Les encyclopédistes amateurs face à la pandémie de Covid-19. Éditorialisation citoyenne des données épidémiologiques sur Wikipédia

Antonin Segault

▶ **To cite this version:**

Antonin Segault. Les encyclopédistes amateurs face à la pandémie de Covid-19. Éditorialisation citoyenne des données épidémiologiques sur Wikipédia. Les Cahiers du numérique, 2020, 16, 10.3166/LCN.2020.008 . hal-03242102

HAL Id: hal-03242102
https://hal.science/hal-03242102

Submitted on 1 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LES ENCYCLOPÉDISTES AMATEURS FACE À LA PANDÉMIE DE COVID-19

Éditorialisation citoyenne des données épidémiologiques sur Wikipédia

Antonin Segault, Laboratoire Dicen-IdF, Université Paris Nanterre

Cet article s'intéresse au processus de science citoyenne que constitue l'éditorialisation de données scientifiques sur l'encyclopédie libre Wikipédia. Il étudie pour cela le développement d'une série de représentations des statistiques épidémiologiques de la Covid-19 sur une période de six mois. À travers l'analyse des historiques de versions et des pages de discussion, il met en évidence l'influence des règles, des outils et de la communauté de Wikipédia sur la qualité des productions. Il montre également que le développement de ces représentations contribue au partage d'un corpus de compétences variées relevant de la littératie des données.

1. Introduction

Depuis quelques années, se confrontent deux visions antagonistes d'Internet : creuset des formes les plus viles de malinformation, mais aussi terreau d'un processus mondial de mise en commun des connaissances. Deux visions qui se réalisent dans des dispositifs bien concrets, tels que les sites web des mouvements suprématistes aux USA (ou la « réinfosphère » de l'extrême-droite française) pour la première, et l'encyclopédie libre Wikipédia pour la seconde. Et chacun de ces dispositifs voit ses usagers faire appel à des concepts, des données et des démarches qui se réclament de la rationalité scientifique. L'expression citoyenne en ligne semble ainsi capable du meilleur comme du pire, soulignant les enjeux de la citoyenneté de données (*data citizenship*), c'est-à-dire de la capacité des citoyens et citoyennes à développer une participation active et critique dans une société gouvernée par les algorithmes et les données (Carmi et al., 2020).

Tout au long du printemps et de l'été 2020, les regards sont restés braqués sur quelques lignes de chiffres : les statistiques épidémiologiques de la pandémie de Covid-19 qui sévit dans le monde entier. À travers ces données, citoyens et citoyennes tentent de comprendre la crise : Le virus se répand-il plus ou moins vite ? La pandémie recule-t-elle ou connaît-elle une deuxième vague ? Le confinement va-t-il être prolongé, levé ou rétabli ? Toutes ces questions essentielles ne peuvent trouver de réponse que dans une compréhension fine de données complexes et parfois trompeuses. Cet article propose par conséquent d'étudier la manière dont ces dernières ont été représentées, expliquées et discutées sur l'encyclopédie en ligne Wikipédia. Nous reviendrons tout d'abord sur les sciences citoyennes et le rôle qu'elles peuvent jouer en situation de crise, puis nous intéresserons plus particulièrement au cas de la contribution sur l'encyclopédie libre Wikipédia. Nous nous focaliserons ensuite sur l'article « Pandémie de Covid-19 en France », terrain de cette étude, afin de décrire les formes qu'y a pris la représentation des données épidémiologique, et les activités qu'y ont déployé les contributeurs depuis sa création (le 29 février) 2020 jusqu'à la fin du mois

d’août 2020. Nous tenterons enfin de déterminer la manière dont le fonctionnement de Wikipédia a affecté le déroulement et les résultats de ce processus particulier de sciences citoyennes.

2. Wikipédia face aux situations de crise

Les personnes confrontées à une situation de crise font face à un déficit d’information. Celui-ci touche évidemment les professionnels et les acteurs institutionnels impliqués dans la gestion de la crise, car « avoir accès à l’information dont on a besoin en temps opportun permet aux gestionnaires de prendre les bonnes décisions et de réussir leurs opérations » (Ait Ouarab-Bouaouli, 2008). Mais la recherche d’information concerne également les citoyens qui sont directement ou indirectement affectés par la situation. En effet, la gestion de crise requiert souvent une action collective (évacuation, confinement, alerte, etc.), impliquant la coopération des citoyens et des professionnels (Fugate, 2011), sur la base d’informations partagées. Plus généralement, la théorie de la dépendance postule que la dépendance des citoyens aux médias tend à s’accroître en période de conflit ou de changement social (Ball-Rokeach & DeFleur, 1976). La crise, caractérisée par une rupture avec le fonctionnement habituel du monde, est génératrice d’incertitude et, par conséquent, d’un besoin accru d’informations nouvelles, susceptibles d’orienter l’action.

2.1. Crises et sciences citoyennes

Longtemps, la production de données précises quant à l’environnement, la population et leurs évolutions multiples a été l’apanage des scientifiques, affiliés à des institutions leur procurant le matériel et la crédibilité nécessaire. Cette hégémonie est cependant bousculée par les sciences citoyennes (*citizen sciences*), des démarches scientifiques caractérisées à la fois par un grand nombre de contributeurs volontaires (plutôt qu’une équipe de chercheurs sélectionnés au début du projet) et par la mise à disposition des outils et des données de manière ouverte (Franzoni & Sauermann, 2014). Au cours des dernières décennies, ces démarches se sont démocratisées du fait de la diffusion des compétences scientifiques, mais aussi du développement de dispositifs (capteurs, ordinateurs, etc.) peu coûteux (Wildschut, 2017). En raison de la place croissante occupée par les données, dans l’activité scientifique tout comme dans le fonctionnement de la société, l’implication des citoyens requiert un certain degré de littératie des données (*data literacy*) (Carmi et al., 2020), incluant des compétences statistiques (*statistical literacy*), mais également communicationnelles, créatives, éthiques, etc. (Wolff et al., 2016). Loin de s’opposer radicalement à la science « académique », les projets de science citoyenne associent souvent l’expertise des citoyens (individuellement ou à travers des associations) et celle des chercheurs universitaires.

Ces projets peuvent mobiliser différentes formes de participation citoyenne. La collecte de données par les citoyens eux-mêmes (*crowdsourcing*) permet la constitution rapide de grand corpus. Ainsi, des dispositifs de suivi médical facilitent l’étude de la santé de larges cohortes (Wiggins & Wilbanks, 2019) tandis que le déploiement de capteurs par des associations offre une mesure plus fine de la qualité de l’air (Parasie & Dedieu, 2019). Mais la science citoyenne ne se limite pas à la production de données : de nombreuses applications s’appuient sur l’expertise citoyenne pour catégoriser des

images (Galaxie Zoo), tester des modèles (Foldit) ou annoter des documents anciens (Scribes of the Cairo Geniza), tandis que d'autres exploitent les ordinateurs inutilisés pour effectuer des calculs scientifiques (SETI@Home, Rosetta@Home, etc.). Encore plus en aval du processus de recherche, les politiques d'Open Data, par lesquelles des chercheurs et des institutions publiques (mais aussi parfois des entreprises) mettent à disposition une partie des données qu'ils ont produites, constituent un terreau favorable pour des projets de science citoyenne portant sur l'analyse, le croisement et l'éditorialisation de données existantes.

Lors de situations de crise, face au besoin d'information, on retrouve cette même diversité des formes de science citoyenne. Après l'explosion de la plateforme pétrolière Deepwater Horizon (2010), une ONG a créé une carte interactive s'appuyant sur le *crowdsourcing* pour documenter l'impact de la marée noire (McCormick, 2012). Après l'accident nucléaire de Fukushima Daiichi, des collectifs d'internautes ont également joué un rôle important dans l'agrégation et l'interprétation de mesures de la radioactivité produites tant par des riverains que par les institutions gouvernementales (Plantin, 2015). En situation de crise, face à l'incertitude, les citoyens font ainsi feu de tout bois, s'appuyant sur une grande variété de sources de données pour s'informer et contribuer à informer les autres.

2.2. Couverture des situations de crise sur Wikipédia

En 2001 déjà, après les attentats du World Trade Center, les citoyens ont massivement eu recours à Internet pour s'informer (Bucher, 2002). Aujourd'hui, les médias sociaux tels que Facebook ou Twitter jouent un rôle central dans la diffusion et la recherche d'informations lors de situations de crise (Sutton et al., 2008). L'encyclopédie en ligne Wikipédia, l'un des dix sites web les plus consultés au monde (Alexa, s. d.), n'est pas en reste. Ses articles, qui apparaissant bien souvent en première position dans les moteurs de recherche, connaissent des pics de consultation lors des événements sportifs, des échéances politiques, mais aussi des catastrophes naturelles ou humaines (Beaude, s. d.). Cependant, l'usage de Wikipédia ne se limite pas à la recherche d'information : les articles de cette encyclopédie libre sont intégralement rédigés, corrigés et mis à jour par les internautes eux-mêmes. Un ensemble de règles et d'outils (pages de discussions associées à chaque article, historiques enregistrant toutes les modifications, bandeaux signalant des articles problématiques, etc.) guident les contributions, dont la qualité est assurée par la « vigilance participative » de la communauté (Cardon & Levrel, 2009).

Ainsi, pour les citoyens faisant face à une situation de crise, Wikipédia constitue-t-elle à la fois une source d'information et un espace de partage. Bien que Wikipédia puisse être considéré comme un « lieu de production et de transmission de connaissances » (Beaude, 2015), les règles de contribution s'opposent à la publication de travaux personnels inédits. L'ensemble des informations intégrées dans un article sont soumises à la neutralité de point de vue[1] et doivent être corroborées par des sources externes : « Wikipédia n'est pas un endroit où publier des idées originales, ni un forum pour défendre ses idées ; toutes les connaissances présentées doivent être vérifiables et

[1] « Neutralité de point de vue », dans les principes fondateurs de l'encyclopédie Wikipédia : https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Neutralit%C3%A9_de_point_de_vue

pertinentes »[2]. Les contributeurs ne pouvant s'appuyer sur des données qu'ils auraient eux-mêmes produites ou récoltées (*crowdsourcing*), leur activité relève principalement de l'agrégation et de la mise en forme de données existantes. Ainsi la participation au projet encyclopédique de Wikipédia peut être assimilée à une forme de science citoyenne située en aval du processus de recherche, focalisée sur l'éditorialisation des résultats de travaux préalables.

Si les règles de Wikipédia encouragent l'usage de sources universitaires (livres, articles, etc.)[3], de telles sources ne sont généralement pas disponibles lorsqu'un article traite d'un événement en cours ou très récent, et des sources d'une fiabilité moindre sont alors employées. Par exemple, dans les premiers jours de l'accident nucléaire de Fukushima Daiichi, l'article Wikipédia qui lui était consacré s'appuyait principalement sur des articles de presse, voire des communiqués publiés par l'exploitant de la centrale lui-même (Moats, 2019). Pour cette raison, la couverture d'événements en cours est découragée dans les règles de l'encyclopédie : « Wikipédia n'est pas un journal d'information et n'a pas pour objectif de relater les faits et événements d'actualité dès qu'ils surviennent »[4]. Cependant, ces recommandations ont peu de poids lorsque les internautes sont confrontés à des situations marquantes (Ferron & Massa, 2011), et les articles rassemblant le plus de contributeurs sont presque systématiquement relatifs à des événements en cours (Keegan et al., 2013). Dans la communauté, le « récentisme » est perçu comme une source d'articles trop longs et sujets aux conflits d'édition[5]. L'étude du processus de rédaction au cours d'événements d'actualité révèle effectivement des dynamiques de collaboration différentes de celles observées sur d'autres sujets (Keegan et al., 2013). Les articles résultant présentent dans un premier temps une représentation désorganisée et instable de l'événement, qui évolue ensuite progressivement vers un point de vue plus neutre (Ferron & Massa, 2011), davantage en accord avec les règles de Wikipédia.

2.3. L'épidémie de Covid 19 sur Wikipédia

L'année 2020 aura été profondément marquée par la pandémie de Covid-19 (la maladie provoquée par le coronavirus SARS-CoV-2) qui, après son émergence en Chine à la fin de l'année 2019, s'est rapidement répandue dans le monde. En France, la première vague épidémique qui déferle à la fin de l'hiver a tué plus de 30 000 personnes[6], éprouvant durement les capacités des services de santé. Les mesures de confinement drastiques, mises en œuvre du 17 mars au 11 mai pour endiguer l'épidémie, ont profondément affecté la vie des individus et de la société dans son ensemble. Le lent déconfinement, les mesures sanitaires persistantes et la perspective (toujours plus réaliste et inquiétante moment où ces lignes sont écrites – plus encore au moment de

---

2 « Travaux inédit », dans les règles de l'encyclopédie Wikipédia : https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Travaux_in%C3%A9dits

3 « Sources fiables », dans les règles de l'encyclopédie Wikipédia : https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Sources_fiables

4 « Ce que Wikipédia n'est pas », dans les recommandations de l'encyclopédie Wikipédia : https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Ce_que_Wikip%C3%A9dia_n%27est_pas#Un_journal_d'information

5 « Recentism », dans les pages d'explication de l'encyclopédie Wikipédia : https://en.wikipedia.org/wiki/Wikipedia:Recentism

6 D'après les chiffres du ministère de la santé au 20 septembre : https://www.gouvernement.fr/info-coronavirus/carte-et-donnees

leur relecture) d'un rebond de l'épidémie sont autant de signes qu'il s'agit là d'une crise majeure, dont l'impact à long terme ne peut encore guère être mesuré.

Cette situation de crise a, évidemment, fait l'objet d'une couverture médiatique intense, mais également d'un grand nombre d'articles sur Wikipédia. Fin août 2020, la page principale consacrée à l'épidémie sur la version française de l'encyclopédie, « Pandémie de Covid-19 », compte près de 40 000 mots. À cette même date, l'encyclopédie française comporte également des articles décrivant la situation dans 135 pays du monde (intitulées selon le modèle « Pandémie de Covid-19 en Afrique du Sud »[7]), ainsi que chacune des régions françaises (en métropole et outre-mer). S'y ajoutent des dizaines d'articles relatifs aux actions de prévention (Port du masque pendant la pandémie de Covid-19, StopCovid, etc.), aux traitements médicaux (Test diagnostique du SARS-CoV-2, Vaccin contre la Covid-19, etc.), aux conséquences de la pandémie (Confinement de 2020 en France, Conséquences économiques, sociales et environnementales de la pandémie de Covid-19, etc.), ainsi qu'à certains groupes spécifiques (Covid-19 pendant la grossesse, Conséquences de la pandémie de Covid-19 sur les communautés LGBT, etc.). Cet effort considérable des contributeurs de Wikipédia a fait l'objet de plusieurs articles de presse en France et à l'étranger[8].

Cette étude se concentrera principalement sur l'article « Pandémie de Covid-19 en France ». Créé le 29 février 2020, cet article a fait l'objet de plus de 7000 modifications en l'espace de six mois, pour une longueur dépassant les 20 000 mots (Figure 1). Il couvre une grande variété de sujet : la chronologie de l'épidémie, la lutte contre sa propagation, son bilan humain, ses conséquences économiques, etc. Du fait de sa proximité avec l'actualité, plusieurs modèles ont été placés en haut de l'article (Figure 2) : un bandeau de « mise en garde médicale », indiquant que « Wikipédia ne donne pas de conseils médicaux ou sanitaires », un bandeau « événement en cours », rappelant que « Cet article concerne un événement récent ou en cours », et un encart « informations officielles », rassemblant plusieurs liens institutionnels juste en dessous de l'infobox[9]. En raison de l'accroissement rapide de la longueur de l'article, un bandeau recommandant sa scission a également été ajouté début août.

---

7 La normalisation des titres a fait l'objet d'une procédure de vote et de consensus dont l'on trouve trace dans la page de discussions de l'article : https://fr.wikipedia.org/wiki/Discussion:Pand%C3%A9mie_de_Covid-19/Archive_2#Choix_du_titre

8 Par exemple, « Sur Wikipédia, le défi du travail d'information sur le coronavirus » (Le Parisien, 6 avril 2020 : https://www.leparisien.fr/societe/sur-wikipedia-le-defi-du-travail-d-information-sur-le-coronavirus-06-04-2020-8294709.php) ou encore « How Wikipedia Prevents the Spread of Coronavirus Misinformation » (Wired, 15 mars 2020 : https://www.wired.com/story/how-wikipedia-prevents-spread-coronavirus-misinformation/)

9 Le terme « infobox » désigne l'encadré synthétique qui se trouve en haut à droite de la majorité des articles de Wikipédia : https://fr.wikipedia.org/wiki/Aide:Infobox

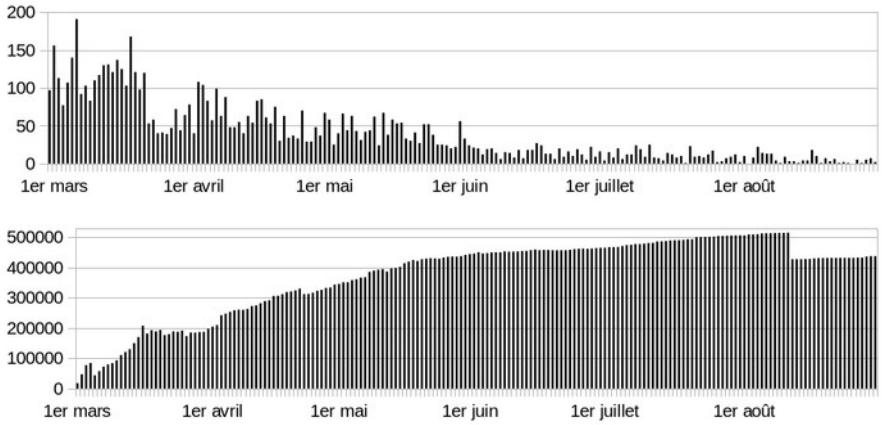


Figure 1 : Nombre de modifications (en haut) et longueur en octets (en bas) de l'article « Pandémie de Covid-19 en France » du 29 février au 31 août 2020

Figure 2 : Haut de l'article « Pandémie de Covid-19 en France », montrant les bandeaux en place au 22 août 2020

3. Éditorialisation des statistiques épidémiologiques de la Covid-19 sur Wikipédia

Parmi les informations relatives à la pandémie de Covid-19 et à son évolution, les statistiques épidémiologiques occupent une place considérable. Chaque soir vers 20 h, le ministère de la santé actualise une lugubre série de métriques : nombre de personnes décédées, hospitalisées, diagnostiquées, etc. Pendant la période de confinement, ces chiffres, souvent présentés en direct à la télévision par le ministre, sont scrutées avec attention, faisant l'objet d'interminables commentaires dans la presse, sur les médias sociaux et, sans aucun doute, dans les salons de millions de citoyens. Ces données sont également publiées sur les sites du ministère de la santé[10] (sous forme d'un tableau de

[10] Tableaux de bord et cartes publiés par le ministère de la santé sur le site du gouvernement : https://www.gouvernement.fr/info-coronavirus/carte-et-donnees

bord interactif conçu spécifiquement par Étatlab) et de l'agence nationale de santé publique[11] (un tableau de bord, mais aussi une section de sa plateforme d'Open Data, Géodes). Elles sont par conséquent reprises dans de nombreuses publications, notamment par les médias nationaux, et font l'objet d'une grande variété de représentations graphiques. Ces courbes et ces histogrammes constituent l'image même de la crise et de sa gestion[12], à tel point que leur impact social peut poser question (Kennedy, 2020).

L'article « Pandémie de Covid-19 en France » comporte un grand nombre de ces graphiques relatifs aux statistiques épidémiologiques. Les sections suivantes se concentrent sur trois des principales modalités de représentation qui sont employées : les histogrammes qui composent la sous-section « Statistiques » de l'article, les cartes des cas qui illustrent l'infobox de l'article, et le modèle synthétique qui rassemble plusieurs séries de données sous l'infobox. L'analyse de l'apparition et de l'évolution de ces représentations graphiques s'appuie sur l'étude du contenu de l'article à la fin août 2020, mais également des versions successives de l'article depuis sa création (accessibles depuis l'historique de versions), ainsi que de la page de discussions où les contributeurs et contributrices échangent et débattent des modifications à apporter à l'article.

3.1. Représentation des séries statistiques sous forme d'histogrammes

Dans la section « Bilan humain » de l'article, une sous-section entière est consacrée à la représentation graphique des séries statistiques relatives aux contaminations, aux hospitalisations, aux guérisons et aux décès liés à la Covid-19. Ces histogrammes sont apparus progressivement, pour la plupart entre mars et avril (Figure 4), au gré de l'évolution des sources de données et de l'intérêt des contributeurs. Les nombres de décès dans les établissements et services sociaux et médico-sociaux (ESMS, qui incluent notamment les EHPAD), en particulier, ont fait l'objet de nombreuses discussions en raison de leur remontée progressive, basée sur des méthodes assez différentes de celles du milieu hospitalier[13]. De fait, l'intégration d'un histogramme présentant leur évolution quotidienne (qui ne correspond donc pas à la date des décès mais à celle de leur remontée), a été plusieurs fois repoussée, et leur combinaison avec les statistiques hospitalières a été tardive.

---

[11] Dossier consacré à la Covid-19 sur le site de Santé Publique France : https://www.santepubliquefrance.fr/dossiers/coronavirus-covid-19

[12] Comme en témoigne l'expression « aplatir la courbe », dans toutes les bouches et dans tous les tweets au début du confinement. Un article Wikipédia lui est même dédié : https://fr.wikipedia.org/wiki/Aplatir_la_courbe

[13] Section « Intégration des chiffres des ESMS » dans la page de discussion de l'article : https://fr.wikipedia.org/wiki/Discussion:Pand%C3%A9mie_de_Covid-19_en_France/Archive_1#Int%C3%A9gration_des_chiffres_des_ESMS

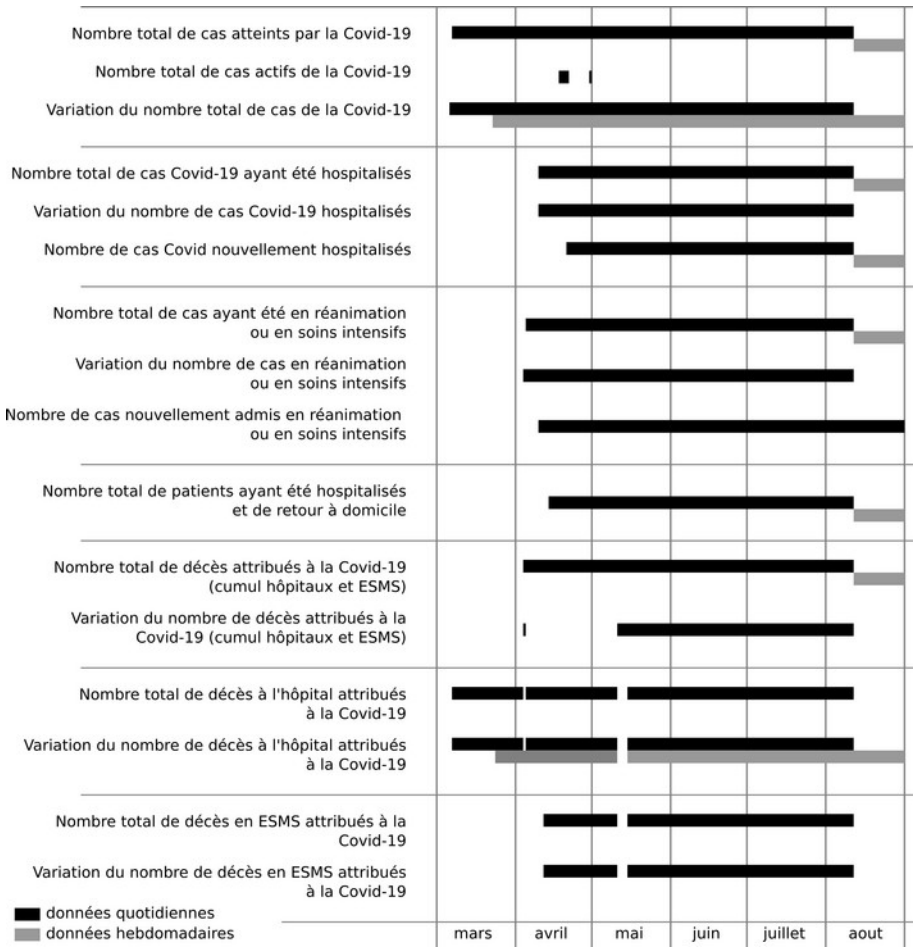
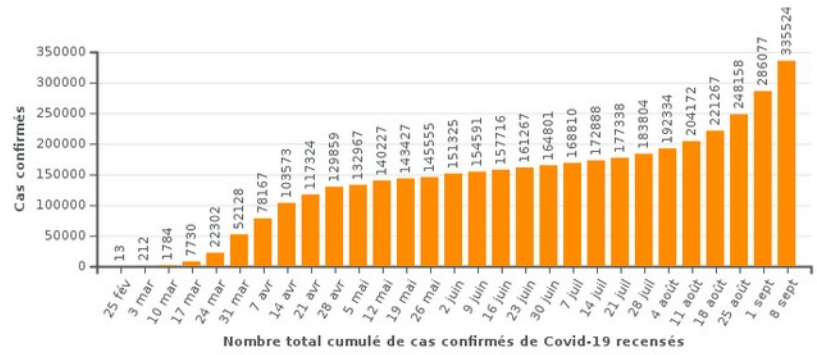


Figure 3 : Présence de différents histogrammes dans l'article « Pandémie de Covid-19 en France » entre le 29 février et le 31 août 2020. La désignation des histogrammes est normalisée et ne reflète pas leurs intitulés dans l'article.

Alors que les articles Wikipédia relèguent généralement les références dans une section dédiée, en bas de page, les sources des données statistiques ont été intégrées dans le corps même de l'article, faisant l'objet d'une longue description au début de la sous-section Statistiques, à la suite d'un échange dans la page de discussion[14]. De nombreuses indications sont également ajoutées autour de chaque histogramme pour tenter d'expliquer la lecture et les spécificités des séries de données. Pour le graphique des décès en ESMS, par exemple, il s'agit d'expliquer les modalités de production des données, afin d'en éviter des interprétations erronées : « Remarque : la remontée des données a été effectuée progressivement par les EHPAD et EMS en avril. Pour ce mois, le tableau indique donc la progression de la mise à jour en complément d'une éventuelle augmentation quotidienne du nombre de décès (surtout les premiers jours) ». Certaines remarques portent sur un point de données en particulier, que des anomalies ou des artefacts méthodologiques affectent particulièrement. Par exemple, dans l'histogramme des cas confirmés : « Note pour le 29 avril : Anomalie dans le comptage des cas de Covid-19 présente sur le site du Gouvernement. »

[14] Section « Sources plus précises pour statistiques SVP » dans la page de discussion de l'article : https://fr.wikipedia.org/wiki/Discussion:Pand%C3%A9mie_de_Covid-19_en_France/Archive_1#Sources_plus_pr%C3%A9cises_pour_statistiques_SVP



Lecture : entre le début du recensement et le 8 septembre 2020, on dénombre 335 524 personnes atteintes de Covid-19 recensées. Sont prises en compte dans ce tableau les personnes atteintes de Covid-19 positivement testées recensées en établissements sociaux et médico-sociaux. Les cas possibles ne sont pas pris en compte.

Note pour le 29 avril : Anomalie dans le comptage des cas de Covid-19 présente sur le site du Gouvernement.

Figure 4 : Un des histogrammes de l'article, capturé en date du 16 septembre 2020, montrant quelques-uns des commentaires visant à en guider la lecture.

Ces explications s'accompagnent d'un ajustement progressif des intitulés des titres, de l'ordre et de la catégorisation des graphiques. Ainsi, les histogrammes présentant l'augmentation quotidienne des cas, des hospitalisations ou des décès sont rebaptisés « variation journalière » lorsque des valeurs négatives y sont apparues. Là aussi, ces chiffres contre-intuitifs (qui ne signalent pas des résurrections), font l'objet d'explications : « Les éventuels chiffres négatifs ou baisses du nombre total de décès sont dus à des rattrapages par Santé publique France d'erreurs de comptage sur les jours précédents ». Au début de la sous-section « Statistiques », une phrase invite également à une lecture distanciée de ces données : « Ces statistiques données chaque jour par les autorités sont imparfaites et incomplètes, notamment parce qu'elles ne prennent pas en compte les décès à domicile, mais aussi parce que la vague épidémique n'est pas terminée. »

Contrairement à la majorité des illustrations de Wikipédia, les histogrammes affichés dans cette section ne sont pas de simples fichiers images. Ils sont générés par le module Graph[15], une extension du logiciel MediaWiki qu'utilise l'encyclopédie, à partir de données actualisées chaque jour directement dans le corps de l'article (selon une syntaxe spécifique) par les contributeurs. Ce module permet une configuration fine des graphiques, qui ont ainsi vu différents aspects de leur présentation (taille, couleur, etc) évoluer au fil des mois. Certains paramètres, tels que l'affichage ou non des valeurs au-dessus des barres des histogrammes, ont fait l'objet de longues discussions, voire de votes[16]. Ces considérations ont également accompagné les débats relatifs à l'addition des décès en hôpitaux et en ESMS : pendant quelques jours, les deux série de données ont été empilées, de deux couleurs différentes, sur un même histogramme, avant qu'une autre présentation ne soit préférée.

---

15 Documentation de l'extension Graph sur le site de MediaWiki : https://www.mediawiki.org/wiki/Extension:Graph

16 Section « Choix graphique discutable pour certains histogrammes » de la page de discussion de l'article : https://fr.wikipedia.org/wiki/Discussion:Pand%C3%A9mie_de_Covid-19_en_France/Archive_2#Choix_graphique_discutable_pour_certains_histogrammes

Le recours au module Graph présente néanmoins un inconvénient : le rendu des graphiques est un processus coûteux en ressources informatiques. Au mois d'août, la longueur de l'article, combinée au grand nombre de graphiques générés, est devenue une source de problèmes techniques, empêchant notamment l'affichage de la palette en bas de l'article. Le 11 août, une profonde refonte de la section Statistiques est entreprise, privilégiant notamment l'utilisation de séries de données hebdomadaires plutôt que quotidiennes. Parmi les graphiques supprimés, bon nombre sont rétablis sur un nouvel article intitulé « Statistiques sur la pandémie de Covid-19 en France ». Cette décision semble néanmoins avoir affecté le travail autrefois bien réglé des contributeurs, aboutissant à une mise à jour plus irrégulière des histogrammes (dont certains se trouvent en doublon sur les deux articles). Début septembre, le déplacement de certains graphiques faisait encore l'objet de débats sur la page de discussion[17] et d'un début de guerre d'édition.

3.2. Représentations cartographiques illustrant l'article

L'infobox de l'article « Pandémie de Covid-19 en France » est introduite par une carte indiquant sous forme de cercles de différents diamètres les nombres d'hospitalisation dans les départements français (visible sur la Figure 2). Fait marquant, à la fin du mois d'août, cette carte présente encore les données du 26 avril, soit près de quatre mois avant. Des données départementales ou régionales sont pourtant régulièrement fournies par l'Agence nationale de santé publique sur ses sites web. Ces données ont été intégrées dans l'article sous forme tabulaire, et régulièrement actualisées jusqu'à la mi-juillet (puis, courant août, ont été déplacées dans le nouvel article consacré aux statistiques). La carte, elle, est restée inchangée.

La cartographie du Covid-19 en France sur Wikipédia est pourtant antérieure à la création de l'article « Pandémie de Covid-19 en France ». Une première carte[18], créée le 26 janvier, indiquait par des aplats de couleur les départements dans lesquels des cas avaient été détectés, sur la base des informations publiées par la presse. Cette carte, hébergée sur Wikimédia Commons, connaît quelques mises à jour, par plusieurs utilisateurs (s'appuyant une des fonctionnalités de Commons, qui permet de « remplacer » un fichier tout en lui laissant le même nom, évitant une actualisation manuelle de tous les articles utilisant l'image ; un historique des versions permet de voir les images successives). Le 29 février, lors de la création de l'article « Pandémie de Covid-19 en France », la carte est complétée et intégrée dans la toute première version de l'infobox de l'article[19]. Le 3 mars, la carte n'ayant pas été mis à jour malgré l'évolution rapide de la situation sanitaire, elle est supprimée de l'article. Vingt-quatre heures plus tard, une nouvelle carte fait son apparition dans l'infobox, présentant cette fois les nombres de cas confirmés dans chaque région, sous la forme d'aplats colorés auxquels sont ajoutés les valeurs numériques[20]. Cette nouvelle carte est régulièrement

---

17 Section « Incohérence » sur la page de discussion de l'article : https://fr.wikipedia.org/wiki/Discussion:Pand%C3%A9mie_de_Covid-19_en_France#Incoh%C3%A9rence

18 Carte « COVID-19 Outbreak Cases in France » sur Commons : https://commons.wikimedia.org/w/index.php?title=File:COVID-19_Outbreak_Cases_in_France.svg

19 Le même jour, son auteur propose également son usage sur la page de discussion de l'article anglophone « COVID-19 pandemic in France » https://en.wikipedia.org/wiki/Talk:COVID-19_pandemic_in_France/Archive_1#Map

actualisée (parfois plusieurs fois par jour) par son créateur et quelques autres contributeurs, à partir des données publiées par le ministère de la santé.

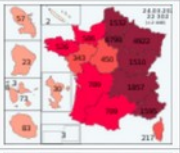
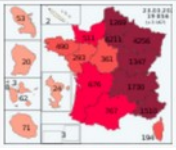
	Date/Time	Thumbnail	Dimensions
current	**19:26, 25 March 2020**		725 × 604 (640 KB)
	19:36, 24 March 2020		725 × 604 (612 KB)
	19:25, 23 March 2020		725 × 604 (611 KB)

Figure 5 : Versions successives d'une carte des cas par région actualisée sur Wikimédia Commons[21]

Le 12 mars, un utilisateur intervient dans la page de discussion, expliquant que les aplats de couleur (il utilise le terme plus précis de « cartes choroplèthes ») ne devraient pas être utilisées pour représenter les nombres de cas, mais seulement pour des ratios tels que le taux de prévalence. Il propose une nouvelle carte, combinant ces deux types d'informations sous forme d'aplats et de cercles de couleurs[22]. Après quelques discussions et modifications, sa carte est intégrée dans l'article le 17 mars au matin en remplacement de la précédente (qui continuera pourtant d'être actualisée sur Commons jusqu'au 25 mars). Dans un premier temps, il dépose tous les jours un nouveau fichier sur Wikimedia Commons puis, à partir du 23 mars, choisir de « remplacer » quotidiennement le fichier[23]. Le 26 mars, l'auteur indique, sur la page de discussion, que les Agences régionales de santé ont cessé la remontée de statistiques régionales[24]. Le même utilisateur renouvelle alors son approche et publie une nouvelle carte présentant cette fois les nombres d'hospitalisations par département, tirés du site du ministère de la santé, sous la forme des cercles de différents diamètres[25]. Cette carte fait l'objet d'actualisations (par remplacement du fichier) du 30 mars au 26 avril, parfois

---

20 L'auteur maintient également une version avec seulement les aplats colorés, sans les nombres, qui lui est parfois préférée dans l'article: https://commons.wikimedia.org/w/index.php?title=File:COVID-19_Outbreak_Cases_in_France_13_Regions.svg

21 Carte « COVID-19 Outbreak Cases in France 13 Regions & DomTom » sur Commons : https://commons.wikimedia.org/w/index.php?title=File:COVID-19_Outbreak_Cases_in_France_13_Regions_%26_DomTom.svg

22 Section « Proposition de cartes » dans la page de discussion de l'article : https://fr.wikipedia.org/wiki/Discussion:Pand%C3%A9mie_de_Covid-19_en_France/Archive_1#Proposition_de_carte

23 Carte « Covid-19, France, nombre de cas et prévalence » sur Commons : https://commons.wikimedia.org/wiki/File:Covid-19,_France,_nombre_de_cas_et_pr%C3%A9valence.png

24 Section « Carte de l'infobox: quelles données ? » sur la page de discussion de l'article : https://fr.wikipedia.org/wiki/Discussion:Pand%C3%A9mie_de_Covid-19_en_France/Archive_1#Carte_de_l'infobox,_quelles_donn%C3%A9es ?

25 Carte « Covid-19, France, nombre d'hospitalisations » sur Commons : https://commons.wikimedia.org/w/index.php?title=File:Covid-19,_France,_nombre_d%27hospitalisations.png

tous les jours, parfois tous les trois ou quatre jours. Cette dernière version est celle qui restera en place plus de quatre mois.

Dans les discussions relatives à la mise à jour des cartes, la génération automatique d'images à partir d'un jeu de données centralisé (par exemple sur Wikidata) a plusieurs fois été évoquée, sans aboutir : les cartes illustrant l'article « Pandémie de Covid-19 en France » ont toujours été créées manuellement par des contributeurs équipés de logiciels appropriés, puis mises en ligne sur Wikimédia Commons. Pourtant, des procédures automatiques ont été mises en œuvre avec succès pour de nombreux autres pays. Les cartes des nombres de cas aux États-Unis[26] et au Royaume-Uni sont automatiquement générées, chaque jour, par un même script Python (mais leur mise en ligne reste manuelle, au vu de l'irrégularité des heures) à partir des données gouvernementales. Le robot XplusMapBot présente un fonctionnement encore plus abouti : chaque jour, les cartes sont générées et mises en ligne automatiquement, tandis qu'un export tabulaire des données est lui aussi déposé sur Wikimédia Commons.[27] Ce robot a d'abord été déployé pour l'Autriche[28] début juillet (ce qui pourrait constituer un indice quant à la localisation de son créateur), puis a rapidement été étendu à d'autres pays européens parmi lesquels la Croatie, la République tchèque, les Pays-Bas, l'Allemagne, l'Italie… mais pas la France.

3.3. Modèle synthétique agrégeant plusieurs séries statistiques

Les modèles (ou templates) sont utilisés sur Wikipédia pour concevoir des bandeaux, des palettes, des infoxbox qui pourront ainsi être intégrés rapidement sur un grand nombre d'articles. Tout comme les articles, les modèles sont créés par les contributeurs de la communauté et disposent d'un historique des versions et d'une page de discussion. Le modèle Coronavirus/Graph/France contient un histogramme présentant de manière empilée les valeurs quotidiennes de cinq séries de données relatives à la Covid-19 : les décès cumulés à l'hôpital, les décès cumulés dans les ESMS, les patients actuellement hospitalisés, les personnes de retour à domicile après hospitalisation, et les cas confirmés mais non hospitalisés (Figure 6). Ce modèle a été créé pour les besoins de l'article « Pandémie de Covid-19 en France », au début duquel il a longtemps été affiché, juste sous l'infobox.

---

[26] Carte des cas aux États-Unis, sur Wikimédia Commons : https://commons.wikimedia.org/wiki/File:COVID-19_outbreak_USA_per_capita_cases_map.svg

[27] Page de description du robot XplusMapBot : https://commons.wikimedia.org/wiki/Commons:Bots/Requests/XplusMapBot L'examen des documents et discussions disponibles n'a pas permis de déterminer comment les données sont agrégées depuis les sources gouvernementales, ni dans quelle mesure cette tâche est automatisée.

[28] Carte des cas en Autriche, sur Wikimédia Commons : https://commons.wikimedia.org/wiki/File:Austria_COVID-19_cases_per_capita.svg

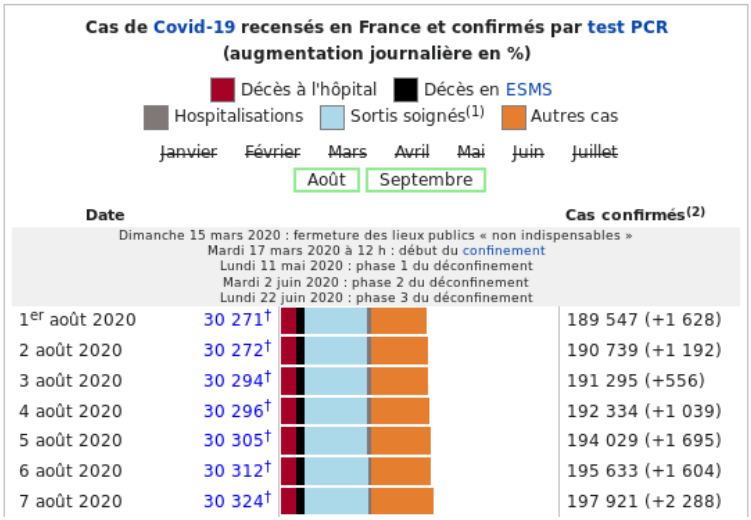


Figure 6 : Extrait du modèle Coronavirus/Graph/France capturé le 13 septembre 2020

La description du modèle indique que les données sont issues des publications du ministère de la santé, de l'agence nationale de santé publique et de l'OMS[29]. Néanmoins, la page de discussion montre que ces sources se sont un temps fait concurrence, en raison de fréquences d'actualisation différentes et de valeurs parfois divergentes. D'autres sources possibles ont également été évoquées, notamment la publication de l'Université Johns-Hopkins, utilisée pour d'autres diagrammes de l'encyclopédie, dont les chiffrent divergeaient parfois en raison d'incohérences dans la prise en compte des Outre-mer ou encore des ESMS[30]. De même, les séries de données affichées ont évolué au cours du temps : par exemple, jusqu'en mai, une série « cas graves », désignant les patients en soins intensifs ou en réanimation, existait en parallèle aux autres hospitalisations. Cette série a été plusieurs fois supprimée (en combinant toutes les hospitalisations) et restaurée, jusqu'à ce qu'un échange dans la page de discussion entérine la suppression[31]. On retrouve, comme pour les histogrammes, de nombreuses discussions relatives aux manières de tenir compte des données manquantes (le plus souvent en laissant la valeur constante pour la journée) ou des baisses des nombres de décès liées à des artefacts méthodologiques (qui sont généralement intégrées au graphique, pour suivre la source).

Tout comme les cartes précédemment décrites, ce modèle s'inscrit dans une série plus large : dans l'encyclopédie francophone, 31 autres pays disposent de représentations similaires. Le modèle français a été créé le 29 février, et la plupart des autres dans le courant du mois de mars. Mais, sur la version anglophone de l'encyclopédie, on trouve des diagrammes semblables plus anciens ; le premier pourrait être celui décrivant la situation sanitaire en Chine, créé le 28 janvier. Côté francophone, le graphique de la

[29] Page de description du modèle Coronavirus/Graph/France : https://fr.wikipedia.org/wiki/Mod%C3%A8le:Coronavirus/Graph/France

[30] Section « Chiffres différents de ceux de John Hopkins University » dans la page de discussion du modèle : https://fr.wikipedia.org/wiki/Discussion_mod%C3%A8le:Coronavirus/Graph/France#Chiffres_diff%C3%A9rents_de_ceux_de_John_Hopkins_University

[31] Section « Cas graves / en réanimation » dans la page de discussion du modèle : https://fr.wikipedia.org/wiki/Discussion_mod%C3%A8le:Coronavirus/Graph/France#Cas_graves_/_en_r%C3%A9animation

France, comme ceux des pays ayant connu une importante vague épidémique tels que l'Italie, l'Espagne ou les États-Unis d'Amérique, présente des nombres importants de modifications, avec une couverture complète de mars à août. Quelques autres, en revanche, n'ont reçu qu'un nombre limité de modification et n'affichent que des données limitées et/ou discontinues. Les pays pour lesquels le modèle a été créé après le mois de mars sont généralement des pays plus éloignés de la France et dans lesquels le pic épidémique est arrivé plus tard (Ouganda, Pérou, etc.).

L'évolution de l'apparence et des fonctionnalités du modèle Coronavirus/Graph/France révèle également des échanges entre plusieurs modèles et plusieurs versions linguistiques de l'encyclopédie. Ainsi, le 29 mars, face à l'allongement du graphique, un contributeur ajoute au modèle des boutons permettant d'afficher ou de masquer les données d'un ou plusieurs mois. La solution semble s'inspirer de celle qui était intégrée dès le 4 mars au modèle relatif à la Chine sur l'encyclopédie anglophone[32]. Ce même utilisateur participe ensuite à la rapide diffusion de cette fonctionnalité sur les modèles francophones d'autres pays (Italie, Espagne, etc). La comparaison avec les autres modèles (de la version francophone) est également invoquée lors d'un débat et d'un vote portant sur le nombre de chiffres significatifs à indiquer dans les pourcentages : « J'ai vérifié pour 6 pays — Italie, Espagne (le cas ne se présente pas), Suisse, Allemagne, États-Unis et Canada — et ce n'était jamais le cas. Donc, par souci d'homogénéité, en plus des raisons invoquées ci-dessus, je pense qu'il vaut mieux ne pas ajouter de décimale. »[33] De manière symétrique, un contributeur intervient sur la page de discussion du modèle Coronavirus/Graph/France pour demander de l'aide sur la mise en forme du modèle Coronavirus/Graph/Venezuela[34].

Le 19 août, le modèle Coronavirus/Graph/France est retiré de l'article « Pandémie de Covid-19 en France » en raison des problèmes de surcharge déjà évoqués. Les données continuent d'être actualisées sur le modèle, bien qu'il ne soit plus utilisé sur aucune page, jusqu'à son intégration dans l'article « Statistiques sur la pandémie de Covid-19 en France » le 31 août. Les échanges sur la page de discussion mentionnent l'idée d'un modèle plus léger, avec moins de données[35], mais le développement ne semble pas commencé fin août. Cette suppression ne semble pas concerner les modèles similaires de l'encyclopédie francophone : même ceux qui comportent beaucoup de données (Espagne, Allemagne, Italie, etc.) semblent toujours affichés sur leur page respective, bien que l'emplacement initial, sous l'infobox, ait généralement été abandonné.

---

[32] Section « Does anyone know how to dynamically show/hide bars for each month? » dans la page de discussion du modèle chinois sur l'encyclopédie anglophone : https://en.wikipedia.org/wiki/Template_talk:COVID-19_pandemic_data/Mainland_China_medical_cases_chart#Does_anyone_know_how_to_dynamically_show/hide_bars_for_each_month?

[33] Section « Erreur ? » dans la page de discussion du modèle : https://fr.wikipedia.org/wiki/Discussion_mod%C3%A8le:Coronavirus/Graph/France#Erreur_?

[34] Section « Aide Venezuela » dans la page de discussion du modèle : https://fr.wikipedia.org/wiki/Discussion_mod%C3%A8le:Coronavirus/Graph/France#Aide_Venezuela

[35] Section « Modèle trop coûteux » dans la page de discussion du modèle : https://fr.wikipedia.org/wiki/Discussion_mod%C3%A8le:Coronavirus/Graph/France#Mod%C3%A8le_trop_co%C3%BBteux

4. Formes de la science citoyenne sur Wikipédia

Les précédentes pages ont décrit le développement et l'évolution de plusieurs types de représentation des données épidémiologiques intégrées à l'article « Pandémie de Covid-19 en France » sur une période de six mois. À travers ces exemples, il apparaît que les contributeurs se livrent à de complexes opérations de recherche, de sélection, de traitement, de mise en forme et d'explication de données issues de sources souvent gouvernementales et ouvertes. De telles opérations peuvent être considérées comme relevant de la science citoyenne, dans la mesure où elles impliquent de nombreux amateurs qui mettent leur production à disposition du grand public (Franzoni & Sauermann, 2014). L'encyclopédie Wikipédia, à travers certaines de ses règles, fonctionnalités et bonnes pratiques semble tour à tour contraindre et faciliter les pratiques de ces contributeurs.

4.1. Représentations multiples mais coordonnées

L'article « Pandémie de Covid-19 en France » frappe tout d'abord par la quantité et la diversité des informations qui le constituent. Durant les trois premiers mois d'existence de l'article, qui recouvrent notamment la phase de confinement en France, les contributions sont nombreuses et la longueur de l'article augmente rapidement (Figure 1). Cette même période voir apparaître un grand nombre de représentations des données épidémiologiques, combinant différentes sources, différentes variables (décès, hospitalisation, guérisons, etc.), différentes granularités (du point de vue temporel et spatial) et différentes modalités visuelles. Chacune présente un certain regard sur un certain aspect de la crise, qui se complètent dans l'article comme une mosaïque complexe. Aux « désirs de données », souvent emprunts d'une inclination au consumérisme ou au contrôle (Carmes & Noyer, 2015), semble ainsi répondre un désir de donner, qui pousse les citoyens et citoyennes à s'engager pour partager leur connaissance et leur compréhension de l'événement qui les frappe. Une réappropriation des données, non pour créer de futiles besoins aux retombées économiques alléchantes, mais pour gracieusement répondre au besoin pressant d'information d'une société plongée dans l'incertitude.

Un tel foisonnement n'est évidemment pas totalement conforme aux visées encyclopédiques de Wikipédia ou même, plus concrètement, à certaines de ses règles de contribution[36]. Il est même caractéristique des dérives imputées aux articles relatifs à des événements récents (*recentism*). Si le caractère ouvert de l'encyclopédie est la cause de cette accumulation, il en apporte également la solution : l'évolution de l'article est émaillée de suppressions, de remplacements, de regroupements d'informations jugées inutiles, redondantes ou obsolètes. Le fonctionnement du wiki, à travers l'apposition de bandeaux, l'organisation de débats dans les pages de discussion, l'annulation de modifications, la création de nouveaux articles connexes, facilite la gestion de l'intense flux de nouvelles contributions. La pandémie de Covid-19 constituant une crise au long cours dont l'issue semble encore lointaine, il n'est pas certain que l'on puisse prochainement observer un retour à la norme wikipédienne, comme ce fut le cas d'articles portant sur des crises plus ponctuelles (Ferron & Massa, 2011). L'important remaniement de la sous-section « Statistiques » à la mi-août montre néanmoins que l'article n'est pas qu'une juxtaposition de contributions isolées, mais bien le produit d'une réflexion collective plus large.

[36] « Article size » dans le guide de contribution : https://en.wikipedia.org/wiki/Wikipedia:Article_size

4.2. Compréhension et mise en avant de la source

Les contributions relatives aux données épidémiologiques de la Covid-19 accordent une place centrale à la description des séries statistiques. Les sources de données, dont le choix a fait l'objet de longs débats dans la page de discussion, sont longuement détaillée dans le corps de l'article et pas seulement dans la section Références. Des instructions de lecture accompagnent chaque histogramme, les intitulés sont soigneusement ajustés pour refléter au mieux la nature des données, des explications méthodologiques sont fournies pour chaque série ou point de données potentiellement contre-intuitif. Les pages de discussion regorgent également de débats sur la manière idéale de représenter différents types de données, le degré de précision à considérer, les stratégies à adopter en cas de valeurs manquantes ou incohérentes, etc. Ces séries statistiques sont en effet difficiles à représenter, parce qu'elles sont le produit d'un suivi épidémiologique complexe par nature, mais aussi parce que la situation exceptionnelle a entraîné des modifications successives et parfois des anomalies dans ce suivi ou dans le processus de remontées d'informations. Les contributeurs ont par conséquent fourni un effort considérable pour rendre ces données intelligibles sans les dénaturer.

La stratégie adoptée face ces données complexes, en accord avec les règles de Wikipédia, consiste à suivre systématiquement la source, quitte à en décrire les subtilités et les anomalies lorsque cela semble nécessaire. Cela nécessite une bonne compréhension des méthodes statistiques employées par les épidémiologistes, mais aussi une capacité à expliquer ces méthodes aux lecteurs de l'article. Cette transparence s'est avérée prudente car, à plusieurs reprises, les contributeurs ont vu leur travail mis en doute par des messages parfois hostiles dans les pages de discussion, tels que : « Je ne m'attendais pas à ce que Wikipedia manipule ! ». L'usage rigoureux des sources permet alors d'apaiser le dialogue, en évitant les attaques contre le travail des contributeurs au profit d'un débat sur les sources (Langlais, 2015). Le bénéfice est alors multiple : pour les critiques, qui comprennent mieux le processus d'édition de l'encyclopédie et les choix effectués dans l'article ; pour les contributeurs, qui perçoivent des faiblesses de leurs sources et peuvent les expliquer, voire découvrir de nouvelles sources plus fiables ; et pour les lecteurs qui bénéficient d'une sélection solidement argumentée des sources les plus pertinentes.

4.3. Acquisition et diffusion de compétences

Le développement des différentes représentations des données statistiques intégrées à l'article fait appel à une large palette d'outils et de techniques, certains appartenant à la plateforme wiki (modèles, extension Graph), d'autres extérieurs, choisis ou développés par des contributeurs (outils de cartographie, robots). La création des cartes, le développement des robots et les modifications des fonctionnalités du modèle semblent n'être chacun l'œuvre que d'un petit groupe de contributeurs réguliers, voir d'un contributeur seul. Cette organisation reflète la tendance à un travail plus distribué qui a été observée dans les articles relatifs à des événements en cours (Keegan et al., 2013). Il faut cependant noter que ces tâches peu partagées (et donc, comme dans le cas des cartes, les plus susceptibles de connaître des interruptions) sont avant tout celles nécessitant des compétences techniques et/ou des outils extérieurs spécifiques. À l'inverse, la mise à jour des histogrammes qui, grâce à l'extension Graph, se fait assez facilement et directement dans l'interface d'édition de l'article, implique un plus grand nombre de contributeurs.

Il apparaît ainsi que l'implication citoyenne pour la publication des données épidémiologiques sur Wikipédia nécessite plus que la seule compréhension des modalités de production et d'interprétation de ces données. Les contributeurs doivent acquérir des compétences techniques (l'usage des différents outils) et créatives (rédaction de texte, apparence des graphiques), mais aussi s'approprier l'ensemble des règles qui régissent l'encyclopédie, corroborant l'idée selon laquelle la littératie des données est plus large que la seule compétence statistique (Wolff et al., 2016). Wikipédia offre différents mécanismes par lesquels ces savoirs et savoirs-faire peuvent se développer et se diffuser : la visibilité du code et des historiques permet de comprendre comment un graphique a pu être réalisé, des bandeaux et des étiquettes signalent les manquements aux règles et les modifications requises, tandis que les pages de discussion sont souvent des lieux d'entraide (certes entre-coupée de débat parfois houleux). Le cas étudié montre en outre que les différentes versions linguistiques de l'encyclopédie ne fonctionnent pas en vase clos, et que les techniques développées sur un wiki peuvent connaître une adoption large et rapide sur les autres.

5. Conclusions et perspectives

Dans cet article, nous nous sommes intéressés aux formes de science citoyenne qui se sont développées sur Wikipédia autour de l'éditorialisation des données épidémiologiques relatives à la pandémie de Covid-19. Une grande diversité d'usages a pu être identifiée autour des histogrammes, des cartes et des modèles, donnant naissance à de nombreuses formes de représentations visuelles. Les discussions et les solutions apportées par les contributeurs révèlent d'importants questionnements méthodologiques, des développements techniques conséquents, et des synergies bénéficiant de la structure internationale de l'encyclopédie.

Il apparaît que, en dépit des réserves de Wikipédia quant à la couverture d'événements en cours, ce dispositif s'est avéré, par bien des aspects, bénéfique pour les pratiques de science citoyenne, les contributeurs qui s'y sont engagés, et les productions qui en sont ressorties. Le niveau d'exigence qui encadre les contributions, défini par un large corpus de règles et souvent rappelé dans les discussions, combiné aux fortes attentes des lecteurs, a induit une vive dynamique d'amélioration des représentations, par la correction d'erreurs, la mise à jour de données obsolète, la rediscussion régulière des sources et des choix graphiques, ainsi que l'explication précise des séries de données. Le recours à des données et des outils pour la plupart ouverts a favorisé l'implication d'un grand nombre de contributeurs et le partage des compétences nécessaires, techniques ou non. Ainsi, les principes, les fonctionnalités et la communauté de Wikipédia semblent contribuer au développement et à la circulation d'une littératie des données riche et rigoureuse chez les contributeurs, y compris en situation de crise.

Il est en revanche bien plus difficile d'estimer l'impact que l'article « Pandémie de Covid-19 en France » pourrait avoir sur la littératie de ses lecteurs. Sa production a fait l'objet d'un grand soin, notamment pour rendre visible et expliquer l'origine des données et de leurs imperfections. Cette exposition inhabituelle aux discours complexes, nuancés et parfois contre-intuitifs qui caractérisent la démarche scientifique ne peut être que positive, comme le suggérait Bruno Latour en juin 2020 : « Le virus a révélé le nombre de choses que vous avez besoin de savoir afin de décider ce qui est factuel et ce qui ne l'est pas. Le grand public apprend beaucoup à propos de la difficulté des statistiques, à

propos des expériences, à propos de l'épidémiologiqe. Dans la vie de tous les jours, des gens parlent de degré de certitude et de marge d'erreur. Je crois que c'est bien. Si vous voulez que les gens comprennent la science, vous devez montrer comment elle est produite. »[37].

Bibliographie et références

Ait Ouarab-Bouaouli, S. (2008). Pour une équité dans l'accès à l'information en cas de catastrophes. *Communication*, *26*(2), 174-193.

Alexa. (s. d.). *The top 500 sites on the web*. Consulté 24 avril 2019, à l'adresse https://www.alexa.com/topsites

Ball-Rokeach, S. J., & DeFleur, M. L. (1976). A dependency model of mass-media effects. *Communication research*, *3*(1), 3–21.

Beaude, B. (s. d.). *Exploring collective attentions*. WikiMaps. http://www.wikimaps.io/

Beaude, B. (2015). De quoi Wikipédia est-elle le lieu? In L. Barbe, L. Merzeau, & V. Schafer (Éds.), *Wikipedia, objet scientifique non identifié*. Presses Universitaires de Paris Ouest (PUPO).

Bucher, H.-J. (2002). Crisis Communication and the Internet : Risk and Trust in a Global Media. *First Monday*, *7*(4). http://uncommonculture.org/ojs/index.php/fm/article/view/943

Cardon, D., & Levrel, J. (2009). La vigilance participative. Une interprétation de la gouvernance de Wikipédia. *Réseaux*, *154*, 51–89.

Carmes, M., & Noyer, J.-M. (2015). Désirs de data. In *Traces numériques et territoires* (p. 179–212). Presses des Mines, Paris.

Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). Data citizenship : Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review*, *9*(2), 1–22.

Ferron, M., & Massa, P. (2011). The Arab Spring| wikirevolutions : Wikipedia as a lens for studying the real-time formation of collective memories of revolutions. *International Journal of Communication*, *5*, 20.

Franzoni, C., & Sauermann, H. (2014). Crowd science : The organization of scientific research in open collaborative projects. *Research Policy*, *43*(1), 1-20.

Fugate, C. (2011). *Understanding the power of social media as a communication tool in the aftermath of disasters*. Federal Emergency Management Agency. https://beta.fema.gov/pdf/about/programs/legislative/testimony/2011/5_5_2011_use_of_social_media_in_disasters.pdf

Keegan, B., Gergle, D., & Contractor, N. (2013). Hot off the wiki : Structures and dynamics of Wikipedia's coverage of breaking news events. *American Behavioral Scientist*, *57*(5), 595–622.

[37] Interview dans le *Guardian*, traduction personnelle : https://www.theguardian.com/world/2020/jun/06/bruno-latour-coronavirus-gaia-hypothesis-climate-crisis

Kennedy, H. (2020). *Simple data visualisations have become key to communicating about the COVID-19 pandemic, but we know little about their impact*. Impact of Social Sciences Blog. https://blogs.lse.ac.uk/impactofsocialsciences/2020/05/04/simple-data-visualisations-have-become-key-to-communicating-about-the-covid-19-pandemic-but-we-know-little-about-their-impact/

Langlais, P.-C. (2015). {{Référence nécessaire}} L'émergence d'une norme wikipedienne. In L. Barbe, L. Merzeau, & V. Schafer (Éds.), *Wikipedia, objet scientifique non identifié*. Presses Universitaires de Paris Ouest (PUPO).

McCormick, S. (2012). After the cap : Risk assessment, citizen science and disaster recovery. *Ecology and society*, *17*(4).

Moats, D. (2019). Following the Fukushima Disaster on (and against) Wikipedia : A Methodological Note about STS Research and Online Platforms. *Science, Technology, & Human Values*, *44*(6), 938–964.

Parasie, S., & Dedieu, F. (2019). À quoi tient la crédibilité des données citoyennes?. L'institutionnalisation des capteurs citoyens de pollution de l'air en Californie. *Revue d'anthropologie des connaissances*, *13*(4), 1035-1062.

Plantin, J.-C. (2015). The politics of mapping platforms : Participatory radiation mapping after the Fukushima Daiichi disaster. *Media, Culture & Society*, *37*(6), 904–921.

Sutton, J., Palen, L., & Shklovski, I. (2008). Backchannels on the front lines : Emergent uses of social media in the 2007 southern California wildfires. *ISCRAM 2008, Proceedings of the 5th International Conference on Information Systems for Crisis Response and Management*, 624–632.

Wiggins, A., & Wilbanks, J. (2019). The rise of citizen science in health and biomedical research. *The American Journal of Bioethics*, *19*(8), 3–14.

Wildschut, D. (2017). The need for citizen science in the transition to a sustainable peer-to-peer-society. *Futures*, *91*, 46-52.

Wolff, A., Gooch, D., Montaner, J. J. C., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics*, *12*(3).