

Le *making of* du suivi de la crise du Covid-19 au *Monde*

Par Les Décodeurs du *Monde*

Journal *Le Monde*

En quelques semaines, une « pneumonie atypique » apparue en Chine est devenue une pandémie mondiale, bouleversant nos vies. Comment rendre compte de l'ampleur, des effets et de la propagation de cette maladie qui allait devenir le Covid-19 ? Depuis le début de la pandémie, le service des Décodeurs du journal *Le Monde* a tenté de suivre au jour le jour l'étendue de cette urgence de santé publique mondiale.

Raconter comment nous avons tenté de rendre compte de l'étendue de l'urgence sanitaire depuis les premiers cas – recensés de manière artisanale –, c'est raconter une histoire de l'épidémie et de l'impréparation des agences sanitaires en matière de données ouvertes. De fait, ces dernières eurent un intérêt capital pour nos lecteurs, puisque c'est en fonction de divers indicateurs – contaminations, R0, hospitalisations, décès, etc. – que se sont décidées des politiques sanitaires qui allaient avoir un impact majeur sur nos vies.

C'est au moment où la France se remettait doucement d'une longue grève contre la réforme des retraites que la nouvelle d'une « pneumonie atypique » apparue en Chine est arrivée jusqu'à nous. Information presque insignifiante en premier lieu, comme il en existe tant dans le flux de l'actualité. Et pourtant. Cette maladie inconnue est devenue en quelques semaines une préoccupation majeure, un sujet total dans lequel nous étions plongés tant comme journalistes que citoyens, d'une ampleur telle qu'elle a bouleversé chaque pan de nos vies et de nos sociétés. En 2020 et 2021, le Covid-19 a été le thème quasi unique de notre travail aux Décodeurs.

En temps normal, notre mission consiste à vérifier les informations qui pourraient être fausses, à expliquer l'actualité, à manier la donnée (publique ou non), à enquêter, à essayer de nouvelles manières d'écrire des articles et de raconter l'actualité. Ce service du journal *Le Monde* créé en 2014 est composé d'une vingtaine de personnes qui ont un profil rare : tous sont journalistes ; la plupart ont également des compétences ou une formation techniques dans le développement informatique, les bases de données, la cartographie, etc.

En sus des articles de vérification et d'explication, ces compétences juxtaposées nous ont amenés à nous intéresser à la question de ce que l'on appelle improprement les « chiffres » d'une épidémie qui allait devenir pandémie mondiale. Nul besoin de refaire la chronologie de cette période dont chacun se souvient – d'autant plus qu'hélas, elle n'est pas close et ne le sera peut-être pas avant des lustres. Mais l'histoire des données liées au Covid-19 raconte en creux la chronologie d'une maladie dont nous avons découvert en temps réel l'ampleur (contaminations), les effets (hospitalisation, mortalité), la propagation géographique et la propension à se diffuser par vagues successives.

PREMIERS SUIVIS DE CE QUI N'ÉTAIT ALORS QU'UNE MALADIE



Figure 1. Le premier tableau de suivi du Covid-19 du *Monde*, publié le 24 janvier 2020, proposait une synthèse des décès dus au coronavirus recensés dans le monde (Sources : OMS, ministère de la Santé de Singapour, NGC/Chine).

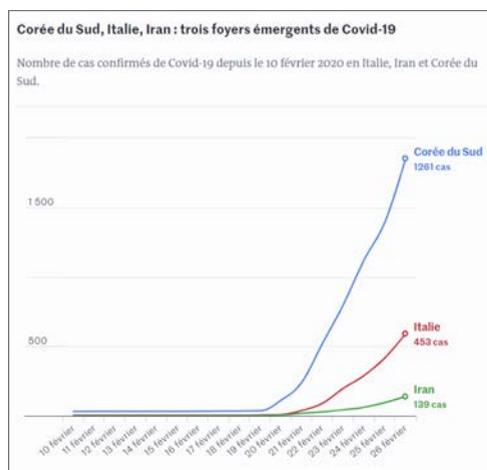


Figure 2. En dehors de la carte, le premier tableau de bord, publié le 27 février 2020, contenait un graphique qui ne suivait « que » les foyers actifs de la maladie, lesquels étaient particulièrement surveillés (Source : Université John Hopkins CSSE).

Aux Décodeurs, tout a commencé le 24 janvier 2020 avec la publication d'un simple tableau (voir la Figure 1 ci-contre), renseigné à la main, qui recensait les quelques centaines de cas déclarés en Chine et les premières contaminations hors du pays. Lorsque nous l'avons publié, nous nous sommes demandé quel serait son destin. Soit cette maladie émergente demeurait à une ampleur limitée comme le fut le Sars-CoV-1 vingt ans plus tôt ; soit le nombre de cas explosait et ce décompte artisanal s'avèrerait rapidement dépassé.

Dans les semaines qui ont suivi, le monde, petit à petit, a pris la mesure de l'ampleur de la crise sanitaire, justifiant son confinement progressif par la diffusion des cas. Comment rendre compte à nos lecteurs de l'avancée d'un virus se jouant des frontières ?

Nous avons publié le 27 février 2020 un premier tableau de bord¹ mettant l'accent sur les cas confirmés des trois foyers actifs hors de Chine : la Corée du Sud, l'Iran, et l'Italie. « La France », écrivions-nous alors, « n'est pour le moment pas considérée comme un foyer épidémique, mais cette situation pourrait changer. »

Le graphique (voir la Figure 2 ci-contre) et le planisphère (voir la Figure 3 page suivante) de cet article (actualisé quotidiennement) étaient basés sur le travail de mise à jour de l'Université Johns Hopkins (JHU, États-Unis), qui était la seule institution proposant des données ouvertes actualisées chaque jour, bien avant que l'Organisation mondiale de la santé ne publie son propre tableau de bord. Ce système artisanal nous a permis dans un premier temps d'automatiser la mise à jour à un rythme important. Cette masse de données est rapidement devenue ingérable : le tableur source n'a cessé de s'agrandir et est devenu un vrai château de cartes, fondé sur une ribambelle de formules, lesquelles

¹ https://web.archive.org/web/20200227235003/https://www.lemonde.fr/les-decodeurs/article/2020/02/27/en-carte-visualisez-la-propagation-mondiale-de-l-epidemie-de-coronavirus_6031092_4355770.html

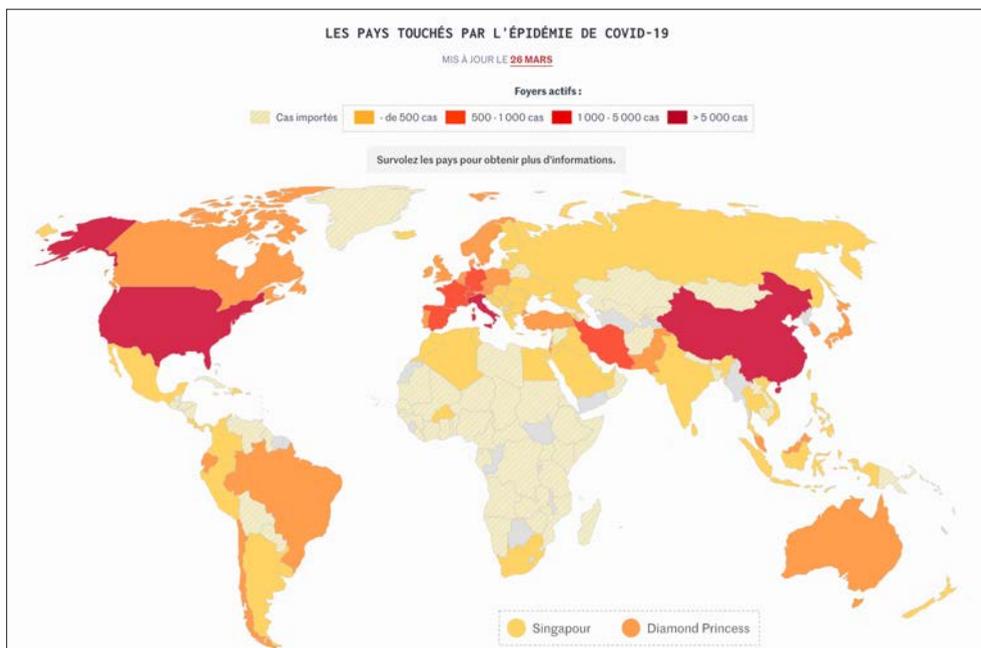


Figure 3. Le premier vrai tableau de bord de suivi du Covid-19, publié le 26 mars 2020, comportait en premier lieu une carte du monde des contaminations, en différenciant à l'époque les cas dits « importés » des transmissions intra-nationales (Source : Université John Hopkins CSSE).

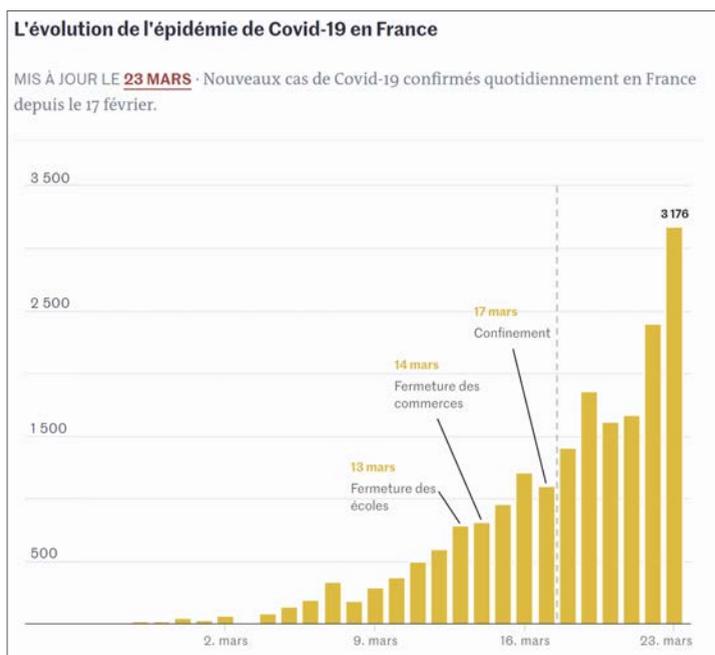


Figure 4. Le premier graphique de suivi des cas en France, publié le 23 mars 2020, d'après les annonces quotidiennes de Jérôme Salomon, le directeur général de la santé, à la presse tous les soirs. Le graphique est annoté des quelques événements qui ont jalonné le mois de mars 2020.

pouvaient faire tomber l'intégralité du flux de données en cas d'irrégularités. D'autant plus que la création et la maintenance n'étaient opérées que par une seule personne.

Entre février et fin mars 2020, les autorités françaises ont rencontré beaucoup de difficultés² pour communiquer les données dont elles disposaient. Pendant les premières semaines, le nombre de cas, d'hospitalisations ou de décès étaient communiqués à l'oral vers 19 heures lors du point presse de Jérôme Salomon et saisies à la main par la rédaction du Monde.fr. Il a fallu attendre la fin mars, soit deux semaines après le début du confinement, et le zèle d'Etalab, le bras armé numérique de Matignon, pour disposer des données collectées par Santé publique France (SPF).

Un premier jeu de données hospitalières³ déposé sur Data.gouv.fr le 27 mars 2020 nous a ainsi permis de publier début avril, en amont du tableau de bord principal, un article détaillant les hospitalisations et décès par département⁴. Ces données locales ont attiré de nombreux lecteurs.

Le même jour, Etalab a mis en ligne le premier “*dashboard*”⁵ du gouvernement sur le Covid-19 et les données brutes sur Data.gouv.fr. Pour la première fois, les données de SPF sont enfin accessibles en *open data* et peuvent rendre compte, dans notre tableau de bord, de l'évolution de l'épidémie en France sans plus passer par les données de JHU, mais bien par des sources nationales et officielles.

Face à l'émergence de ces nouveaux indicateurs et au succès rencontré par le tableau de bord, il a été décidé mi-avril de construire une page de suivi plus complète, plus réfléchie et plus ambitieuse, aussi bien techniquement qu'éditorialement.

CRÉER UN REPÈRE DURABLE POUR LES LECTEURS

Pour ce nouveau tableau de bord, l'objectif a été de proposer un objet durable aux lecteurs, ce qui impliquait plusieurs choses : il fallait qu'il soit aussi facile que possible à maintenir, que le flux de données soit plus robustement géré, et que son design soit pérenne et utilise des codes visuels adaptés au suivi d'une crise sanitaire.

Les flux ont été gérés par des scripts automatisés qui aspiraient directement les sources officielles. Ils sont accessibles *via* des URL d'un sous-domaine du Monde.fr capable de supporter un trafic important.

Figure 5. Reprenant ce que proposait le premier tableau de bord, le second tableau de bord publié le 6 mai 2020 s'ouvrait par une synthèse de la situation mélangeant texte et graphiques mis à jour automatiquement (Source : Université John Hopkins CSSE).



² https://www.lemonde.fr/les-decodeurs/article/2020/04/15/covid-19-le-sursaut-des-administrations-pour-fournir-des-donnees-chiffrees_6036687_4355770.html

³ <https://www.data.gouv.fr/fr/datasets/donnees-hospitalieres-relatives-a-lepidemie-de-covid-19/>

⁴ https://www.lemonde.fr/les-decodeurs/article/2020/04/01/coronavirus-visualisez-le-nombre-de-personnes-hospitalisees-departement-par-departement_6035199_4355770.html

⁵ <https://web.archive.org/web/20200328082306/https://dashboard.covid19.data.gouv.fr/>

Quelles données pour suivre l'épidémie ?

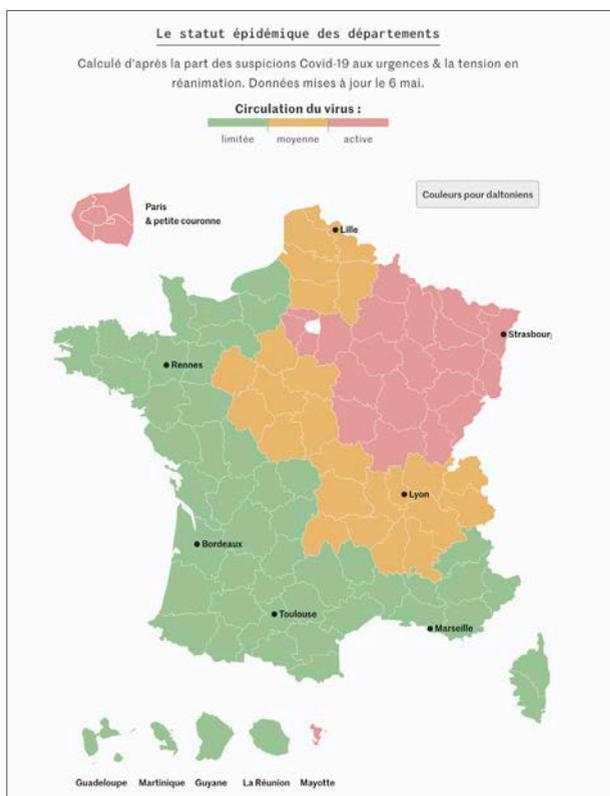
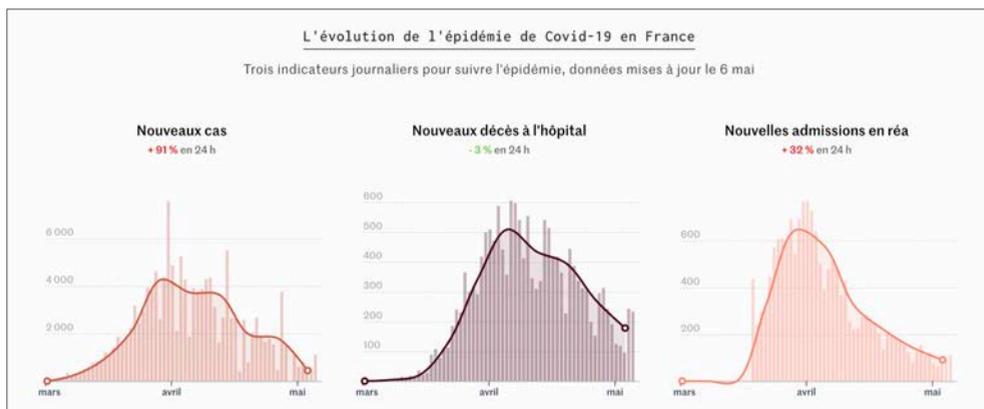


Figure 6 ci-contre. Désormais obsolète, cette carte, publiée le 6 mai 2020, était au début du second tableau de bord dévolue à synthétiser chaque jour le « statut épidémique » des départements décrété par le gouvernement. Une métrique imparfaite qui a été abandonnée au bout de quelques mois.

Figure 7 ci-dessous. Ce second tableau de bord, publié le 6 mai 2020, proposait aussi de suivre les statistiques gouvernementales en distinguant les nouvelles données quotidiennes en barres de tendances à moyen terme (les moyennes glissantes sur sept jours) incarnées par des lignes. Une combinaison qui deviendra rapidement un standard dans les visualisations dédiées au Covid-19.



L'accent a aussi été mis sur le design de la page, conçu avec les équipes de la direction artistique du journal pour harmoniser les visuels et proposer une lecture simple, claire, et agréable aux lecteurs. C'est passé par une refonte des codes couleur qui différenciait les multiples données (les contaminations, les hospitalisés, admis en réanimation, et les décès, principalement). Les nuances ont été choisies afin de refléter une « gradation » de gravité, du simple cas jusqu'au décès, en prenant soin d'obtenir un « espacement » chromatique similaire entre chaque.

Après trois semaines de télétravail (nous étions alors tous confinés), de design et de « couture » tant éditoriale que technique entre les différentes parties du tableau de bord, le projet a fini par prendre forme.

Alors, quand le tableau de bord a été publié, le 5 mai 2020⁶, des doutes nous ont encore traversés : cette page de suivi n'arrivait-elle pas trop tard ? Serait-elle utile aux lecteurs et lectrices, alors que certains scientifiques professaient qu'il n'y aurait « pas de deuxième vague » ? Ces inquiétudes ont été vite démenties par le succès du tableau de bord.

LA NÉCESSITÉ DE S'ADAPTER, DE MAINTENIR ET DE FAIRE ÉVOLUER

La publication du tableau de bord n'a cependant pas signifié la fin du travail autour de cet objet éditorial. Il s'agissait désormais de maintenir les flux de données au gré des pannes techniques ou des modifications de la structure des données fournies par les organisations publiques, en particulier Santé publique France (SPF). C'étaient d'ailleurs parfois des lecteurs assidus qui pointaient les *bugs*, notamment quand ces derniers survenaient le soir ou en fin de semaine.

Pour certains, il suffisait d'adapter nos scripts d'importation des données, en les rendant plus stricts et en prévoyant la détection et le traitement des erreurs possibles. Pour d'autres, il s'agissait de modifier le code de nos graphiques en fixant un maximum dans l'échelle des ordonnées pour éviter qu'un rattrapage statistique fait par un pays n'aplatisse le graphique des cas confirmés correspondant à celui-ci. En mai 2020, des corrections concernant les données des décès en EHPAD ont aussi donné lieu à des adaptations pour éviter d'afficher des valeurs négatives de morts. Les modifications plus légères des données de SPF, par exemple en mars 2021, après que l'agence a décelé une surestimation de ses indicateurs de suivi des tests virologiques, ont été automatiquement apportées dans nos graphiques de suivi, ceux-ci étant branchés sur des fichiers régulièrement mis à jour *via* des *cron jobs*, c'est-à-dire des actions automatisées depuis notre serveur.

Dans plusieurs cas, un article d'explication a été publié en parallèle du tableau. Pour les informations moins importantes, un système de « messagerie » a permis de placer un encart au-dessus d'un graphique, donnant du contexte sur les changements ou pannes. L'objectif était toujours d'expliquer ce qui s'était passé aux lecteurs et lectrices, de leur rappeler constamment les définitions, les limites et les précautions à prendre avec les données qui étaient partout dans le discours public.



Figure 8.

Le premier tableau de bord, publié le 30 mai 2020, proposait une synthèse de l'état de la pandémie dans le monde avec un graphique en barres empilées au temps T (inspiré par Reuters Graphics) et un graphique historique.

⁶ https://www.lemonde.fr/les-decodeurs/article/2020/05/05/coronavirus-age-mortalite-departements-pays-suivez-l-evolution-de-l-epidemie-en-cartes-et-graphiques_6038751_4355770.html

Nous avons d'ailleurs toujours gardé en tête que les données restaient au cœur des enjeux politiques : régulièrement, nous avons décidé d'ajouter ou de retirer des graphiques de notre page de suivi. Nous avons par exemple décidé d'intégrer, début 2021, des données réclamées aux autorités sanitaires depuis longtemps : la distinction entre les hospitalisations de personnes soignées « pour » Covid-19 et celles de personnes porteuses du Covid-19 mais hospitalisées pour un autre motif.

Au total, entre le 5 mai 2020 et le 15 décembre 2022, le tableau de bord a été republié – et donc modifié – 549 fois, soit en moyenne tous les deux jours. La majorité de ces republications était destinée à réparer des *bugs*, le reste à modifier, ajouter ou retirer des visuels ou des textes. Certaines ont été de plus grande envergure : en mai 2021, pour le premier anniversaire de notre outil, une refonte importante a été publiée, incluant des évolutions graphiques : placement de la partie française en haut de page, évolution de la présentation des visuels, optimisation sur écrans mobiles, ajout de certains indicateurs et retrait d'autres, comme le profil des morts.

UN TABLEAU DE BORD PEUT EN CACHER D'AUTRES

Mi-juillet 2020 a été créé sur le même modèle un article de suivi des tests virologiques par département, accompagnant la montée en puissance du dispositif national de dépistage pendant l'été, puis un autre mi-novembre (à la faveur du second confinement) classant les 101 départements français selon la tendance des hospitalisations. En 2021, deux autres “trackers” ont été publiés : un article sur l'évolution des variants du coronavirus Sars-CoV-2, mais surtout un suivi du fait majeur de la première moitié de 2021 : la campagne de vaccination.

Après la première injection, le 27 décembre 2020, les jeux de données sur la vaccination sont arrivés rapidement en libre accès (toujours sur Data.gouv.fr, le portail officiel d'*open data* de l'État français), ce qui nous a permis de lancer un nouveau tableau de bord, aussi important que celui sur le suivi de la pandémie en elle-même. Les mêmes règles de design ont été appliquées afin d'en faire une page jumelle de celui du printemps 2020. Dans les semaines et les mois qui ont suivi le déploiement de la vaccination contre le Covid-19 en France, ce nouveau tableau de bord a connu un succès d'audience certain.

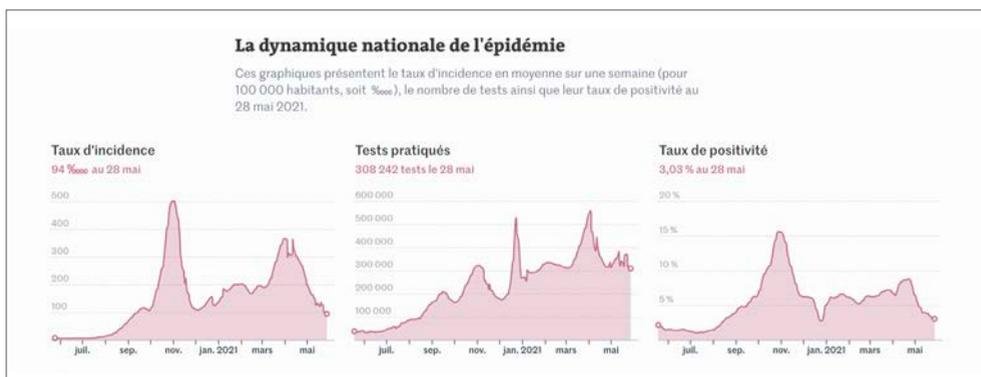


Figure 9. Les données virologiques, publiées le 28 mai 2021, font partie des nouvelles métriques mises progressivement en *open data* par les services gouvernementaux, et que l'on a choisies de représenter afin de mieux comprendre les évolutions du nombre de cas détectés.

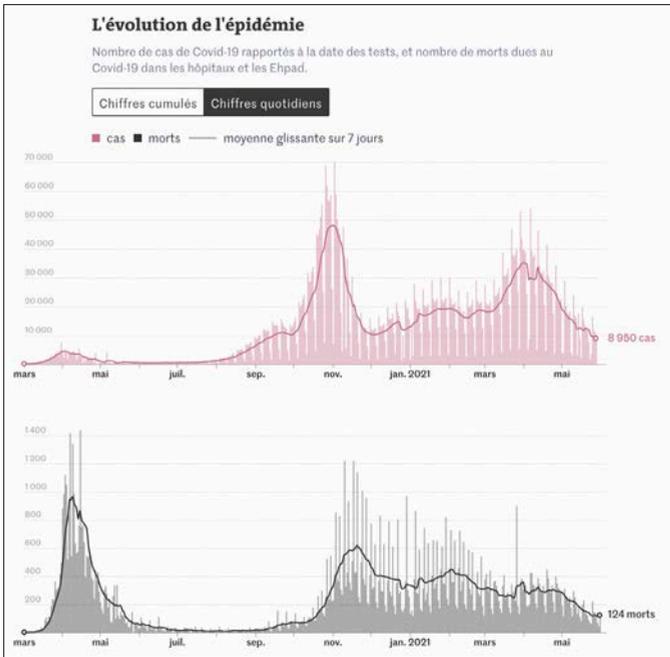


Figure 10.
Un an après sa première publication, le 18 juin 2021, le principal tableau de bord de suivi a subi une mise à jour majeure de son design, plus cohérent et plus travaillé.

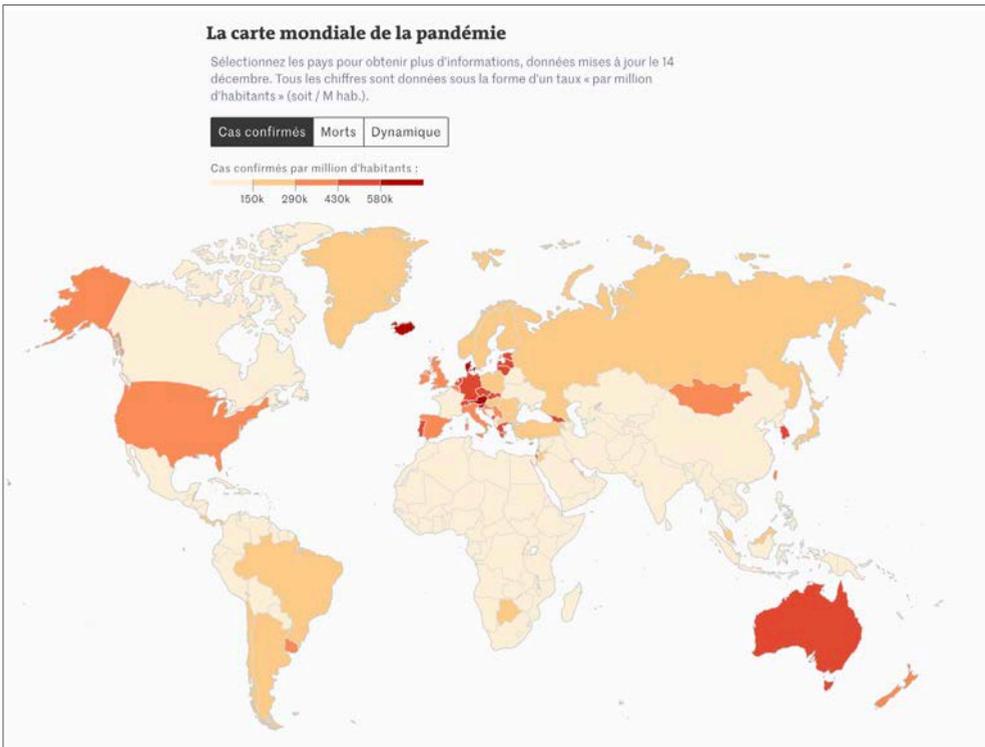


Figure 11. Le planisphère a été revu et corrigé pour proposer, le 14 décembre 2022, en plus des cas et des morts, un onglet « dynamique » qui visualise les tendances récentes de chaque pays.

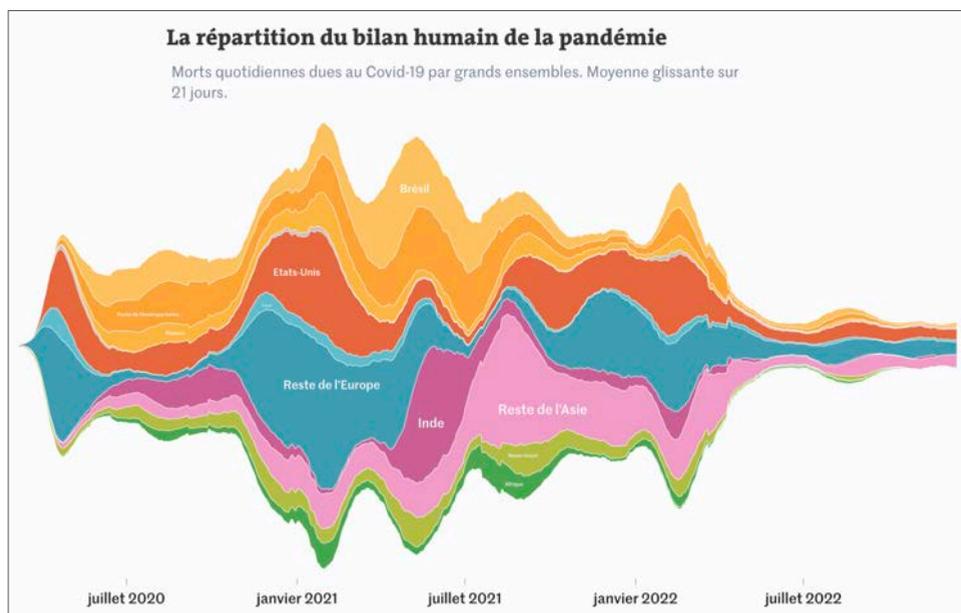


Figure 12. Ce tableau de bord mis à jour s'est enrichi, le 14 décembre 2022, de quelques visualisations, dont ce "streamchart" permettant de visualiser la répartition des décès dus au Covid-19, par grand bloc géographique (inspiré de l'excellent travail du *Financial Times* en la matière).

UN SUCCÈS D'AUDIENCE CERTAIN ET UNE MONTÉE EN COMPÉTENCES

Le tableau de bord principal dédié au Covid-19 est l'article le plus lu de l'histoire du site du *Monde*, notamment grâce à sa longévité. Au 15 décembre 2022, il cumule plus de 30 millions de visites. Plus largement, sur la même période, les huit articles de suivi du Covid-19 publiés aux Décodeurs en 2020 et 2021 ont attiré près de 60 millions de visites.

Si l'exploitation de données était déjà au cœur des pratiques éditoriales des Décodeurs, la création et la maintenance des tableaux de bord du Covid nous ont appris beaucoup sur la visualisation des données, sa mise en contexte et sa hiérarchisation.

Grâce à la montée en compétences induite par ce projet, le format « tableau de bord » est également entré dans nos habitudes. Nous l'avons ainsi décliné sur l'élection présidentielle, ou la production d'électricité.