sur les morts à domicile », *Mediapart*. Adresse : <https://www.mediapart.fr/journal/france/190520/suivi-de-l-epidemie-le-cafouillage-de-l-insee-sur-les-morts-domicile?onglet=full>

32. Etalab (2020), « Comment les administrations ont collaboré à l'ouverture des données du Coronavirus : le cas français », *Le blog d'Etalab*. Adresse : <https://www.etalab.gov.fr/comment-les-administrations-ont-collabore-a-louverture-des-donnees-du-coronavirus-le-cas-francais>

33. Données de la carte de vigilance COVID-19 – data.gouv.fr. Adresse : <https://www.data.gouv.fr/fr/datasets/donnees-de-la-carte-de-vigilance-covid-19/>

34. Samuel Goëta sur Twitter, 2020. Adresse : <https://twitter.com/samgoeta/status/1323341113965580289>

35. Ce chiffre a été calculé à partir de l'outil de statistiques de fréquentation, stats.data.gouv.fr, qui comptabilise le trafic sur l'ensemble des sites d'Etalab. Les données brutes ayant permis ce calcul sont publiées par Dataactivist sur data.gouv.fr : <https://www.data.gouv.fr/fr/datasets/nombre-de-visiteurs-par-jour-depuis-2014-des-url-de-data-gouv-fr-1/>

Data.gouv.fr : des pics de fréquentation liés aux confinements

Les périodes de confinement sont indiqués dans les zones grisées

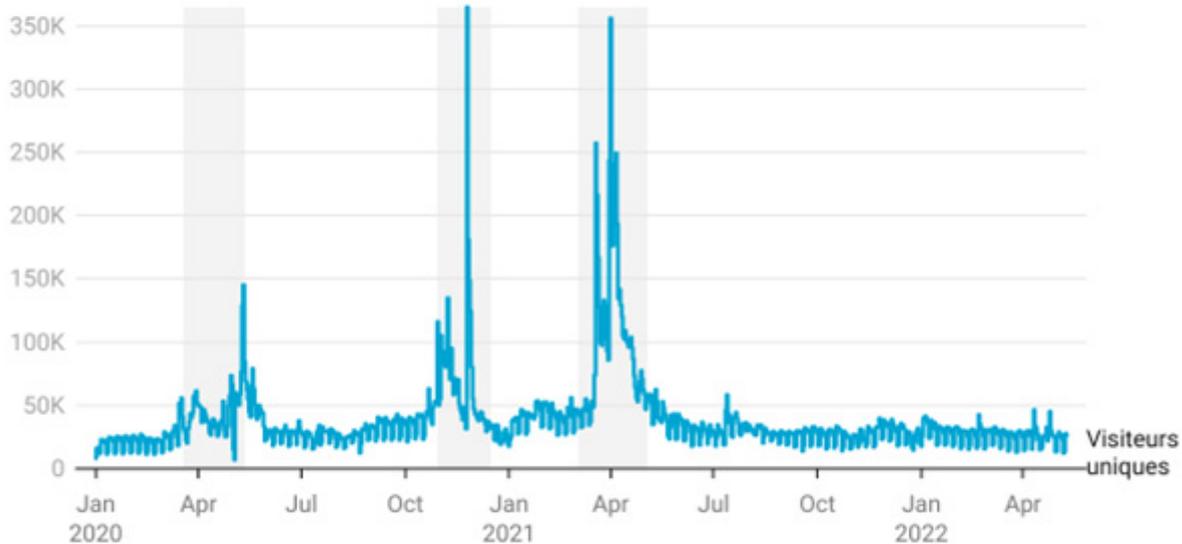


Chart: Samuel Goëta • Source: Stats.data.gouv.fr • Created with Datawrapper

Figure 4 – Évolution de la fréquentation de data.gouv.fr (Source des données : stats.data.gouv.fr)

En mai 2020, les données ont servi en particulier à calculer les indicateurs de la carte de vigilance qui déterminaient département par département les conditions de la reprise des activités. Le 24 novembre 2020, le président Emmanuel Macron déclarait lors d'un discours télévisé envisager le déconfinement d'ici au 15 décembre 2020 si la situation sanitaire continuait de s'améliorer avec des objectifs de 5000 nouveaux cas positifs quotidiens et moins de 3000 patients en réanimation. Les données liées à la propagation du virus ont alors été très suivies pour prévoir si le déconfinement allait se réaliser et si les réunions lors des fêtes de fin d'année étaient envisageables.

La forte croissance de la fréquentation de data.gouv.fr en 2020 et l'importance dans le débat public des visualisations issues de données ouvertes a contribué à relancer la politique d'open data du gouvernement. Avant la pandémie, l'ouverture des données semblait sur la sellette depuis la nomination en décembre 2018 de Nadi Bou-Hanna à la Direction interministérielle au numérique (DINUM), à laquelle Etalab est rattachée. S'inscrivant en rupture avec son prédécesseur Henri Verdier, Nadi Bou-Hanna avait placé l'ouverture des données en retrait face aux enjeux d'échange de données entre administrations. Cela s'est reflété en particulier dans la stratégie et la feuille de route 2019-2022 de la DINUM³⁶ intitulée « TECH.GOUV » qui, avant sa mise à jour en septembre 2021, ne contenait aucune mention de l'ouverture des données ni de data.gouv.fr. Le manque d'appui du directeur interministériel au numérique à l'open data pouvait être perçu par certains comme une remise en cause du gouvernement du principe d'open data par défaut prévu par la loi pour une République numérique de 2016.

36. TECH.GOUV : l'État lance un programme pour accélérer la transformation numérique du service public. Adresse : <https://www.numerique.gouv.fr/actualites/tech-gouv-accelerer-la-transformation-numerique-du-service-public/>



Figure 5 – Remise du rapport du député Eric Bothorel au Premier ministre Jean Castex
(Photo : Service d'information du Gouvernement)

Le rapport du député Éric Bothorel remis au Premier ministre le 23 décembre 2020, en tentant de tirer les leçons de la pandémie, a donné lieu à une relance vigoureuse de l'ouverture des données au sein de l'État. Le 22 juin 2020, le Premier ministre Édouard Philippe lui a confié une lettre de mission qui débute par une référence à la pandémie : « *La gestion de l'épidémie de Covid-19 l'a démontré [...] la prise de décision dépend plus que jamais de la maîtrise des données, publiques comme privées, donc de leur circulation et de leur partage.* » Le député de la 5^e circonscription des Côtes-d'Armor a été sélectionné pour guider cette mission du fait de ses nombreuses expériences professionnelles dans l'informatique. Renaud Vedel, préfet coordonnateur national pour l'intelligence artificielle, et Stéphanie Combes, directrice du Health Data Hub, ont aussi composé la mission à laquelle ont contribué plusieurs agents de l'administration. S'interrogeant sur « *notre capacité collective à toujours mieux organiser l'ouverture, le partage et l'exploitation des données et des codes sources* », le Premier ministre a chargé le député d'analyser les opportunités et les freins à lever pour favoriser l'ouverture par défaut des données. Le député a aussi eu pour mission d'accélérer le partage des données dites d'intérêt général collectées par des acteurs privés et l'ouverture des codes sources des applications et logiciels développés par les administrations. Au terme de six mois d'auditions et de concertation, le député a rendu son rapport le 23 décembre 2020³⁷. Le rapport a abouti à une circulaire le 28 avril 2021 par le Premier ministre imposant notamment la nomination d'un administrateur des données, des algorithmes et des codes sources (AMDAC) dans chaque ministère et la publication par chaque ministère d'une feuille de route sur l'ouverture des données et des codes sources le 28 septembre 2021³⁸. Les quinze feuilles de route publiées ont totalisé plus de 500 actions sur les enjeux des AGDAC. Certains acteurs de la société civile ont jugé pusillanimes les actions sur l'ouverture des données³⁹. Quoi qu'il en soit, la mobilisation des agents dans les ministères pour l'élaboration de ces feuilles de route a replacé le sujet de l'open data comme un sujet d'importance pour l'administration.

37. Remise du rapport sur la politique publique de la donnée, des algorithmes et des codes sources, *Gouvernement.fr*. Adresse : <https://www.gouvernement.fr/remise-du-rapport-sur-la-politique-publique-de-la-donnee-des-algorithmes-et-des-codes-sources>

38. « Données, algorithmes et codes sources : une mobilisation générale sans précédent, à travers 15 feuilles de route ministérielles », 2021, *numerique.gouv.fr*. Adresse : <https://www.numerique.gouv.fr/actualites/donnees-algorithmes-codes-sources-mobilisation-generale-sans-precedent-15-feuilles-de-route-ministerielles/>

39. « Data, code, algos : 15 feuilles de route et plus de 500 actions », 2021, #TeamOpenData. Adresse : <https://teamopendata.org/t/data-code-algos-15-feuilles-de-route-et-plus-de-500-actions/31297>

5. L'émergence des médiateurs et le rôle de CovidTracker dans l'ouverture des données de l'administration

Le rapport du député Bothorel a aussi contribué à mettre en lumière le rôle des initiatives de la société civile pour « accélérer » l'innovation de la sphère publique et renforcer la transparence des décisions. Deux initiatives font l'objet d'une étude de cas dans le rapport : OpenCovid19-fr, évoqué précédemment, et CovidTracker. Cette dernière initiative mérite qu'on s'y attarde tant sa visibilité auprès du grand public a été importante et son rôle central dans l'ouverture des données liées à la pandémie à partir de la fin de l'année 2020. CovidTracker est une initiative lancée par Guillaume Rozier, un ingénieur en sciences des données de 24 ans qui, au moment de la première vague, terminait ses études à Telecom Nancy⁴⁰. En mars 2020, stagiaire dans une banque luxembourgeoise, ne pouvant pas télétravailler ni rester au Luxembourg, il a décidé d'occuper son temps libre en analysant les données de la pandémie fournies par l'université Johns Hopkins. Début mars, il publiait sur Twitter un graphique pour montrer un décalage de neuf jours⁴¹ entre la situation en France et celle en Italie où la croissance du nombre de cas était exponentielle.

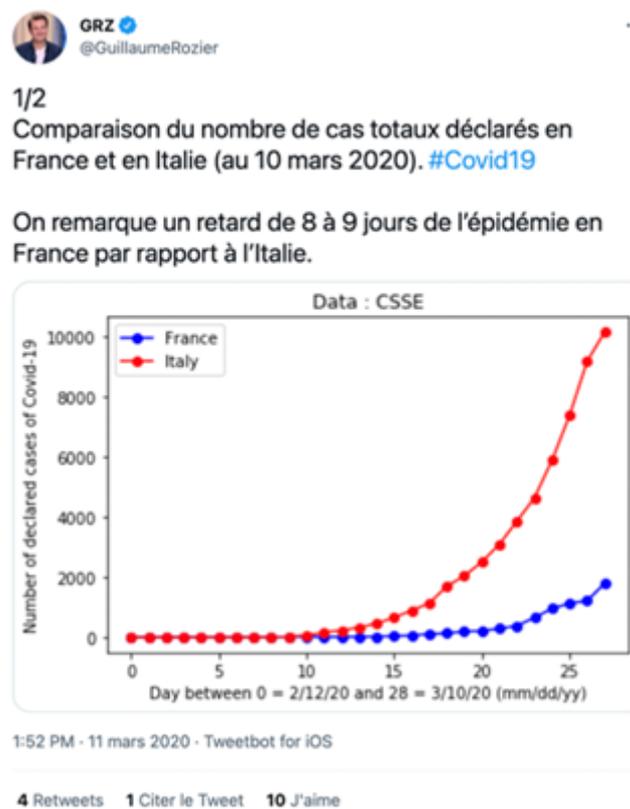


Figure 6 – Tweet de Guillaume Rozier le 11 mars 2020 comparant l'évolution du nombre de cas en France et en Italie

Au fur et à mesure de l'avancée de la première vague, ses tweets ont connu un grand succès aboutissant à la publication de son site intitulé CovidTracker en avril 2020. Le 13 avril 2020, 15 000 personnes l'avaient consulté⁴². De nombreux médiateurs, comme lui issus de formation

40. Bembaron E. et R. Enguérand (2020), « Comment est né le site très suivi CovidTracker », *Le Figaro*.

41. Rozier G. (2020), « Comparaison du nombre de cas totaux déclarés en France et en Italie (au 10 mars 2020). #Covid19 On remarque un retard de 8 à 9 jours de l'épidémie en France par rapport à l'Italie. », *Twitter*. Adresse : <https://twitter.com/GuillaumeRozier/status/1237723117419716608>

42. « Épidémie covid-19 : un étudiant de Télécom Nancy réalise un outil de visualisation de données », 2020. Adresse : <http://telecomnancy.univ-lorraine.fr/fr/big-data-0>

en sciences des données, ont contribué à visualiser et à expliquer les données de la pandémie. En novembre 2020, *Le Parisien* publiait un article sur « ces jeunes geeks passés maîtres en suivi de l'épidémie » dressant le portrait des « 3G » : Germain Forestier, Guillaume Saint-Quentin et Guillaume Rozier, trois spécialistes des données analysant quotidiennement sur Twitter les données de la pandémie⁴³. Selon une personne de la rédaction de *Quotidien* – l'émission de grande écoute de TMC animée par Yann Barthès – que j'ai interrogée, cet article a convaincu les journalistes d'inviter Guillaume Rozier sur le plateau pour présenter CovidTracker au million et demi de téléspectateurs de l'émission. L'ingénieur y a livré un plaidoyer en faveur de l'ouverture des données présentées comme une « mine d'or sous-exploitée » et un moyen d'améliorer la qualité de nos vies⁴⁴.

En plus de fournir des outils de visualisation des données, CovidTracker a joué un rôle essentiel dans l'ouverture de nouvelles données liées à la pandémie. En novembre⁴⁵ 2020, Elias Orphelin, un des premiers contributeurs de CovidTracker, relayait avec Guillaume Rozier sur Twitter une demande d'ouverture des données de détection de la concentration du SARS Cov-2 dans les eaux usées. Étudiant en dernière année d'école de commerce à l'ESCP Paris et analyste de données dans le cadre de son apprentissage, Elias Orphelin souhaitait intégrer ces données dans CovidTracker pour compléter les données de tests issues de SI-DEP qui sont très sensibles à l'évolution de la politique de tests et aux pratiques des laboratoires. Par exemple, les jours fériés réduisent la capacité de tests et créent artificiellement dans les données des baisses du taux de contamination, encore plus quand ils s'enchaînent en novembre ou en avril. Accompagné par l'équipe de madada dans la formulation et le ciblage de sa demande, Elias Orphelin a transmis sa demande au ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation (MESRI) qui a consacré 3 millions d'euros au projet d'Observatoire épidémiologique dans les eaux usées (OBEPINE). Lancé en juillet, OBEPINE vise à réaliser un suivi en temps réel des traces de Covid-19 dans les eaux usées sur tout le territoire national afin d'anticiper les évolutions de la pandémie. Cette demande s'appuyait sur l'obligation d'accès ouvert pour les données issues de recherches financées sur projets prévue par le Plan national pour la science ouverte annoncé par la ministre Frédérique Vidal, le 4 juillet 2018. La demande formulée par Elias Orphelin le 20 novembre 2020 a reçu une réponse favorable le 9 janvier 2021 de la part d'Isabelle Blanc, l'administratrice ministérielle des données du MESRI⁴⁶, le ministère ayant aussi largement communiqué sur Twitter cette réponse. Les bénévoles de CovidTracker ont découvert les modalités d'exercice du droit d'accès et aussi ses lenteurs. La réponse favorable n'a conduit à l'ouverture des données que six mois après la demande initiale alors que le Code des relations entre le public et l'administration (CRPA) prévoit un délai d'un mois pour le traitement des demandes d'accès aux documents administratifs.

Les bénévoles de CovidTracker ont aussi joué un rôle essentiel dans l'ouverture des données liées aux variants. À la fin du mois de décembre, avec l'émergence du variant dit « anglais » (renommé au printemps « alpha » par l'OMS), les craintes étaient fortes d'une diffusion de cette souche plus contagieuse sur le territoire français, ce qui est confirmé le 25 décembre 2020. Pour suivre cette propagation, le 7 février 2021, la Direction générale de la santé (DGS) a exigé que tout test positif fasse l'objet d'un criblage pour déterminer s'il s'agit d'une contamination

43. Berrod N. (2020), « Covid-19 : ces jeunes "geeks" passés maîtres en suivi de l'épidémie », *Le Parisien*.

44. *Quotidien* sur Twitter : « "L'analyse de données peut sauver des vies !" @GuillaumeRozier, le fondateur de CovidTracker, dans #Quotidien » <https://t.co/iVjf3xbInj> / Twitter.

Adresse <https://twitter.com/Qofficiel/status/1328790471531966464>

45. Orphelin E. (2020), « L'analyse des données des eaux usées pourrait devenir un indicateur essentiel dans le suivi de l'épidémie. Seul bémol : pas d'OpenData à court terme, pour des raisons de propriété des résultats, partagées entre de nombreux acteurs... Quel dommage ! », *Twitter*. Adresse : <https://twitter.com/eorphelin/status/1329366338058465280>

46. Données relatives à la concentration de SARS-CoV-2 dans les eaux usées – Une demande d'accès à l'information à Ministère de l'enseignement supérieur, de la recherche et de l'innovation, *Ma Dada*. Adresse : https://madada.fr/demande/donnees_relatives_a_la_concentra

par un variant d'intérêt⁴⁷. Le 19 février, Elias Orphelin sur Twitter⁴⁸ repérait un article de La Voix du Nord qui évoque un tableau Excel sur le bureau du Premier ministre Jean Castex avec la part de variant et l'évolution du taux d'incidence dans chaque département⁴⁹. Le contributeur de CovidTracker demandait sur Twitter la publication de ces données toujours pas accessibles au citoyen et formulait trois jours plus tard une demande d'accès aux documents administratifs⁵⁰ au ministère de la Santé via madada.fr. Guillaume Rozier s'est fait l'écho de la demande d'Elias Orphelin en lançant sur Twitter le hashtag #ExcelDeCastex repris une centaine de fois et en regrettant dans l'émission C à Vous sur France 5 la non-publication de ces données⁵¹. Le 8 mars 2021, les données sur la part des variants par département ont été publiées par Santé Publique France sur data.gouv.fr en réponse à cette mobilisation.

CovidTracker s'est aussi illustré dans l'ouverture des données liées à la vaccination. Le 27 décembre 2020, la première dose de vaccin Covid est administrée à une septuagénaire en Seine-Saint-Denis. La campagne de vaccination s'appuie sur un système d'information nommé VAC-SI (Vaccin Covid-SI) lancé le 4 janvier 2021. Le suivi de la campagne de vaccination était devenu essentiel, sa lenteur a beaucoup été critiquée en comparaison avec d'autres pays européens. Le 27 décembre 2020, Guillaume Rozier avait lancé VaccinTracker, un site présentant des graphiques et des cartes pour comprendre l'évolution de la campagne vaccinale⁵². Mais l'accès aux données du nombre de vaccinations en France a pris plusieurs semaines. Lors de la présentation de VAC-SI à la presse le 4 janvier 2021, le directeur général de l'Assurance maladie, Thomas Fatome, avait promis que les autorités sanitaires feraient un bilan quotidien de la campagne de vaccination dès le 5 janvier 2021⁵³. Pourtant, pendant la première semaine de janvier, seuls des chiffres arrondis ont été communiqués au public sur Twitter. Le ministère avait justifié ce choix par une volonté de privilégier le lancement de la vaccination à la remontée des chiffres. Le 30 décembre 2020, Guillaume Rozier avait été invité au ministère de la Santé à une réunion pour lui présenter la stratégie vaccinale en France et les difficultés rencontrées par les autorités sanitaires dans la remontée des chiffres⁵⁴. Quelques jours plus tard, le ministre de la Santé Olivier Véran s'est entretenu avec Guillaume Rozier le félicitant sur BFM TV pour son travail. À partir du 30 décembre, son cabinet a transmis à Guillaume Rozier, chaque jour, par SMS, les chiffres quotidiens de la vaccination. Le fondateur de CovidTracker publiait alors quotidiennement les chiffres de la campagne de vaccination, étant alors la seule source d'informations publiques sur ce sujet. Le 11 janvier, il mettait fin à cette exclusivité en publiant un message sur le site VaccinTracker relayé sur Twitter⁵⁵ : « *Le ministère de la Santé a contacté CovidTracker le 30 décembre afin de lui fournir des données officielles, plus exhaustives et à jour. Depuis, nous recevons régulièrement un nouveau chiffre du nombre de vaccinés de sa part, et nous le remercions pour cela. Cependant, cette situation n'est pas conforme avec nos principes d'OpenData. VaccinTracker ne sera désormais mis à jour qu'à partir de données publiques officielles, dès que celles-ci seront disponibles.* »

47. STRATEGIE DE FREINAGE DE LA PROPAGATION DES VARIANTES DU SARS-COV-2 : RENFORCEMENT SPECIFIQUE SUR LES VARIANTES D'INTERET 20H/501Y.V2 ET 20J/501Y.V3, 2021, DGS Urgent. Adresse : https://solidarites-sante.gouv.fr/IMG/pdf/dgs-urgent_no12_strategie_variant.pdf

48. Orphelin E. (2021), « Sur le bureau de @JeanCASTEX, un Excel comparant pour chaque département l'incidence et la part de variant. Monsieur le Premier Ministre, pouvez-vous svp transférer ce document aux citoyens ? Car ces chiffres ne sont toujours pas communiqués !! 😞 », *Twitter*.

49. « Covid-19 : face à "un ennemi qui change", l'exécutif obligé de naviguer à vue », 2021, *La Voix du Nord*.

50. « Circulation des variants de la Covid-19 sur le territoire - Une demande d'accès à l'information à Ministère des solidarités et de la santé », 2021, *Ma Dada*. Adresse : https://madada.fr/demande/circulation_des_variants_de_la_c

51. « C à vous sur Twitter », 2021, *Twitter*. Adresse : <https://twitter.com/cavousf5/status/1363916869414092802>

52. Rozier G. (2020), « 📄 Voici un tout nouvel outil, VaccinTracker ! Suivez l'évolution de la vaccination contre la #Covid19 en France. Nombre de personnes vaccinées, types de vaccins autorisés, phases de vaccination... », *Twitter*. Adresse : <https://twitter.com/GuillaumeRozier/status/1343209878064525312>

53. « Vaccination : un bilan quotidien à partir du mardi 5 janvier », 2021, *Le Figaro*. Adresse : <https://www.lefigaro.fr/sciences/vaccination-un-bilan-quotidien-a-partir-du-mardi-5-janvier-20210104>

54. « "Vaccintracker" : le site indépendant qui publie les chiffres de la vaccination en France », 2021, *Le Figaro*. Adresse : <https://www.lefigaro.fr/sciences/vaccintracker-le-site-independant-qui-publie-les-chiffres-de-la-vaccination-en-france-2021010>

55. Rozier G. (2021), « GRZ sur Twitter : "Message important à propos des données de vaccination. VaccinTracker ne sera désormais mis à jour qu'à partir de données publiques officielles #OpenData." », *Twitter*. Adresse : <https://twitter.com/GuillaumeRozier/status/1348576613928230912>



Figure 7 – Tweet de Guillaume Rozier annonçant la fin de la reprise des chiffres envoyés par le ministère des Solidarités et de la Santé

Se revendiquant des principes de l'open data, il refusait de mettre à jour le site tant que les données ne seraient pas publiques. Un premier pas est franchi par le ministère de la Santé le jour même, le 11 janvier, avec un communiqué de presse annonçant « plus de 138 000 » vaccinations en France depuis le début de la campagne⁵⁶ avec les chiffres pour chaque région. Les données sur le nombre quotidien de vaccinations en France et dans chaque région sont finalement publiées sur data.gouv.fr le 12 janvier 2021. Elles sont complétées le 23 janvier par des données sur les livraisons et les stocks de vaccins. À partir du 27 janvier, les données sont directement alimentées depuis VAC-SI alors qu'elles provenaient auparavant de remontées des ARS.

6. Le Hackathon Covid : renforcer et afficher la collaboration entre l'État et la société civile

Au début de l'année 2021, CovidTracker occupait une place croissante dans la médiation des données liées à l'évolution de la pandémie. Par exemple, le 25 février 2021 lors d'une conférence de presse, le Premier ministre Jean Castex s'appuyait sur un graphique représentant l'évolution du taux d'incidence des plus de 80 ans. La source affichée était CovidTracker⁵⁷. Les statistiques de recherches sur Google indiquent aussi la popularité de cette initiative, les recherches « covidtracker » ont dépassé par exemple celles sur le ministre de la Santé Olivier Véran entre mars et décembre 2021.

56. Ministère de la santé (2021), « Vaccination contre la COVID en France : au 11 janvier 2021, plus de 138 000 personnes ont été vaccinées en France ».

Adresse : https://solidarites-sante.gouv.fr/IMG/pdf/cp_dgs_chiffres_vaccination_110121_.pdf

57. Libeau D. (2021), « Pourquoi "CovidTracker" a été adopté par le gouvernement et les médias (entre autres) », *blog.DavidLibeau.fr*.

Adresse : <https://blog.davidlibeau.fr/pourquoi-covidtracker-a-ete-adopte-par-le-gouvernement-et-les-medias-entre-autres/>



Figure 8 – Évolution du nombre de recherches pour covidtracker, Olivier Véran et Oliver Veran entre mars 2020 et mai 2022 (Source : Google Trends)

À partir d'avril 2021, CovidTracker a étendu son champ d'action au-delà de la visualisation et de l'analyse de données en intervenant directement dans la campagne de vaccination avec le lancement de « Vite Ma Dose ! », un nouvel outil qui détecte automatiquement le prochain rendez-vous disponible à proximité sur Doctolib puis, progressivement, sur l'ensemble des plateformes de prise de rendez-vous. En proposant ce service répondant à la grande difficulté pour trouver un créneau à l'époque, la notoriété de CovidTracker a alors explosé. Au même moment, le site Covidliste est lancé à l'initiative de Martin Daniel, data scientist et Mathieu Ripert, ingénieur, en collaboration avec Dr Antoine Roux, pneumologue. Covidliste permet aux centres de vaccination de toucher rapidement les personnes à proximité inscrites sur une liste d'attente pour leur proposer les doses restantes à la fin de journée ou proches de leur expiration. Ces sites entièrement maintenus et animés par des bénévoles jouent progressivement un rôle important dans la campagne de vaccination. Ils ont aussi été critiqués pour renforcer les inégalités d'accès à la vaccination en favorisant les personnes les plus à l'aise avec les outils numériques.

L'émergence de ces nouveaux outils interrogeait au sein du gouvernement et de l'administration. D'une part, ils représentaient une forme d'idéal de l'open data dans lequel la société civile s'approprie rapidement les données ouvertes et propose de nouveaux services adaptés aux besoins de la population. D'autre part, Covidliste et Vite Ma Dose ont développé des services qui auraient pu être proposés par le service public. Ces plateformes ont aussi dû produire ou collecter des données privées (disponibilités des rendez-vous sur les plateformes) ou personnelles (candidats à la vaccination). Malgré sa proximité d'objet avec Covidliste, Vite Ma Dose s'est différencié en ne collectant pas de données à caractère personnel et en publiant en open data les données des rendez-vous disponibles. Face au succès de Vite Ma Dose, Guillaume Rozier a pu être présenté comme celui qui « ringardise le ministère de la Santé »⁵⁸. Plusieurs

58. « Guillaume Rozier : le CV du jeune ingénieur qui ringardise le ministère de la Santé », 2021, *L'Express.fr*. Adresse : https://www.lexpress.fr/actualite/societe/guillaume-rozier-le-cv-du-jeune-ingenieur-qui-ringardise-le-ministere-de-la-sante_2148411.html

articles opposaient l'agilité et l'innovation d'acteurs comme CovidTracker aux « lenteurs » de l'administration. Pourtant, Guillaume Rozier et les bénévoles de CovidTracker semblaient plutôt demandeurs de collaboration avec l'administration. Lors d'un échange en mars 2021 avec l'équipe de la Direction interministérielle de la transformation publique (DITP) en charge du plan d'action « gouvernement ouvert », portant sur la transparence de l'administration et la collaboration avec la société civile dans la mise en œuvre des politiques publiques, Elias Orphelin avait proposé l'idée d'un hackathon portant sur les données ouvertes liées à la pandémie. Un hackathon est un événement qui, sur une durée courte (souvent deux jours), rassemble un public divers qui contribue de manière bénévole à résoudre un certain nombre de défis, en développant de nouveaux outils (Goëta, 2018).

La proposition du hackathon a attiré l'attention de la DITP qui, début avril, apportait son soutien à l'initiative afin de faire émerger de nouveaux outils pour lutter contre la pandémie, mais aussi montrer le caractère innovant et collaboratif de l'administration et donner de la visibilité à la collaboration entre administration et société civile dans le cadre du programme « gouvernement ouvert. » Avec plusieurs collègues de Dataactivist, nous avons collaboré avec la DITP pour organiser et animer le hackathon Covid les 23 et 24 avril 2021, en lien avec de nombreuses administrations et acteurs de la société civile. Selon sa présentation officielle, ce hackathon visait à « unir [les] forces pour contribuer à mettre fin à la pandémie », « amplifier le mouvement d'ouverture des données et du code source déjà engagé par plusieurs administrations » et montrer que « l'ouverture et la collaboration répondent de manière plus efficace et rapide aux enjeux sans cesse renouvelés de la crise⁵⁹ ».

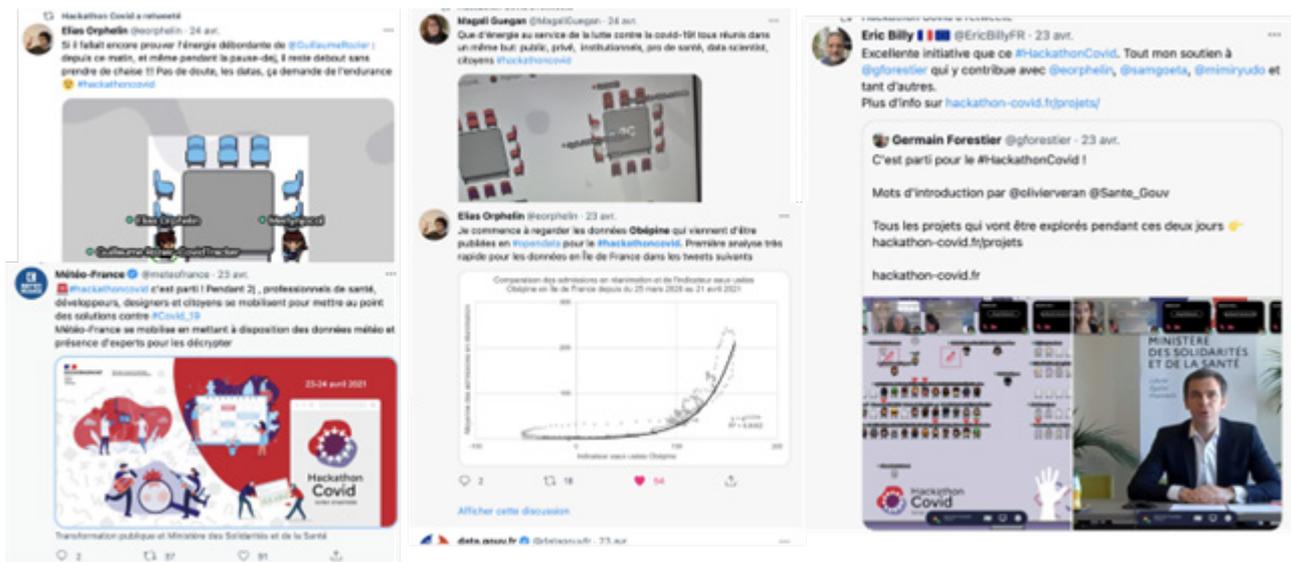


Figure 9 – Sélection de tweets publiés à l'occasion du hackathon Covid en mai 2021

Les projets des participants devaient répondre à un des quatre défis : mieux accompagner les victimes de la Covid à court et long terme, fluidifier la campagne de vaccination, suivre différemment l'évolution de la pandémie, mesurer les risques et favoriser les comportements vertueux. L'événement rassemblait de nombreuses administrations (Santé Publique France, Assurance Maladie, ARS Île-de-France...) et des organisations citoyennes (CovidTracker, Covidliste, collectif Du Côté de la Science...). Pendant deux jours dans un espace en ligne facilitant l'interaction entre les participants, les équipes ont fait émerger une dizaine de projets comme un mémorial des victimes de la Covid ou un outil d'analyse de l'effet des conditions

59. Hackathon Covid. Adresse : <https://hackathon-covid.fr/>

climatiques sur la propagation du virus. Certains projets ont été accompagnés par le ministère de la Santé par la suite comme un quizz sur les idées reçues liées à la vaccination qui a été relayé par le ministère de la Santé⁶⁰ ou un projet visant à faciliter l'échange de doses entre centres de vaccination qui a été soutenu dans son développement par une subvention visant à l'étendre potentiellement à d'autres produits sanitaires aux contraintes similaires.

Le hackathon a permis l'ouverture de nouvelles données comme celles d'OBEPINE qui, malgré l'appui du MESRI, n'étaient toujours pas publiées. Predict Services, une entreprise spécialisée dans la prévision des risques climatiques ayant Météo France pour actionnaire, a publié à l'occasion du hackathon l'Index PREDICT de Transmissivité Climatique de la Covid-19 (IPTCC). Fin mai 2021, l'Assurance Maladie a publié un portail intitulé Data vaccin Covid⁶¹ qui présente des données complémentaires de celles de Santé Publique France en indiquant les taux de vaccination par département de résidence et les taux de couverture vaccinale pour 50 pathologies graves et/ou chroniques⁶². À l'occasion du hackathon, avec Elias Orphelin de CovidTracker, nous avons formulé une demande d'ouverture des données sur le nombre de cas et d'hospitalisations selon le statut vaccinal des personnes. L'idée avait émergé pendant le hackathon que ces données puissent permettre d'évaluer l'impact de la vaccination sur les contaminations ou les hospitalisations, voire le nombre de vies « sauvées » par le vaccin. Ces données n'étaient pas collectées, elles nécessitaient de relier les systèmes VAC-SI, SI-DEP et SI-VIC qui n'avaient pas d'identifiants communs⁶³. De premiers résultats du croisement de ces bases ont été publiés en juillet par la Direction de la recherche, des études, de l'évaluation et des statistiques (DREES) du ministère de la Santé⁶⁴, puis ouverts sur le portail de la DREES⁶⁵. Les données indiquent à l'échelle nationale la part des personnes vaccinées ou non selon leur classe d'âge parmi les dépistages, hospitalisations et placements en soins intensifs pour infection au coronavirus. Elles ont donné lieu à la création par CovidTracker de VaxImpact, un site qui vise l'impact de la vaccination sur l'épidémie et calcule la réduction des risques d'hospitalisation du fait de la vaccination. Consécration de ses travaux, le 20 juillet 2021, Guillaume Rozier était décoré Chevalier de l'ordre national du Mérite par le président de la République. Il a dédié sa médaille aux nombreux contributeurs qui l'ont aidé à développer CovidTracker et les services associés comme ViteMaDose ou VaccinTracker.

7. Conclusion

Que retenir de cette rétrospective de l'ouverture des données sur la pandémie de Covid-19 en France ? Bien que j'aie proposé un cadrage sur la France, à plusieurs reprises j'ai présenté des initiatives internationales qui ont servi de modèles aux acteurs investis dans l'ouverture des données. On a pu aussi voir l'émergence d'acteurs comme l'université Johns Hopkins ou Our World In Data qui a récupéré et visualisé les données d'un grand nombre de pays pour les comparer.

60. Ministère de la santé (2021), « Ministère des Solidarités et de la Santé sur Twitter : "#COVID19 | Le #HackathonCovid organisé en avril a permis de faire émerger des projets citoyens pour lutter contre la pandémie. Le Vaccino Quiz en est un ! Il permet de tester ses connaissances & de s'informer sur la vaccination Lancez-vous 📱 », *Twitter*. Adresse : https://mobile.twitter.com/sante_gouv/status/1408681433359699975

61. Assurance Maladie (2021), « Accueil - Data vaccin Covid ». Adresse : <https://datavaccin-covid.ameli.fr/pages/home/>

62. Assurance Maladie, « Covid-19 : lancement de Data vaccin Covid, un site pour visualiser la couverture vaccinale en France | ameli.fr | Assuré ». Adresse : <https://www.ameli.fr/bouches-du-rhone/assure/actualites/covid-19-lancement-de-data-vaccin-covid-un-site-pour-visualiser-la-couverture-vaccinale-en-france>

63. Gouthière F. (2021), « Sur quoi se base Olivier Véran pour affirmer que 93% des patients Covid admis dans 30 grands hôpitaux de France ne sont pas vaccinés ? », *Libération*. Adresse : https://www.liberation.fr/checknews/sur-quoi-se-base-olivier-veran-pour-affirmer-que-93-des-patients-covid-admis-dans-30-grands-hopitaux-de-france-ne-sont-pas-vaccines-20210719_L7T6UG4UGVE4FDCRHLJZ7P5XRY

64. Drees (2021), « Les appariements SI-VIC, SI-DEP et VAC-SI | Direction de la recherche, des études, de l'évaluation et des statistiques ». Adresse : <https://drees-site-v2.cegedim.cloud/sources-outils-et-enquetes/les-appariements-si-vic-si-dep-et-vac-si>

65. « Covid-19 : résultats par âge issus des appariements entre SI-VIC, SI-DEP et VAC-SI », 2021, *DATA.DREES*. Adresse : <https://data.drees.solidarites-sante.gouv.fr/explore/dataset/covid-19-resultats-par-age-issus-des-appariements-entre-si-vic-si-dep-et-vac-si/information/>

Cet article est centré sur les données de veille sanitaire, mais nous aurions pu aussi insister sur le fait que la pandémie a aussi joué un rôle essentiel dans le développement en France et à l'international de la science ouverte, en particulier de l'ouverture des données de la recherche. Plus largement, nous avons pu voir comment l'ouverture des données liées à la pandémie a constitué un tournant dans la politique d'open data du gouvernement et aussi dans la diffusion du concept, devenu parfois un sujet d'intérêt pour des émissions de télévision généralistes à grande écoute. L'accès aux données sur la pandémie a mis en lumière les faiblesses de la mise en œuvre du droit d'accès aux documents administratifs et du principe d'ouverture des données par défaut prévu par la loi pour une République numérique. Le rapport du député Bothorel au Premier ministre s'est appuyé sur ces failles révélées par la pandémie pour relancer la politique d'ouverture des données de l'État qui semblait subir une forme d'essoufflement. Dans *Les Échos*, Éric Bothorel présentait même le Covid-19 comme un décideur politique à part entière : « *La pandémie est le meilleur secrétaire d'État au Numérique qu'on ait eu depuis quinze ans, c'est un formidable accélérateur pour la coopération entre les administrations et les secteurs d'activité.* » Les initiatives de la société civile comme OpenCovid19-fr ou CovidTracker (il y en aurait bien d'autres à citer) ont aussi renouvelé le plaidoyer de l'ouverture des données en faisant de l'open data un outil essentiel de réponse à la crise sanitaire. L'émergence de médiateurs de données a été vue comme un « bouleversement » ou un « accélérateur » de changements dans certaines administrations⁶⁶. Elle a pu aussi, à bas bruit, faire l'objet de controverses comme on a pu le voir dès le début de la première vague avec les prévisions de propagation de l'épidémie et les préconisations réalisées par Tomas Pueyo, un entrepreneur français dont les billets de blog sur le Covid-19 publiés sur Medium ont cumulé plus de 40 millions de fois et été traduits dans plus de quarante langues⁶⁷.

Si l'on s'intéresse maintenant plus directement aux données, cette rétrospective souligne l'importance des systèmes d'information dans la collecte et la circulation des données liées à la pandémie. Dans les premiers temps de la crise sanitaire, les autorités se sont appuyées sur des outils existants (SI-VIC, Sentinelles, réseau « 3labos »...) qui ont été complétés progressivement par des systèmes d'information dédiés au Covid-19 comme SI-DEP, VAC-SI ou le réseau OBEPINE. Là où le plan de prévention de pandémie grippale de 2009 édictait des principes et donnait une marche à suivre, les autorités ont dû développer de véritables infrastructures de données pour faire face à la situation inédite d'extension de l'épidémie sur tout le territoire. Derrière une grande partie des frictions (Edwards *et al.*, 2011) ou des retards dénoncés par certains acteurs et journalistes comme des « couacs » ou des « cafouillages⁶⁸ », se cachent certainement des situations complexes et inédites où la circulation des données n'allait pas de soi et demandait un travail important en amont. La question de la fiabilité des données est aussi revenue régulièrement, les producteurs comme les usagers n'hésitant pas à avertir sur les incertitudes ou les limites des données utilisées. Cette situation s'est présentée par exemple pendant la première vague quand les tests n'étaient pas disponibles en nombre suffisant et que les chiffres de dépistage rendaient compte d'une réalité très éloignée de la situation épidémique. La fermeture des laboratoires lors des jours fériés a aussi donné l'impression de baisses factices de la circulation du virus. Les acteurs ont appris progressivement à composer avec ces incertitudes ou à s'adapter à une situation mouvante, par exemple avec l'émergence des variants ou des vaccins qui ont changé la donne à plusieurs reprises.

Enfin, on peut s'interroger sur les effets de l'ouverture de ces données. Guillaume Rozier a

66. Godeluck S. (2021), « L'open data, une révolution silencieuse dans l'administration », *Les Echos*, n° 23488.

67. Martin N. (2021), « Coronavirus : une épidémie de faux articles scientifiques », *France Culture*. Adresse : <https://www.franceculture.fr/emissions/radiographies-du-coronavirus-la-chronique/radiographies-du-coronavirus-du-lundi-16-mars-2020>

68. Le Borgne B. (2021), « Urgence, cafouillages, "mille-feuille" d'indicateurs... Dans les coulisses des données très stratégiques sur le Covid-19 », *France Info*.

Adresse : https://www.francetvinfo.fr/sante/maladie/coronavirus/urgence-cafouillages-mille-feuille-d-indicateurs-dans-les-coulisses-des-donnees-tres-strategiques-sur-le-covid-19_4099757.html

répété à plusieurs reprises que l'open data était une « arme anti-complotisme » expliquant que « si les autorités disent "il y a 300 morts aujourd'hui", vous aurez toujours des gens qui diront "ils mentent, ce n'est pas vrai !" Les données publiques, détaillées par département, par jour, par âge, par sexe, permettent de vérifier ce qu'il en est.⁶⁹ » Une affirmation qui rejoint celle d'Amélie de Montchalin, la ministre de la Transformation et de la Fonction publiques, sur le site de *L'Opinion* : « Les gens pensent que les chiffres mis en ligne par CovidTracker constituent la vérité. Si c'est nous qui les donnons, on pense que l'on ment. » Ouvrir des données à un haut niveau de précision et permettre à quiconque de les analyser renforce-t-il la confiance dans les faits ? Les premiers travaux parus sur la pandémie ont alerté sur la pollution cognitive que l'explosion des indicateurs et des visualisations de données génère, phénomène décrit sous le terme d'*infodemics* (Grusd et al., 2021). D'autres ont montré que la circulation des données sur la pandémie au contraire alimente l'ère dite de « post-vérité » qui réduit l'influence des faits objectifs par rapport aux appels à l'émotion ou aux opinions personnelles (Shelton, 2020). Dans un article traduit dans cette revue, William Davies (2017) avait déjà alerté contre le ressentiment que pouvait créer le gouvernement par les nombres. La période que nous traversons, où une série d'indicateurs sanitaires déterminent notre quotidien, a probablement entretenu ce phénomène. L'ouverture des données semble être une condition nécessaire à la confiance, je doute toutefois qu'elle soit suffisante.

Toutes ces questions seront au cœur du projet de recherche DoMeSCO (Données et Médiation scientifique : leçons du Covid-19) pensé autour d'un partenariat entre le Centre de sociologie de l'innovation (Mines ParisTech, i3 – UMR 9217) et Dataactivist. Soutenu par l'ANR dans le cadre de l'appel à projets « Science avec et pour la société : Médiation et communication scientifiques », ce projet de recherche a pour objectif de structurer et de diffuser les enseignements qui peuvent être tirés de cette période afin d'outiller les actions de médiation futures, de guider les modalités de production et d'ouverture des données au sein de l'administration publique et de faciliter les collaborations entre les deux⁷⁰.

Références

Chignard S. (2021), « L'open data de crise : entre mobilisation citoyenne et communication gouvernementale », *Annales des Mines*, n° 14, p. 73-77.

Davies W. (2017), « Comment la statistique a perdu son pouvoir - et pourquoi nous devrions craindre ce qui va suivre », *Statistique et Société*, vol. 5, n° 1, pp. 11-20.

Dong E., H. Du et L. Gardner (2020), « An interactive web-based dashboard to track COVID-19 in real time », *The Lancet Infectious Diseases*, vol. 20, n° 5, pp. 533-534.

Edwards P. N., M. S. Mayernik, A. L. Batcheller, G. C. Bowker, et al. (2011), « Science friction: Data, metadata, and collaboration », *Social Studies of Science*, vol. 41, n° 5, pp. 667-690.

Goëta S. (2018), « "Données recherchent publics" : les politiques d'open data à l'épreuve de la réutilisation », in G. Gourgues et A. Mazeaud (éds.), *L'action publique saisie par ses « publics » : Gouvernement et (dés)ordre politique*, Villeneuve d'Ascq, Presses universitaires du Septentrion, pp. 137-154.

69. Floc'h B. (2021), « Covid-19 : l'enjeu crucial des données publiques », *Le Monde*. Adresse : https://www.lemonde.fr/planete/article/2021/03/06/covid-19-l-enjeu-crucial-des-donnees-publiques_6072158_3244.html

70. Un carnet de recherche sur la plateforme Hypotheses permet de suivre les actualités du projet : <https://domesco.hypotheses.org/>

Gruzd A., M. De Domenico, P. L. Sacco, and S. Briand (2021), « Studying the COVID-19 infodemic at scale », *Big Data & Society*, vol. 8, n° 1, <https://doi.org/10.1177/20539517211021115>

Ronai M. (2021), « La numérisation à marche forcée du système de santé face à la Covid-19 », *Annales des Mines*, n° 14, p. 52-61.

Shelton T. (2020), « A post-truth pandemic? », *Big Data & Society*, vol. 7, n° 2, <https://doi.org/10.1177/2053951720965612>

Strasser B. J. (2011), « The Experimenter's Museum GenBank, Natural History, and the Moral Economies of Biomedicine », *Isis*, vol. 102, n° 1, pp. 60-96.