

L'*open data* de crise : entre mobilisation citoyenne et communication gouvernementale

Par **Simon CHIGNARD**

Expert de la valorisation des données

Le tremblement de terre qui a frappé Haïti le 12 janvier 2010 a été qualifié à l'époque par les Nations unies de « pire crise humanitaire liée à un désastre naturel » ; 230 000 personnes y ont perdu la vie, et plus d'1,3 million d'habitants de l'île se sont retrouvés sans abri. Les infrastructures du pays ont été durement touchées, ce qui a rendu encore plus difficile l'accès des secours aux zones les plus reculées. Mais Haïti représente aussi, pour nombre d'observateurs de l'*open data*, un cas d'école de mobilisation et d'utilisation des données comme facteur de résilience. En effet, en quelques jours et partout dans le monde, des centaines de bénévoles se sont mobilisés, notamment pour cartographier les zones sinistrées. Les images satellites mises à disposition ont été utilisées pour créer des plans détaillés, permettant ainsi aux équipes de secours de disposer d'une information actualisée.

L'équipe Humanitarian OpenStreetMap Team (HOT), constituée en association dès le mois d'août 2010, s'est particulièrement illustrée dans l'usage de la cartographie libre OSM (OpenStreetMap) comme réponse aux situations de crise et comme facteur de résilience. Depuis, HOT est intervenue sur la plupart des crises humanitaires de la dernière décennie en lien ou non avec des catastrophes naturelles (tsunamis, tremblements de terre, inondations), et s'est développée autour d'autres thématiques du développement durable.

La crise haïtienne a permis de démontrer, si besoin était, que l'accès à une information fiable et à jour est l'un des éléments de réponse à des situations d'urgence. Elle a aussi souligné, en creux, que les États ne sont pas nécessairement les mieux placés – ni les mieux dotés – pour fournir l'information. Enfin, que le numérique porte aussi en lui des capacités de mobilisation des foules ou, pour le moins, d'une armée de bénévoles qui savent collecter, traiter et exploiter des données. Mais la crise déclenchée par le tremblement de terre de 2010 a une particularité : elle s'est déroulée dans une unité de lieu (Haïti).

La crise sanitaire provoquée par la Covid-19 a une autre particularité, qui la distingue et la singularise parmi les crises de ces dernières décennies : son ampleur. L'épidémie s'est étendue en quelques mois à l'ensemble de la planète. Quasiment aucun continent n'a été épargné. Il s'agit bien en cela d'un évènement massif et global, qui touche indifféremment des pays où les infrastructures, notamment informationnelles, sont développées, et d'autres où elles manquent. Nous visons dans cet article à comprendre comment la donnée, et en particulier la donnée ouverte (*open data*), a pu être mobilisée comme un facteur de résilience, tant par les gouvernements et la communauté scientifique que par la société civile. En effet, cette crise illustre une nouvelle fois la capacité de mobilisation de ceux qui, tout en n'étant ni médecins, ni décideurs politiques, ont toutefois jugé qu'ils avaient un rôle à jouer pour faire face au coronavirus. Entre mobilisation citoyenne et communication gouvernementale, l'*open data* a de multiples facettes, parfois en tension.

Mobilisation générale

Le 11 mars 2020, à Genève, le directeur général de l'Organisation mondiale de la santé (OMS) vient de qualifier la Covid-19 de « pandémie ». Les Français ne le savent pas encore, mais ils vont bientôt faire connaissance avec un terme qui accompagnera encore leur vie quotidienne un an plus tard : le confinement. En Italie, le gouvernement a déjà mis en place des mesures en vue de lutter contre l'épidémie.

À cette époque, depuis quelques semaines déjà, les autorités publiques françaises diffusent des informations sur la progression de l'épidémie dans notre pays. Mais l'information est très parcellaire et difficilement exploitable. Santé publique France et les agences régionales de santé communiquent lors de points presse, publient parfois des communiqués de presse, au mieux des documents au format PDF.

En résumé : rien que l'on ne puisse qualifier d'*open data*, au sens de « donnée librement réutilisable ». Aucune donnée officielle n'est à l'époque non plus publiée sur la plateforme nationale « data.gouv.fr ».

Face à ce constat, des individus, bientôt regroupés en communautés, se mettent au travail dès la mi-mars 2020. Leur ambition : compiler et mettre à disposition l'ensemble des données éparées sur la situation sanitaire en France. La tâche n'est pas simple, compte tenu de la très grande diversité des sources de données. Certains d'entre eux en sont rendus à regarder les vidéos des points presse du ministre chargé de la Santé ou du directeur général de la Santé pour y glaner les informations actualisées, puis les transformer en données.

Il faut s'arrêter un instant sur le rôle des outils numériques qui, comme lors de la crise en Haïti, ont permis la mobilisation et la coordination entre les initiatives, alors même que le pays était confiné.

Le premier outil utilisé est la messagerie de groupe Slack, plébiscitée par les communautés de développeurs. Slack permet de créer des espaces d'échange *via* des canaux de discussion (*channel*) dédiés à une thématique. Un exemple parmi d'autres, qui illustre l'ampleur de la mobilisation : "Data Against Covid-19", présenté comme un « espace d'échange autour de la collecte et de l'analyse des données, de leur représentation et de la compréhension des enjeux liés à la donnée pour les personnels soignants et la société civile ». Ce sont près de 90 000 messages qui seront échangés *via* ce canal thématique, attirant près de 1 200 membres. Qui sont-ils ? Il est difficile de se faire une idée précise. On peut cependant noter que, pour une partie d'entre eux, ils travaillent pour des structures publiques, à l'image de Lior Perez, responsable des développements informatiques à Météo-France, ou encore Bastien Guerry et Jérôme Desboeufs, respectivement agent et prestataire d'Etalab, la structure interministérielle en charge de l'ouverture des données publiques.

Le second outil majeur est GitHub, service d'hébergement et de gestion de développement de logiciels. La page de dépôt de fichiers intitulée "OpenCovid19-fr / Data" est ainsi utilisée pour consolider les données de sources officielles françaises concernant l'épidémie. Une cinquantaine d'individus – parmi lesquels on retrouve là encore plusieurs personnes gravitant autour d'Etalab – y contribuent. Ils ne se contentent pas de récupérer des données, ils les consolident et les rediffusent, notamment *via* un site de visualisation « veille-coronavirus.fr » qui propose, entre autres, un tableau de bord de l'épidémie. Cette proximité avec Etalab se révélera être un atout : ce qui était initialement une initiative d'une mobilisation citoyenne va progressivement être repris pour devenir « le » tableau de bord officiel, publié sur « data.gouv.fr » puis intégré sur le site « gouvernement.fr », géré par le service d'information du gouvernement (SIG).

Au même moment, Santé publique France et Etalab se mettent en ordre de marche pour diffuser des données les plus complètes possibles. Après les données sur les hospitalisations et les

contaminations viendront celles sur les tests de dépistage, et, enfin, les données concernant la campagne de vaccination. Mais l'*open data* ne se limite pas à la sphère sanitaire : le site « data.gouv.fr » accueille aussi des données sur les aides en soutien à l'activité économique (dispositif d'activité partielle, fonds de solidarité). L'objectif étant ici de donner à voir l'action gouvernementale en réponse à la crise.

Ailleurs dans le monde

Dès le mois d'avril 2020, l'OMS appelle les gouvernements à faire de l'*open data* l'un des leviers de leurs actions dans la lutte contre le coronavirus. De nombreuses organisations internationales ont entrepris depuis un recensement des initiatives et des expérimentations dans le cadre de la crise sanitaire. Ainsi, l'Observatoire de l'innovation dans le secteur public de l'OCDE a recensé plus de 450 initiatives au niveau mondial⁽¹⁾, dont une moitié liées aux données (collecte, production et exploitation, réalisation de tableaux de bord). Ce recensement met en lumière la très grande diversité des données mises à disposition, mais aussi les différences de maturité en matière de numérique, voire de transparence. Ainsi, à Taiwan et en Corée du Sud, les autorités ont mis à disposition une API (interface de programmation) donnant accès en temps réel aux stocks de masques disponibles dans les pharmacies. Ces API, accessibles librement et gratuitement, ont permis l'intégration de ces données dans de très nombreuses applications. Le contraste est saisissant avec d'autres pays développés où l'information sur la disponibilité des masques était plus que parcellaire, voire absente.

Une étude, réalisée conjointement par la Data Unit de l'OCDE et le Gov Lab (de l'Université de New York), a permis d'étudier en particulier le rôle de l'*open data* dans la première phase de la pandémie, c'est-à-dire de mars à juillet 2020 : 85 projets ont ainsi été analysés en détail. Les tableaux de bord et la visualisation de données représentent ensemble plus de la moitié de ces initiatives, suivis par les dépôts de données et notamment les portails *open data* (37 %). La prépondérance des tableaux de bord illustre leur importance dans les stratégies de communication des gouvernements, tant nationaux que locaux. En ce sens, les visualisations de données participent de la communication officielle autour du virus, avant d'être pensées comme des outils permettant l'action, tant au niveau des gouvernements que de la société civile elle-même.

Entre transparence et redevabilité, la crise sanitaire comme révélateur des tensions de l'*open data*

La crise sanitaire a aussi révélé et illustré les tensions qui entourent la notion d'*open data*. Dans un article remarqué de 2012, Harlan Yu et David Robinson (à l'époque chercheurs dans les universités américaines de Stanford et de Georgetown) pointaient déjà les ambiguïtés de l'*open data*⁽²⁾.

Depuis les années 1950, expliquaient-ils, la notion de « gouvernement ouvert » (*open government*) était associée à l'émergence du droit d'accès à l'information. Selon eux, l'*open data* illustre la dérive de ce concept initial de « gouvernement ouvert » vers une approche plus technologique. Mettre des données à disposition (les ouvrir) est à la portée de n'importe quel gouvernement, mais cela suffit-il à rendre ces gouvernements redevables ? « Ainsi, les “données gouvernementales ouvertes” peuvent faire référence à des données qui rendent le gouvernement dans son ensemble plus ouvert (c'est-à-dire plus responsable devant le public), mais peuvent également faire référence

(1) OPSI Covid-19 Innovative Response Tracker, <https://oecd-opsi.org/covid-response/>

(2) YU H. & ROBINSON D. (2021), “The new ambiguity of ‘open government’”, *UCLA Law Review*, 178, pp. 81-91.

à des divulgations politiquement neutres du secteur public qui sont faciles à réutiliser, mais qui n'ont peut-être rien à voir avec la responsabilité publique. Aujourd'hui, un régime peut se dire "ouvert" s'il construit le bon type de site *web* – même s'il ne devient pas plus redevable », exposent Yu et Robinson.

Cette ambiguïté est parfaitement illustrée par la crise sanitaire de la Covid-19. Les gouvernements – en Europe et ailleurs – revendiquent la publication des données comme une preuve de la transparence de leurs actions dans la lutte contre la crise sanitaire. En France, le gouvernement, tant par la voix du Premier ministre ou celle du ministre chargé de la Santé, mentionne régulièrement les données publiées par Santé publique France sur le portail « data.gouv.fr ». L'audience du portail atteint d'ailleurs des records : au mois de novembre 2020, la plateforme a accueilli plus de 2,3 millions de visiteurs uniques, c'est-à-dire la plus forte audience depuis son lancement en 2011.

Cet effort de transparence sur les données de la Covid-19 signifie-t-il pour autant que l'action publique est redevable ? Peut-on considérer que l'*open data* permet à « la Société [...] de demander compte à tout agent public de son administration » comme le mentionne dans son article 15 la Déclaration des droits de l'homme et du citoyen de 1789 ? Les questions restent ouvertes.

Ce qui est acquis, c'est que la transparence revendiquée n'empêche pas la défiance, bien au contraire. Plusieurs sondages d'opinion⁽³⁾ indiquent que près de 60 % des Français ne font pas confiance au gouvernement dans sa gestion de la crise sanitaire. Pire, selon une enquête d'opinion de février 2021, 73 % des personnes interrogées estiment que « le gouvernement ne dit pas la vérité aux Français »⁽⁴⁾. De toute évidence, le lien entre transparence et confiance est sûrement plus complexe qu'une simple relation de cause à effet, comme l'illustre de nombreux travaux de recherche (notamment en sciences politiques).

Les leçons de la crise sanitaire

À l'heure où nous écrivons ces lignes (mars 2021), la crise sanitaire est encore une réalité. Il est cependant déjà possible d'identifier quelques enseignements de la mise en œuvre de *open data* pour faire face à l'épidémie.

La première leçon est la capacité de mobilisation en période de crise : la société civile comme les acteurs publics (au niveau local et national) se sont mobilisés pour publier des données, parfois dans des conditions d'urgence et de pression importante. Il est remarquable de constater l'écart entre la situation de départ en mars 2020, telle que décrite au début de cet article, et la situation un an plus tard. L'offre, en matière de données, est incommensurablement plus grande qu'elle ne l'était au début de la crise sanitaire. Le portail data.gouv.fr met ainsi en avant plus de 70 jeux de données issus de divers producteurs publics tant nationaux (Santé publique France, ministère de la Santé et des Solidarités, direction générale des entreprises, Insee, Etalab, URSSAF...) que locaux (villes de Nantes, Marseille, Grenoble, Issy-les-Moulineaux...).

La seconde leçon est précisément liée au rôle de la société civile, et à l'émergence de nouveaux médiateurs. Les communautés qui se sont mobilisées pour collecter et analyser les données ont sans aucun doute joué un rôle de déclencheur, et par la suite d'aiguillon de l'ouverture des données par les gouvernements. Nous avons vu aussi l'importance de nouveaux médiateurs, comme la plateforme « CovidTracker », dont l'un des fondateurs, le *data scientist* Guillaume Rozier a acquis une grande visibilité dans les médias. Les visualisations de « CovidTracker » sont même

(3) Par exemple, l'enquête Ifop pour le JDD du 3 mars 2021 et l'enquête Odoxa pour *Le Figaro* et France Info du 4 février 2021.

(4) Sondage Odoxa pour *Le Figaro* et France Info, 4 février 2021.

parfois reprises dans les actions de communication officielle du gouvernement. Ces nouveaux médiateurs présentent aussi l'intérêt d'être perçus comme plus crédibles que la communication gouvernementale – même si, *in fine*, ils partagent la même source de données (Santé publique France).

La troisième leçon est que le plein potentiel de l'*open data* n'est pas encore atteint. Des progrès restent à faire sur l'utilisation et l'exploitation des données, en particulier pour appuyer les décisions publiques. Il doit être possible de développer bien mieux que des tableaux de bord, mais aussi une véritable culture de la décision publique, qui soit factuelle, basée sur les données (*data-driven*) et surtout transparente. On notera d'ailleurs avec grand intérêt les recommandations de la mission confiée par le Premier ministre au député Éric Bothorel, présentées dans son rapport publié en décembre 2020. Elles invitent notamment à l'ouverture des modèles épidémiologiques de l'Institut Pasteur utilisés par le conseil scientifique. Un pas de plus vers une action publique plus transparente ?