

SANTÉ PUBLIQUE FRANCE FACE AU DÉFI DE L'OPEN DATA // SANTÉ PUBLIQUE FRANCE RISES TO THE CHALLENGE OF OPEN DATA

Étienne Lucas (etienne.lucas@santepubliquefrance.fr), Aurélie Fouquet, Delphine Jezewski-Serra, Jebraïel Ben Raies, Perrine de Crouy-Chanel, Caroline Alleaume

Santé publique France, Saint-Maurice

Soumis le 18.09.2023 // Date of submission: 09.18.2023

Résumé // Abstract

L'expression « open data » fait référence à un concept dans lequel des données sont mises à disposition de tous, sans restriction de copyright ni de brevet. Il est défini par la disponibilité, la réutilisation des données, et la participation universelle. En santé publique, il permet notamment l'information de tous sur la situation sanitaire, l'enrichissement de la recherche scientifique et l'appui à la gestion ; besoins qui ont été renforcés par la crise de la Covid-19. Cet article vise ainsi à présenter les outils utilisés par Santé publique France pour y répondre, les enjeux, les difficultés rencontrées et les perspectives envisagées.

Santé publique France produit de nombreux indicateurs sanitaires qu'elle met à disposition en toute transparence, ceci en répondant aux obligations réglementaires. Début 2019, l'observatoire cartographique Géodes s'ouvre en ligne avec 300 indicateurs épidémiologiques (900 en juin 2023), facilitant l'exploration et la récupération des données. Les premiers indicateurs liés à la Covid-19 y sont introduits en mars 2020, démultipliant la fréquentation du portail et incitant à recourir à d'autres outils comme la plateforme data.gouv d'Etalab (Direction interministérielle du numérique) puis le tableau de bord de l'agence, InfoCovidFrance. L'utilisation de ces outils pendant la pandémie a permis de placer la France parmi les pays ayant le plus facilité l'accès à leurs données, la classant troisième sur 41 pays évalués sur différents critères.

Le souci de transparence ne fait pas oublier l'impératif de garantir à l'échelle géographique la plus fine la confidentialité et la sécurité des données de santé. Ces éléments et les moyens nécessaires alimentent la réflexion actuelle sur la refonte de l'open data à Santé publique France.

Open data refers to a concept in which data are made available and accessible to all, without copyright or patent restrictions. It is defined by the availability, re-use and redistribution of data, as well as universal participation. In the field of public health, this represents a way of informing society as a whole about the health situation, enriching scientific research, and supporting management – needs that were made more acute by the COVID-19 crisis. The aim of this article is to present the tools used by Santé publique France, the French national public health agency, to meet these needs, as well as the issues at stake, the difficulties encountered and future opportunities under consideration.

Santé publique France produces numerous health indicators, which it makes available in a fully transparent manner while meeting regulatory obligations. In early 2019, the Géodes cartographic observatory was launched online with 300 epidemiological indicators (900 by June 2023), facilitating data exploration and extraction. The first COVID-19 indicators were introduced in March 2020, boosting portal traffic and encouraging the use of other tools such as Etalab's data.gouv platform (French Interministerial Digital Department) and the InfoCovidFrance dashboard. The provision of these tools during the pandemic placed France among the countries that most facilitated access to their data, ranking it third out of 41 countries evaluated on various criteria.

Commitment to transparency does not overlook the need to guarantee the confidentiality and security of health data related to a very fine geographical scale. These considerations, as well as the resources available, contribute to the current discussions on the development of open data at Santé publique France.

Mots-clés : Open data, Portail ouvert, Covid-19, Santé publique France
// Keywords: Open data, Open portal, COVID-19, French National Public Health Agency

Introduction

Dans le cadre de ses missions de surveillance et de prévention des risques sanitaires, Santé publique France, l'Agence nationale de santé publique, produit de nombreux indicateurs renseignant sur

l'état de santé des populations. L'augmentation du nombre d'indicateurs calculés, selon leur disponibilité à différents niveaux géographiques, ainsi que les évolutions en matière d'obligation réglementaire (directive Inspire¹) invitant à la mise à disposition plus large d'informations et de données dans

le champ de la santé ont conduit Santé publique France à se doter en 2019 d'un observatoire cartographique : Géodes⁽¹⁾.

Avec la pandémie de Covid-19 et l'amplification de la production d'indicateurs pour le suivi de l'épidémie, de nouvelles problématiques se sont imposées : réactivité, fréquence et finesse des informations produites, volumétrie, etc. La mise à disposition d'indicateurs sanitaires par Santé publique France a donc dû évoluer avec une mise à jour quotidienne, une diffusion à des échelles fines d'une information sensible, et l'accompagnement de tous les utilisateurs, qu'ils soient issus du grand public, des médias, de la sphère scientifique ou bien des professionnels de santé et de santé publique. L'open data devient en 2021 une partie intégrante de la structuration du programme de travail de Santé publique France via l'enjeu « Numérique en santé publique », un des six enjeux stratégiques réaffirmés en 2023.

Cet article vise à présenter les outils employés par Santé publique France pour répondre aux besoins préexistants et émergents durant la pandémie. Il met en évidence les contraintes et limites des solutions d'open data développées dans un contexte d'urgence en rapport avec l'organisation mise en place (figure). Enfin, il propose des perspectives pour améliorer la stratégie d'open data de Santé publique France afin de mieux satisfaire à l'avenir les attentes d'un public varié.

⁽¹⁾ <https://geodes.santepubliquefrance.fr/>

Matériel et Méthodes

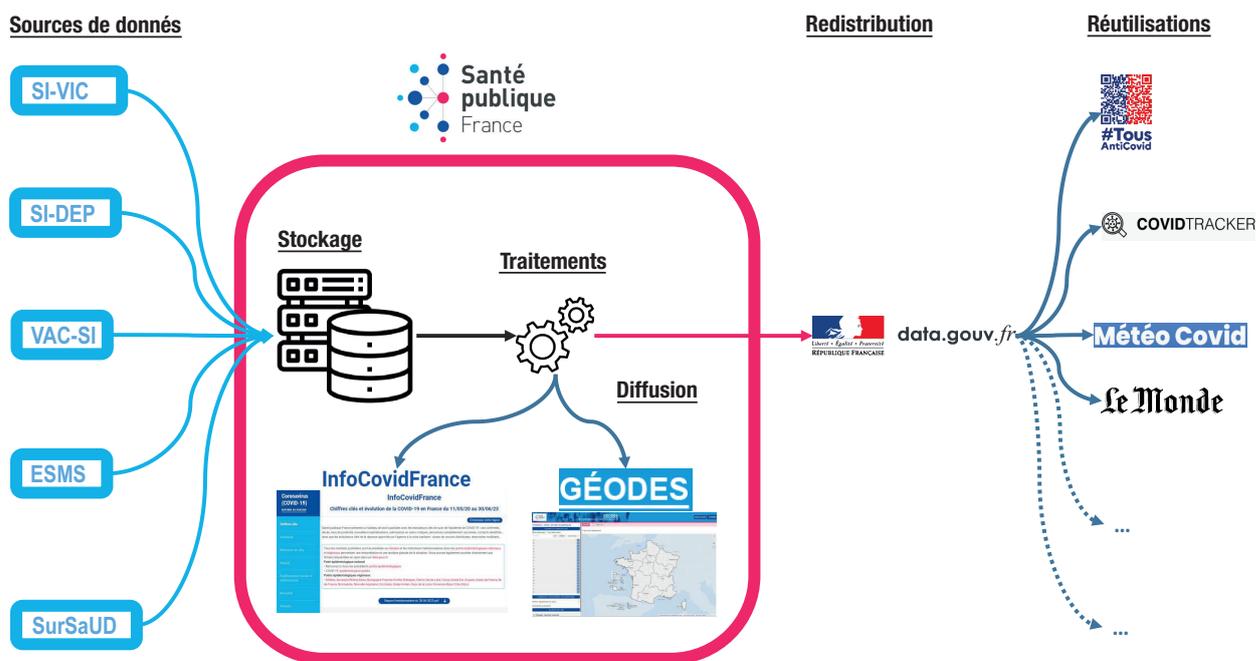
Définition

L'open data fait référence à un concept dans lequel les données sont mises à disposition de manière libre et accessible à tous, sans restriction. Dans ce cadre, les données peuvent être utilisées, réutilisées et redistribuées par n'importe qui, pour n'importe quel but. L'open data est définie par plusieurs principes^{2,3} :

- La disponibilité des données : celles-ci doivent être facilement accessibles par tous ;
- La réutilisation : elle favorise l'innovation, l'efficacité et la transparence en permettant aux individus et aux organisations d'exploiter pleinement le potentiel des données. Par réutilisation, on entend ici les possibilités et fonctionnalités qui vont permettre à l'utilisateur de réutiliser, hors de Géodes, les indicateurs qu'il a consultés, soit sous forme de graphiques, de cartes, ou même en téléchargeant les tables. En effet, « l'utilisation primaire des données de santé renvoie à la collecte de données à l'occasion de la prise en charge des patients, quand l'usage secondaire désigne l'utilisation de ces données pour d'autres finalités comme la recherche et l'innovation, le pilotage ou encore l'amélioration de la qualité des soins. Ainsi, la conception d'algorithmes pour mieux détecter des pathologies, la réalisation d'études pour mesurer les effets en vie réelle de médicaments ou encore l'élaboration de tableaux de bord pour piloter une crise sanitaire s'appuient sur des usages

Figure

Outils d'open data mis en place par Santé publique France pendant la pandémie de Covid-19



SI-VIC : Système d'information pour le suivi des victimes ; SI-DEP : Système d'information de dépistage populationnel ; VAC-SI : Système de suivi des vaccinations ; ESMS : Système de surveillance de la Covid-19 en établissements sociaux et médico-sociaux ; SurSaUD : Surveillance sanitaire des urgences et des décès ; Géodes : Géo données en santé publique.

secondaires des données de santé »⁴. Toutefois il n'est pas toujours garanti que les données obtenues permettent d'aboutir à des résultats probants, fiables, valorisables. En effet, le risque de mauvaise utilisation existe, qu'il relève du simple mésusage ou de l'utilisation à mauvais escient. C'est en rendant disponible une documentation la plus complète et la plus intelligible possible que l'on tend à minimiser ce risque sans pouvoir le supprimer totalement. Afin de tirer le meilleur parti des données, un travail considérable est donc nécessaire pour comprendre leur organisation, la façon dont elles sont collectées, les biais et les limites qu'elles présentent⁵.

- La redistribution : l'open data implique la possibilité de partager les données. Elle peut se faire par le biais de licences ouvertes et encourage ainsi le partage de connaissances et l'innovation. Le concept d'open data est intrinsèquement lié à celui de « *data innovation* »⁶ désignant le fait d'utiliser la donnée pour créer des projets et services à forte valeur ajoutée. Cela implique de mettre les données à disposition de toutes les parties prenantes qui peuvent les réutiliser. Ouvertes et partagées, elles permettent de créer de nouveaux usages innovants.
- La participation universelle : l'open data doit être accessible à tous, sans discrimination ni exclusion. Les données doivent être présentées dans un format compréhensible.

Cadre légal de l'open data

La loi pour une république numérique⁷, adoptée en octobre 2016, constitue un pilier important du cadre légal de l'open data en France. Cette loi introduit le principe de la « donnée d'intérêt général » et établit l'obligation pour les administrations publiques de publier certaines de leurs données dans des formats ouverts et réutilisables. Elle définit également les principes de réutilisation des données publiques, en encourageant la gratuité, la facilité d'accès et la non-discrimination.

La réglementation a été enrichie par d'autres textes et notamment par la directive européenne de 2019⁸ sur les données ouvertes qui a été transposée en droit français en 2021, renforçant ainsi les obligations des administrations publiques en matière de publication de données ouvertes.

Santé publique France, avec sa Direction appui, traitements et analyses des données (Data) créée en 2017, s'organise pour répondre à ces nouvelles obligations.

Résultats

2019 – Géodes

Au début de l'année 2019, Santé publique France inaugure l'observatoire cartographique Géodes (Géo données en santé publique)⁽¹⁾ développé à partir

de la solution Géoclip Air, débutant ainsi le processus de mise en open data des données. Il a été conçu dans le but de rendre accessibles et exploitables les données de santé de l'agence. Le nombre d'indicateurs disponibles est passé de 300 à son ouverture, à plus de 900 en 2023.

Il s'agit d'une plateforme en ligne regroupant et rendant accessible un large éventail d'informations permettant de caractériser l'état de santé des populations en France : des indicateurs de santé issus des données de surveillance, des enquêtes nationales, des bases de données hospitalières, des registres de santé, et des informations sur les déterminants de santé (facteurs qui déterminent l'état de santé des individus). Les thématiques couvertes sont celles de l'agence : maladies infectieuses, maladies chroniques, santé en lien avec l'environnement et santé au travail, comportements de santé, causes de décès, etc. L'objectif de Géodes est de fournir aux chercheurs, professionnels de santé, décideurs et au grand public des informations fiables et actualisées pour la compréhension et l'analyse de différents aspects de la santé.

Géodes offre également des fonctionnalités avancées pour faciliter l'exploration et l'utilisation des données. Les utilisateurs peuvent rechercher, filtrer et télécharger les données en fonction de leurs besoins spécifiques. La plateforme propose des outils de visualisation tels que des cartographies ou des diagrammes en barres pour présenter les données de manière claire et compréhensible, favorisant ainsi l'analyse et la diffusion des informations.

Depuis son ouverture au public, l'outil n'a cessé de s'enrichir de nouvelles thématiques ou des mises à jour de celles existantes. Santé publique France travaille en étroite collaboration avec d'autres acteurs du secteur de la santé : ministère de la Santé, Caisse nationale de l'assurance maladie (Cnam), laboratoires, agences régionales de santé (ARS), sociétés savantes, etc., pour garantir la qualité, la pertinence et la fiabilité des données disponibles sur la plateforme.

Dans un souci de réutilisation des données dans le domaine de la santé publique, les données disponibles sur la plateforme sont mises à disposition sous des licences ouvertes, permettant leur utilisation libre et leur réutilisation par des tiers. Cela contribue à améliorer la compréhension et la gestion de la santé publique.

2020 – L'impact de la Covid-19

La pandémie de Covid-19 a eu un impact significatif sur de nombreux aspects de notre société, y compris le développement de l'open data dans le domaine de la santé, en soulignant l'importance critique de disposer de données de santé fiables, les plus réactives possible et accessibles au grand public. Les informations précises et actualisées sur la propagation du virus, les taux d'infection, les mesures de prévention et les protocoles de traitement sont essentiels pour prendre des décisions

⁽¹⁾ <https://geodes.santepubliquefrance.fr/>

éclairées et mettre en place des politiques de santé efficaces. En conséquence, de nombreux gouvernements et organisations ont intensifié leurs efforts pour rendre les données de santé disponibles au public, encourager la transparence et promouvoir la collaboration entre les différentes parties prenantes.

Les premiers indicateurs Covid-19 ont été disponibles via Géodes en mars 2020. Cette ouverture a permis de garantir la transparence et la réutilisation par le public tout en tenant compte de l'enjeu de protection des données personnelles. Santé publique France a dû trouver l'équilibre dans son offre d'open data en rendant le maximum de données accessibles à l'échelle géographique la plus fine possible, tout en garantissant la non ré-identification des individus. Les indicateurs présentés à une maille géographique fine ne sont ainsi jamais représentés en effectifs et les taux sont uniquement présentés en classes. Il est de ce fait impossible de connaître le nombre de personnes incluses dans le calcul du taux et ainsi de ré-identifier une personne dans un territoire aussi restreint.

La fréquentation du portail, qui restait précédemment limitée et concernait un public averti (médecins, scientifiques, etc.), a alors augmenté de façon exponentielle (de 1 million de visiteurs en 2019 à 15 millions en 2020) et la demande du public s'est élargie. Géodes a vu le nombre d'indicateurs disponibles exploser : environ 200 indicateurs consacrés à la pandémie y sont actualisés quotidiennement entre 2020 et 2023. Toutefois, l'administration du portail s'est révélée contraignante en période de crise. Par ailleurs l'outil Géodes présentait des limites, notamment une navigation qui se complexifie au fur et à mesure que le nombre de thématiques augmente, des possibilités de représentations graphiques relativement réduites, une interface vieillissante et une ergonomie peu intuitive. Enfin, le format des données exportables était trop contraint pour la communauté des *data scientists*⁹ experts en sciences des données, et ne leur permettait pas d'alimenter directement et automatiquement leurs propres outils de visualisation en raison de l'absence d'API⁽²⁾.

La nécessité de passer par d'autres voies s'est imposée : la plateforme data.gouv⁽³⁾ est rapidement apparue comme incontournable, puis la mise à disposition des indicateurs sous forme de tableau de bord (InfoCovidFrance) s'est avérée utile pour compléter par d'autres formes de datavisualisation les informations diffusées sur l'observatoire cartographique Géodes.

data.gouv

Etalab est un département de la Direction interministérielle du numérique (Dinum), chargé de coordonner les efforts de l'administration française pour

rendre les données du secteur public accessibles à tous, dans un format ouvert et réutilisable. Dans cet objectif, Etalab a développé un portail permettant aux administrations, aux entreprises et aux citoyens de publier leurs données : data.gouv. Ce portail est le point focal de l'open data en France depuis sa création en 2011.

Avec l'appui d'Etalab, Santé publique France a ouvert sa page en 2021 sur data.gouv afin de mettre à disposition, de façon quotidienne, les données en lien avec la surveillance de la pandémie de Covid-19. Ce portail a trouvé une forte visibilité auprès de la communauté des *data scientists*, et a permis une excellente diffusion ainsi qu'une large et fréquente réutilisation de ces données.

La mise à disposition sur data.gouv a ainsi permis à de nombreux acteurs de valoriser la donnée à travers d'autres outils de restitution. On compte ainsi plus de 250 réutilisations des données de l'agence (144 réutilisations pour les données d'hospitalisations, 44 pour l'activité de dépistage, 52 pour l'activité SOS Médecins et des urgences, 46 pour le suivi des campagnes de dépistage, etc.). Ces réutilisations prennent plusieurs formes, comme des tableaux de bord, des API ou des data-visualisations. Elles sont produites par un éventail d'acteurs, qu'ils soient publics, privés ou citoyens. On peut citer : Covidtraker⁽⁴⁾, Météo-Covid⁽⁵⁾, les Décodeurs du Monde⁽⁶⁾, TousAntiCovid, le Tableau de bord du gouvernement, Architecture&Performance⁽⁷⁾.

L'utilisation de data.gouv par Santé publique France pour publier les données de surveillance de la Covid-19 a été efficace. La France a été remarquée pendant la pandémie comme l'un des pays ayant le plus facilité l'accès à ses données¹⁰. Elle est classée troisième sur 41 pays évalués par différents critères parmi lesquels : description des données, date de publication et rythme de mise à jour, forums et formulaires de contact, documentation et tutoriels, visualisation et statistiques, partage des données, etc. La France, l'Autriche et le Portugal sont les 3 pays qui apparaissent le plus fréquemment, pour chaque critère, parmi les 5 pays en tête de ce classement.

InfoCovidFrance

Ainsi, la crise sanitaire a vu fortement augmenter l'offre de tableaux de bord accessibles au public présentant une description de la situation sanitaire au prime des indicateurs d'infections, d'hospitalisation, de décès, ainsi que de vaccination liés à la pandémie. Dans un contexte de forte médiatisation, ces indicateurs sont rapidement devenus des éléments essentiels de communication des médias, des établissements, ainsi que des autorités de santé.

⁽²⁾ L'API ou Application Programming Interface (Interface de programmation d'application) est l'ensemble normalisé de classes, de méthodes, de fonctions, de constantes – des données – par lequel un logiciel offre des services à un autre logiciel qui les consomme pour alimenter ses propres fonctionnalités.

⁽³⁾ <https://www.data.gouv.fr/fr/organizations/sante-publique-france/>

⁽⁴⁾ <https://covidtracker.fr/>

⁽⁵⁾ <https://www.meteo-covid.com/>

⁽⁶⁾ https://www.lemonde.fr/les-decodeurs/article/2022/11/29/covid-19-les-chiffres-de-l-epidemie-en-france-et-dans-le-monde-en-cartes-et-en-graphiques_6038751_4355777.html

⁽⁷⁾ https://www.architecture-performance.fr/ap_blog/tests-incidence-hospitalisations-reanimations-et-deces-du-au-covid-en-france-par-departements-et-par-regions/

Les tableaux de bord Covid-19 destinés au grand public sont ainsi plus largement démocratisés que pour tout autre problème de santé¹¹. Si la mise à disposition des indicateurs de suivi de l'épidémie de Covid-19 a pu dès 2020 être réalisée via Géodes puis par data.gouv, permettant l'émergence de ces nouvelles plateformes de data-visualisation, Santé publique France développe en 2021 son propre tableau de bord. InfoCovidFrance⁽⁸⁾, accessible depuis le site Internet de l'agence, permet de publier pour la première fois de nouveaux indicateurs sur la pandémie issues du *contact-tracing*, sur l'activité de la Réserve sanitaire, de distribution des vaccins ou encore sur les résultats d'enquête sur l'adhésion aux comportements de prévention. Plus qu'un outil de valorisation de l'activité de Santé publique France, InfoCovidFrance vise à être pédagogique pour être accessible à un plus grand nombre. Les visuels varient entre des graphiques illustrant la donnée, des courbes de tendances, des formats chiffres clés permettant d'interpréter en un coup d'œil l'indicateur, et enfin des infographies illustrant simplement l'histoire racontée par la donnée¹². Ainsi le tableau de bord offre un outil très visuel, interactif, favorisant l'appropriation des informations, ainsi que des entrées à la fois thématiques, populationnelles et géographiques. Enfin, InfoCovidFrance propose également une version anglaise visant à augmenter la visibilité et à faciliter le partage d'information dans le cadre de collaborations internationales.

Accompagnement des utilisateurs

Dans un contexte de réutilisation des données publiées sur différents supports (dont les réseaux sociaux) et d'interprétation des résultats par un public de plus en plus expert, l'agence a très vite accompagné les utilisateurs dans la compréhension de la donnée.

Les données fournies par Santé publique France avaient pour rôle d'informer en totale transparence la population de l'évolution de l'épidémie. La quantité d'informations a soulevé de nombreuses interrogations et l'agence a dû répondre aux nombreuses sollicitations reçues par le biais des formulaires de contact ou tweets émanant de citoyens, de scientifiques ou de journalistes. Ce sont en tout près de 4 000 formulaires qui ont été remplis et envoyés via Géodes entre mars 2020 et juillet 2022, soit 4 par jour en moyenne. En comparaison, entre juillet 2022 et juillet 2023 nous avons reçu seulement 432 formulaires de contact.

Les outils de visualisation ont été un mode de consommation privilégié des données par le grand public. On peut ainsi penser aux différents graphiques produits par des data-journalistes qui sont facilement consultables sur les réseaux sociaux, comme sur Twitter (aujourd'hui X). La mise à disposition de la donnée

sur data.gouv et sa consommation sur les différents outils de restitution ont suscité un intérêt plus vif du grand public. Certains utilisateurs développaient une réelle expertise et se tournaient naturellement vers le producteur de la donnée (Santé publique France). Ainsi, l'onglet discussion de la page data.gouv a été fortement sollicité pendant cette période.

Pour les publications sur InfoCovidFrance, data.gouv et Géodes, chaque jeu de données et chaque indicateur étaient accompagnés par une description précise dans l'onglet dédié. Parfois, des notes méthodologiques étaient diffusées. C'est le cas pour la page SI-DEP, et sa fiche méthodologique téléchargeable¹³ pour préciser les changements qui ont eu lieu sur les indicateurs issus de la surveillance virologique à compter de mai 2022. Cette dernière a été téléchargée plus de 600 fois. Des moments d'accompagnement pédagogique étaient planifiés à l'attention des principaux utilisateurs (journalistes et *data scientists* en particulier) avant la publication de nouveaux indicateurs ou tout autre évolution de l'offre disponible. Ces moments fournissaient l'occasion de présenter les nouvelles publications, de répondre aux questions des utilisateurs, de prendre en considération leurs commentaires et remarques pour améliorer la qualité de la publication officielle.

Discussion

Optimisation des outils

La multiplication des plateformes open data complique la lisibilité pour les citoyens, et les différences méthodologiques de construction entre les indicateurs produits par les différents acteurs peuvent susciter des questionnements voire même de la défiance, comme cela a pu être constaté dans les messages reçus via les formulaires de contact Géodes. Forte de ce constat, Santé publique France s'organise pour engager une refonte de ses outils d'open data venant pallier les limites de Géodes en vue d'améliorer l'expérience utilisateur, de proposer une palette plus large de représentations graphiques, de renforcer la lisibilité et l'intelligibilité des informations mises à disposition et d'en faciliter l'administration.

Protection des données personnelles

Les données de santé sont extrêmement sensibles, contenant des informations personnelles et confidentielles. Lorsque ces données sont rendues accessibles au public dans le cadre de l'open data, il est impératif de garantir leur confidentialité et leur sécurité, notamment en empêchant la ré-identification par une anonymisation⁽⁹⁾ rigoureuse avant leur publication.

En assurant la protection des données de santé et en se préoccupant de leur qualité, l'open data

⁽⁸⁾ <https://www.santepubliquefrance.fr/dossiers/coronavirus-covid-19/coronavirus-chiffres-cles-et-evolution-de-la-covid-19-en-france-et-dans-le-monde>

⁽⁹⁾ L'anonymisation de données consiste à modifier le contenu ou la structure de ces données afin de rendre très difficile ou impossible la ré-identification des personnes ou des entités concernées.

est devenu une ressource précieuse pour la santé publique. Les chercheurs, les professionnels de santé et les décideurs politiques peuvent exploiter ces données pour améliorer les soins, identifier les tendances épidémiologiques et prendre des décisions de gestion et de prévention éclairées basées sur des preuves.

Data.gouv étant un point focal de l'open data en France, Santé publique France ne pourra pas, à l'avenir, se priver de sa visibilité. Une réflexion doit cependant être menée pour savoir ce qui doit y être diffusé. Les données Covid-19 devaient y être déposées, compte tenu du rythme quotidien de mise à jour et d'un fort intérêt pour ces données dans l'espace médiatique. La question est ouverte concernant les autres données produites par l'agence (issues des systèmes de surveillance ou d'enquêtes) et leur place sur un portail généraliste d'open data. La complexité de certains indicateurs et de leurs méthodes de construction pose également la question du public auquel l'open data s'adresse, et du niveau de littératie en santé des utilisateurs, pour les rendre accessibles et compréhensibles par le plus grand nombre.

Refonte du système pour une homogénéisation des pratiques

Sortie du contexte de crise de la pandémie, à moyens et ressources diminués, Santé publique France ne peut pertinemment pas gérer deux outils de diffusion d'indicateurs en open data en complément de la mise à disposition de données sur data.gouv. Une réflexion actuellement en cours va probablement mener vers une disparition des outils Géodes et InfoCovidFrance au profit d'un environnement rassemblant les avantages des deux outils avec une qualité de service équivalente, en différenciant deux profils d'utilisateurs. Un premier profil « grand public » pourra naviguer intuitivement pour visualiser les données et indicateurs directement dans l'application avec une documentation vulgarisée. Les experts préféreront consulter une documentation scientifique et interroger les bases de données via les API. L'objectif de cette refonte sera d'avoir à disposition un outil incluant un tableau de bord avec des chiffres clés sur les surveillances sanitaires dont l'actualité mérite une mise en avant (par exemple : la grippe en période hivernale, la surveillance des décès suite à des vagues de chaleurs, des résultats d'enquêtes...), des indicateurs couvrant l'ensemble des champs de surveillance de l'agence.

Conclusion

La nécessité urgente, induite par la pandémie de Covid-19, de partager des informations en masse et de collaborer à l'échelle mondiale a incité de nombreux pays et organisations à adopter des approches plus ouvertes en matière de données de santé. De nouvelles initiatives et plateformes ont vu le jour, permettant aux chercheurs et aux décideurs politiques d'accéder à des ensembles de données précieux pour mieux comprendre la propagation de

l'épidémie, évaluer l'efficacité des mesures de santé publique et développer de nouveaux traitements. En France la stratégie d'open data est impulsée au niveau national depuis les années 2010. Il existe des normes d'interopérabilité qui se sont structurées, permettant de bénéficier d'outils de mise à disposition des indicateurs produits par l'agence et d'accéder à des jeux de données d'autres institutions. La diversité des systèmes de santé à travers le monde et les variations dans la collecte et la normalisation des données ont rendu difficile la mise en place d'un cadre universel d'open data en santé dans le contexte d'urgence induit par la pandémie. Les différences de terminologie, de format et de méthodologie de collecte des données engendrent des défis lors de l'agrégation, de l'analyse et du croisement des informations issues de sources très diverses. Pour maximiser l'impact de l'open data dans le domaine de la santé et associer les données à l'action, il sera crucial de développer des normes et des protocoles communs pour la collecte, la mise en qualité, le traitement et le partage des données^{4,6}. ■

Remerciements

Nous remercions tous les acteurs de Santé publique France qui ont participé à la mise en place, la collecte, l'organisation des flux de données, aux tests, à la communication et la maintenance des différents outils d'open data : Johnny Platon, Benjamin Taisne, Édouard Chatignoux, Camille Pelat, Cécile Sommen, Céline Caserio-Schönemann, Jérôme Guillevic, Andrea Guajardo Villar, Yann Le Strat, Guillaume Spaccaterra, Alima Marie-Malikitte, Laëtitia Huiart, Stéphane Nardy, Adel Arfaoui, Paul-Henri Lampe, Sophie Mallejac, Minh-Canh Quan, Michel Slimane, Céline Zaccarini, Anne Robion, Marie Hamsany, Aurélien Dousseron, Anouk Tabai.

Nous remercions également nos partenaires et leurs équipes : EtaLab, TousAntiCovid, CirilGroup et DigDash.

Liens d'intérêts

Les auteurs déclarent ne pas avoir de liens d'intérêt au regard du contenu de l'article.

Références

- [1] Ministère du Partenariat avec les territoires et la Décentralisation. La directive européenne Inspire. 2019. <https://www.ecologie.gouv.fr/politiques-publiques/directive-europeenne-inspire>
- [2] James L. Defining Open Data. 2013. <https://blog.okfn.org/2013/10/03/defining-open-data/>
- [3] Commission nationale informatique et libertés. Guide pratique de la publication en ligne et de la réutilisation des données publiques (« open data »). Paris: Cnil; 2019. 28 p. <https://www.cada.fr/lacada/guide-pratique-de-la-publication-en-ligne-et-de-la-reutilisation-des-donnees-publiques>
- [4] Marchand-Arvier J, Allasonnière S, Hoang A, Jannot AS. Fédérer les acteurs de l'écosystème pour libérer l'utilisation secondaire des données de santé. Paris: ministère de la Santé et de la Prévention; 2023. 150 p. https://sante.gouv.fr/IMG/pdf/rapport_donnees_de_sante.pdf
- [5] Health Data Hub. Innover avec l'ensemble des acteurs. 2024. <https://www.health-data-hub.fr/innover-avec-lensemble-des-acteurs>
- [6] Ministère de la Santé et de la Prévention. 18. Développer la recherche en santé numérique et en particulier l'utilisation secondaire des données de santé. In: Feuille de route du

numérique en santé 2023-2027. Paris: ministère de la Santé et de la Prévention; 2023. p. 39. <https://sante.gouv.fr/actualites/actualites-du-ministere/article/lancement-de-la-feuille-de-route-du-numerique-en-sante-2023-2027>

[7] LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique. <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033202746/>

[8] Union européenne. Directive (UE) 2019/1024 du Parlement européen et du Conseil du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public (refonte). <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:32019L1024>

[9] Concepteurs d'avenir. Data scientist. 2024. <https://www.concepteursdavenir.fr/fiche-metier/data-scientist-0>

[10] Nikiforova A, McBride K. Open government data portal usability: A user-centred usability analysis of 41 open government data portals. *Telematics and Informatics*. 2021;(58):101539.

[11] Dasgupta N, Kapadia F. The future of the public health data dashboard. *Am J Public Health*. 2022;112(6):886-8.

[12] Chau D, Parra J, Santos MG, Bastías MJ, Kim R, Handley MA. Community engagement in the development of health-related data visualizations: A scoping review. *J Am Med Inform Assoc*. 2024;31(2):479-87.

[13] data.gouv. Données de laboratoires pour le dépistage (à compter du 18/05/2022) – SI-DEP. 2023. <https://www.data.gouv.fr/fr/datasets/donnees-de-laboratoires-pour-le-depistage-a-compter-du-18-05-2022-si-dep/>

Citer cet article

Lucas É, Fouquet A, Jezewski-Serra D, Ben Raies J, de Crouy-Chanel P, Alleaume C. Santé publique France face au défi de l'open data. *Bull Épidémiol Hebd*. 2024;(20-21):475-81. http://beh.santepubliquefrance.fr/beh/2024/20-21/2024_20-21_5.html

> ARTICLE // Article

SURVESMS : UN DISPOSITIF DE SURVEILLANCE DE LA COVID-19 EN ÉTABLISSEMENTS SOCIAUX ET MÉDICO-SOCIAUX

// SURVESMS: A SYSTEM FOR MONITORING COVID-19 IN LONG-TERM CARE FACILITIES

Côme Daniau¹ (come.daniau@santepubliquefrance.fr), Jérôme Naud¹, Sabrina Tessier², Éline Hassan³

¹ Santé publique France, Saint-Maurice

² Santé publique France – Bourgogne-Franche-Comté, Dijon

³ Santé publique France – Antilles, Gourbeyre

Soumis le 28.02.2024 // Date of submission: 02.28.2024

Résumé // Abstract

Lors de la phase de diffusion du virus SARS-CoV-2, un des enjeux majeurs de la surveillance épidémiologique a été de disposer de données actualisées et exhaustives sur les personnes atteintes de Covid-19 dans les établissements sociaux et médico-sociaux (ESMS). Cet article décrit la mise en place du dispositif de surveillance de la Covid-19 en ESMS (SurvESMS) et son adaptation pour répondre aux besoins des acteurs de la surveillance et des décideurs, ainsi qu'à l'évolution des pratiques face à l'acquisition progressive des connaissances sur la Covid-19. Le développement de ce dispositif de surveillance met en lumière la nécessité de se doter de plateformes informatiques performantes et agiles, permettant d'accéder aux données de santé des résidents et d'être en capacité de développer rapidement des systèmes de surveillance en situation d'urgence.

During the outbreak of the SARS-CoV-2 virus, one of the major challenges for epidemiological surveillance was to obtain up-to-date, exhaustive data on people with COVID-19 in long-term care facilities. This article describes the implementation of the COVID-19 surveillance system in long-term care facilities (SurvESMS). It also describes how the system was adapted to meet the needs of people involved in surveillance and decision-makers, as well as the changes in practices resulting from the gradual acquisition of knowledge about COVID-19. The development of this surveillance system highlights new requirements for agile IT platforms that provide access to residents' health data and that can be used to develop surveillance systems rapidly in the event of an outbreak.

Mots-clés : Dispositif de surveillance, Covid-19, Établissements sociaux et médico-sociaux, Solution informatique
// Keywords: Surveillance system, COVID-19, Residential care, IT solution