

COMMUNIQUER LES DONNÉES EN TEMPS DE CRISE SANITAIRE

Rapport à destination des acteurs impliqués dans
la production, la diffusion ou la médiatisation de données

Auteurs : Nathalie Pignard-Cheynel, Diego Antolinos-Basso, Adrian Bangerter, Valéry Bezençon, Giuseppe Gruttad'Auria, Adrian Holzer, Abdessalam Ouazki, Michael Puntiroli, Andrew Robotham

Projet PANDA - Pandemic Data
Production, diffusion et compréhension des données en temps de pandémie

Soutenu par le Fonds national suisse (FNS) dans le cadre du Programme national de recherche PNR 80 « Covid-19 et société »

Mai 2026

unine
Université de Neuchâtel



**Fonds national
suisse**



Covid-19 et société
Programme national de recherche

Sommaire

Synthèse du rapport	3
Avant-propos. L'information est insuffisante : nous devons aussi la comprendre, par Samia Hurst-Majno	5
Le projet PANDA	6
Introduction. La pandémie de Covid-19, moment critique pour les données ?	7
Chapitre 1. Ce que les données de pandémie représentent	8
Une donnée n'est pas un simple chiffre	8
Des indicateurs utiles, mais instables	9
Chapitre 2. Produire et faire circuler les données dans un écosystème sous tension	11
Les sources des données en temps de pandémie	11
Une forte concentration autour des sources officielles	12
Journalistes et institutions : dépendance, friction et collaboration	13
Des données publiées aux données exploitables	14
L' <i>Open Government Data</i> : se préparer avant la crise	15
Chapitre 3. Médias et journalistes face à l'urgence des données	17
La pandémie, catalyseur du journalisme de données	17
Des rédactions réorganisées autour des données.....	17
<i>Dashboards</i> et outils du journalisme de données	18
Chapitre 4. Couvrir la pandémie par les données : ce que les médias ont rendu visible	21
Une couverture médiatique fortement quantifiée	21
Chronologies, cartes, tableaux : les formes graphiques de la crise	22
Les cadrages de la pandémie via les indicateurs	24
Chapitre 5. La compréhension des données par les publics en contexte d'incertitude	26
S'informer en pandémie : entre besoin de repères et fatigue informationnelle.....	26
Littératie des données : tous les publics ne lisent pas les chiffres de la même manière	27
Représenter l'incertitude sans affaiblir la confiance	27
Favoriser la transparence et l'accompagnement de la réception des données	29
Concevoir des visualisations adaptées	30
Vers des expériences de lecture adaptatives	32
Chapitre 6. Perspectives pour une future crise sanitaire	35
Mettre en place un système de <i>monitoring</i> public multithématique.....	35
Adapter la communication des données aux publics et aux objectifs.....	36
Structurer les interfaces entre institutions, médias et experts	37
Renforcer la littératie des données chez les professionnels et les publics	37
Conclusion	38

Chapitre 7. Et si une pandémie arrivait aujourd'hui, que ferait-on différemment ?	39
« La donnée ne peut plus rester exclusivement biologique », par Duc-Quang Nguyen.....	40
« La collaboration entre scientifiques, journalistes et gouvernements est centrale », par Mathieu Rebeaud.....	41
« Les données sont celles que l'autorité veut bien communiquer », par Nicolas Willemin	42
« Ouvrir les données publiques n'affaiblit pas les institutions », par Titus Plattner.....	43
« L'architecture institutionnelle reste largement inchangée depuis la pandémie », par Olga Baranova	44
« Les rédactions devraient s'engager dans les outils collaboratifs et <i>open source</i> », par Colin Porlezza	45
« Les médias doivent s'affranchir du travail ingrat de récolte et de mise au propre des données », par Gaël Hürlimann	46
« Le datajournaliste devient un opérateur, capable de combiner des instruments hétérogènes », par Antonin Descampe	47
« La littératie des données exige un changement de culture sociétal fondamental, profond et continu », par Monique Lehky Hagen et Diego Kuonen	48
« Notre responsabilité est de faire preuve d'une extrême prudence, voire d'assumer une part d'incertitude », par Charles-Henry Groult.....	49
« Une lecture brute des chiffres nourrit des interprétations trompeuses », par Valentin Tombez	50
« Il est essentiel de transmettre des chiffres avec l'analyse nécessaire à leur interprétation », par Laurence Senn	51
« L'IA pourrait éclairer de manière significative les décisions dans le cadre d'une future pandémie », par Antoine Flahault	52
Quelques lectures pour aller plus loin	53

Table des tableaux et figures

Tableau 1. 11 catégories de sources de données identifiées dans les datavisualisations Covid-19 publiées par cinq médias de Suisse romande entre janvier 2020 et décembre 2022 (projet PANDA)	12
Figure 1. Extrait du <i>dashboard</i> Covid du <i>Temps</i>	19
Tableau 2. Statistiques sur les corpus du projet PANDA	21
Tableau 3. Répartition des principaux types de datavisualisation, hormis la catégorie « autres » comptant pour 2,9 % des DVU et 1,7 % des MT (données du projet PANDA)	23
Figure 2. Taux de présence hebdomadaire des sous-catégories d'indicateurs sanitaires dans les articles du corpus Covid-19 (par 3 500 signes)	25
Figure 3. Datavisualisation avec intervalles de confiance, utilisée dans le cadre de l'expérience sur les effets de la visualisation de l'incertitude	28
Figure 4. Capture d'écran du dispositif interactif Vizualink élaboré dans le cadre du projet PANDA	33
Figure 5. Capture d'écran du dispositif de <i>chatbot</i> NewsChat élaboré dans le cadre du projet PANDA.....	34
Figure 6. Copie d'écran du tableau de bord Covid-19 proposé sur le site de l'OFSP.....	35

Pour citer ce rapport : Pignard-Cheynel N., Antolinos-Basso D., Bangarter A., Bezençon V., Gruttad'Auria G., Holzer A., Ouazaki A., Puntiroli M., Robotham A. (2026). *Communiquer les données en temps de crise sanitaire. Rapport à destination des acteurs impliqués dans la production, la diffusion ou la médiatisation de données.* Université de Neuchâtel, <https://libra.unine.ch/handle/20.500.14713/100202>

Synthèse du rapport

La pandémie de Covid-19 a fait des **données un élément central de la décision politique, du travail journalistique et de la compréhension publique**. Cas, décès, hospitalisations, taux de reproduction, couverture vaccinale ou occupation des soins intensifs ont structuré la manière dont la crise a été suivie, discutée et gouvernée. Pourtant, une donnée disponible n'est pas nécessairement fiable, réutilisable ou correctement comprise. Les chiffres ne sont pas de simples reflets du réel : ils reposent sur des définitions, des choix de collecte, des contraintes techniques et des arbitrages qui orientent leur interprétation.

Le projet PANDA (*Pandemic Data : Production, diffusion et compréhension des données en temps de pandémie*), mené dans le cadre du PNR 80 du Fonds national suisse, analyse la production, la médiatisation et la réception des données pandémiques. Il s'appuie sur 36 724 articles publiés par cinq médias suisses romands entre janvier 2020 et décembre 2022, dont 813 datavisualisations interactives, sur 34 entretiens avec des journalistes et représentants institutionnels, ainsi que sur des expériences et enquêtes consacrées à la réception des visualisations par les publics.

Ce projet nous rappelle que **les données de pandémie sont des constructions**. Un « cas », un « décès » Covid ou un « lit disponible » semblent être des unités évidentes ; ils dépendent pourtant de définitions, de capacités de collecte, de temporalités de remontée et parfois de choix politiques. Certains indicateurs ont aussi eu une dimension performative : ils ont servi à justifier des restrictions, à fermer ou rouvrir des lieux, à orienter des comportements collectifs. Dans une démocratie semi-directe comme la Suisse, l'accès à des données fiables et contextualisées constitue donc un enjeu sanitaire et démocratique.

La circulation médiatique des données a reposé sur un **faible pluralisme des sources**, avec une forte concentration autour des autorités publiques suisses, représentant près de 65 % des mentions totales. Cette centralité tient au rôle de l'OFSP dans la consolidation fédérale des indicateurs sanitaires, mais elle a rendu

les médias dépendants des modalités de publication retenues par les institutions : formats, temporalités, granularité, archivage, protection des données et possibilités de réutilisation.

Les premiers mois ont révélé les **limites d'un écosystème des données** peu préparé à une demande aussi soudaine de chiffres. Les journalistes ont dû composer avec des données dispersées, parfois publiées sous forme non éditable, sans historique consolidé ou structure stable. Plusieurs rédactions ont constitué leurs propres bases ou se sont appuyées sur des agrégateurs. La mise en place progressive de tableaux de bord et d'accès par API a marqué une amélioration importante dans la gestion des données.

La pandémie a agi comme un **catalyseur pour le journalisme de données**. Les datajournalistes, souvent situés en périphérie des rédactions, se sont rapprochés du cœur de la production éditoriale : ils ont validé des graphiques, identifié des sources, alerté sur les limites de certains indicateurs, automatisé des flux et parfois formé des collègues. Cette reconnaissance s'est accompagnée d'un travail largement invisible : maintenance de tableaux de bord, nettoyage de données, réparation de scripts ou mise à jour continue des visualisations. La pandémie a aussi conduit des journalistes généralistes à travailler avec des données sans y être toujours bien préparés.

L'analyse des contenus montre que **la pandémie a été fortement quantifiée**. Les articles consacrés au Covid-19 contiennent davantage de chiffres que les articles hors Covid, et les indicateurs sanitaires dominent largement. Cas, décès, hospitalisations, soins intensifs, taux de positivité et vaccinations ont structuré une grande partie du récit médiatique, grâce à leur disponibilité, leur actualisation fréquente et leur compatibilité avec les *dashboards*. Cette centralité a toutefois produit un effet de cadrage : ce qui était le plus visible était aussi ce qui était le mieux mesuré, le plus facilement mis à jour et le plus aisément représentable. Des dimensions comme la précarité, l'isolement, le décrochage scolaire, la santé mentale

ou les violences domestiques ont été moins présentes, car plus difficiles à quantifier, moins suivies et moins centrales dans la gestion politique de la crise.

Les visualisations ne sont pas des contenants neutres. Chronologies, cartes, tableaux et graphiques en barres ne rendent pas visibles les mêmes dimensions de la crise. Les chronologies ont soutenu une lecture de la pandémie comme succession de vagues et de mesures. Les cartes ont facilité les comparaisons, avec des risques de perception liés à la taille des territoires. Les tableaux enrichis ont permis de suivre plusieurs indicateurs simultanément et ont été fortement réutilisés. Représenter des données, c'est donc toujours choisir, hiérarchiser et orienter l'attention vers certains aspects de la crise.

Du côté des publics, l'accès aux données ne garantit pas leur compréhension. Lire une courbe, comparer des taux, interpréter une moyenne, comprendre une marge d'erreur ou distinguer corrélation et causalité suppose des compétences inégalement distribuées. Les publics interprètent aussi les données à partir de leur confiance dans les institutions et les médias, de leurs croyances préalables, de leur expérience de la crise. Une même visualisation peut être lue comme une information utile, une preuve de sérieux ou un signal anxiogène voire suspect.

Communiquer l'incertitude est un enjeu central, en particulier en temps de crise. Son affichage peut être perçu comme un signe de rigueur, notamment par les publics les plus familiers des statistiques, mais il peut aussi compliquer la compréhension du message. Il ne s'agit pas de montrer davantage d'information, mais de la rendre lisible : message principal clair, incertitude présentée de manière sélective, explications courtes et accès progressif aux détails méthodologiques. En pratique, la transparence progresse, mais reste incomplète. La mention des sources est presque systématique dans les datavisualisations analysées, mais la transparence méthodologique demeure limitée : beaucoup indiquent une source sans lien vers le jeu de données original, et peu explicitent les choix de traitement, les limites ou les incertitudes. Une communication responsable doit dire ce qui est mesuré, ce qui ne l'est pas, comment les données ont

été produites et ce qu'elles permettent - ou non - de conclure.

Une meilleure communication des données peut aussi passer par des **dispositifs de lecture plus adaptés et interactifs**. Des formats liant texte et graphique, ou permettant à un assistant conversationnel d'expliquer les données d'un article, peuvent améliorer la fluidité de lecture et l'engagement de certains publics, notamment face à des contenus complexes ou jugés peu attractifs. Ces outils doivent toutefois rester encadrés : l'interactivité ne doit pas masquer le message principal, et l'IA générative soulève des questions de crédibilité, de biais, d'hallucinations et de dépendance à l'outil.

Les **perspectives pour une future crise sanitaire** retiennent quatre priorités. Premièrement, préparer en amont un *monitoring* public multithématique, intégrant les données sanitaires, mais aussi les effets sociaux, psychologiques et économiques de la crise. Cela suppose des données réellement réutilisables : structurées, documentées, historisées et ouvertes. Deuxièmement, structurer les interfaces entre institutions, médias, chercheurs et experts, par des briefings techniques, des interlocuteurs identifiés et des canaux d'échange stables. Troisièmement, adapter la communication des données aux publics et aux objectifs, en tenant compte de la littératie des données, de la confiance et des usages. Quatrièmement, faire de la littératie des données une compétence partagée, pour les citoyens comme pour les journalistes, les porte-parole institutionnels, les administrations et les responsables politiques.

La pandémie a montré que les données sont indispensables à la compréhension et à la gouvernance des crises sanitaires, mais qu'elles peuvent aussi être mal produites, contextualisées, visualisées ou comprises. Le principal enseignement du rapport est simple : l'enjeu n'est pas seulement de disposer de plus de données, mais de **construire les conditions sociales, techniques et démocratiques de leur intelligibilité**. En Suisse, cette exigence concerne directement la décision publique, le fédéralisme, la participation démocratique et la confiance.

Avant-propos

L'information est insuffisante : nous devons aussi la comprendre

Par **Samia Hurst-Majno**, vice-présidente de la *Swiss National COVID-19 Science Task Force* pendant la pandémie. Elle était membre du comité de direction du programme *Covid-19 et société* (PNR 80) dont a fait partie le projet PANDA à l'origine de ce rapport.

En juin 2022, alors que la crise pandémique avait déjà été officiellement déclarée comme close, j'ai participé à titre personnel à la deuxième *CoronaMahnwache* (veillée Corona) organisée sur la place fédérale, devant le Parlement à Berne. Depuis le début de la crise, une association citoyenne faisait le tour de la Suisse en allumant devant les parlements cantonaux, et à deux reprises devant le Parlement fédéral, un nombre de bougies correspondant au nombre de victimes du SARS-CoV-2 décédées dans leur canton – ou dans le pays – jusqu'alors. En juin 2022, le bilan national était de près de 14 000 morts : nous étions nombreux, mais cela nous a quand même pris des heures de disposer ces bougies, puis de les allumer, une à une. Des passants sont venus nous aider et c'est de leur réaction que je souhaite vous parler. Ils étaient bouleversés. Ce jour-là, ce que nous avons le plus entendu était : « Je ne me rendais pas compte qu'il y en avait eu autant ».

Les morts, nos médias en avaient pourtant beaucoup parlé : quotidiennement, presque exclusivement, jusqu'au point où il était très vite devenu évident pour la plupart d'entre nous que cette information risquait de devenir la seule et que c'était un problème. Des journaux ont alors commencé à publier des biographies personnelles des victimes. On a parlé, un peu plus, d'autres aspects de la crise. On n'a cependant jamais cessé d'égrener chaque jour le nombre de cas et le nombre de morts. Malgré cela, même là, l'information n'avait pas véritablement été intégrée. La place fédérale couverte de bougies -

techniquement « seulement » une autre manière de présenter les mêmes données - était capable d'arracher des larmes aux personnes qui, quelques minutes auparavant, croyaient déjà savoir ce qu'elles venaient en fait de comprendre.

Quel espoir, donc, pour une intégration plus nuancée et complète de l'information nécessaire à la compréhension publique dans une crise future ? Le projet *Pandemic Data* (PANDA) livre ici des résultats importants. Car oui, bien sûr : chaque message aura des effets différents sur des personnes différentes. Rendre les données disponibles est important, mais pour véritablement parler d'information du public, pour permettre à chacun une compréhension véridique de la situation, c'est clairement insuffisant. Ces données doivent aussi pouvoir être intégrées par des personnes différentes, et les nécessaires variations dans la présentation des données ne doivent jamais se transformer en tentatives de manipulation. Rendre tout cela actionnable est loin d'être évident. Cela requiert des outils permettant aux lecteurs de cibler la présentation des données qui leur importent, ainsi que des connaissances permettant de construire ces outils selon des paramètres pertinents.

Les résultats de ce projet, et les techniques qu'il propose, promettent donc une avancée cruciale : rendre les données plus intelligibles à un plus grand nombre de personnes différentes, lors d'une prochaine crise mais également en temps ordinaire, même lorsque ces données sont complexes et entourées d'incertitude. Ne sous-estimons pas ce progrès. Nous parlons ici de soutien à la citoyenneté, et de démocratie.

Le projet PANDA

Ce rapport présente les résultats du projet PANDA (*Pandemic Data : Production, diffusion et compréhension des données en temps de pandémie*), mené pendant trois ans au sein de l'Université de Neuchâtel dans le cadre du PNR 80 du Fonds national suisse. Il s'adresse aux journalistes, aux administrations, aux décideurs et à toute personne impliquée dans la production, le traitement, l'interprétation ou la diffusion des données en contexte de crise sanitaire.

Le projet vise à comprendre comment, en période de pandémie, les données sont produites et mises en récit, notamment par les journalistes, mais aussi comment elles sont reçues par les publics. Il examine le rôle de la confiance dans les médias, de la littératie des données et des croyances préexistantes dans l'interprétation des visualisations et des indicateurs.

Trois questions structurent le projet : comment les données sont-elles produites et mises en récit par les professionnels de l'information ? Dans quelle mesure les publics interprètent-ils ces données différemment selon leurs dispositions et leurs compétences ? Comment améliorer la présentation des données afin d'en faciliter la compréhension ?

Le projet s'appuie sur une analyse des productions médiatiques de cinq médias suisses romands entre janvier 2020 et décembre 2022, soit 36 724 articles

dont 813 datavisualisations interactives liées à la pandémie. Il mobilise également 34 entretiens, dont 27 en rédaction et 7 auprès d'institutions suisses, trois expériences sur la réception des datavisualisations, une enquête sur un échantillon représentatif de la Suisse romande portant sur les attitudes des individus face aux datavisualisations ainsi qu'une exploration de nouvelles formes de narration de données interactives, testées auprès d'utilisateurs.

Le rapport ne cherche pas à documenter exhaustivement ce qui s'est passé pendant la pandémie de Covid-19. Il retient les résultats les plus utiles pour mieux préparer à une éventuelle future crise sanitaire : comment produire des données plus fiables, les rendre plus accessibles, les contextualiser sans les affaiblir et permettre à des publics différents de les comprendre sans les réduire à des chiffres isolés ?

Le propos se veut didactique, ancré dans les résultats du projet, mais aussi dans d'autres travaux scientifiques mobilisés à des fins de mise en perspective. Le dernier chapitre rassemble une quinzaine de contributions externes, que nous avons sollicitées pour leur expérience et leur expertise. Ces courts textes fournissent un regard complémentaire à celui des chercheurs, orienté vers l'avenir.

L'équipe du projet

Prof. Nathalie Pignard-Cheyne, Académie du journalisme et des médias, UniNE, Direction de recherche
 Prof. Valéry Bezençon, Institut de management, UniNE, Co-requérant
 Prof. Adrian Holzer, Institut du management de l'information, UniNE, Co-requérant
 Dr Michael Puntiroli, Institut de management, UniNE, Co-requérant
 Prof. Adrian Bangerter, Institut de psychologie du travail et des organisations, UniNE, Conseiller de projet
 Dr Andrew Robotham, Académie du journalisme et des médias, UniNE, Postdoctorant
 Diego Antolin-Basso, Académie du journalisme et des médias, UniNE, Collaborateur scientifique
 Giuseppe Gruttad'Auria, Institut de management, UniNE, Doctorant
 Céline Dupuis, Académie du journalisme et des médias, UniNE, Doctorante (04.2023-07.2024)
 Abdessalam Ouaazki, Institut du management de l'information, UniNE, Collaborateur scientifique
 Flavia Gillioz, Académie du journalisme et des médias, UniNE, Collaboratrice scientifique (09.2024-06.2025)
 Jean Friedrich, Académie du journalisme et des médias, UniNE, Collaborateur scientifique (02-07.2025)

Introduction. La pandémie de Covid-19, moment critique pour les données ?

Plusieurs chercheurs en journalisme ont qualifié la pandémie de Covid-19 de *critical incident*, au sens de moment pivot durant lequel les routines, les rôles, les normes et les frontières mêmes du journalisme ont été discutés et renégociés. Plutôt qu'une rupture nette, il s'agit d'un moment de tension inscrit dans des dynamiques préexistantes, sans que l'on puisse nécessairement conclure à une transformation profonde et durable du champ journalistique.

Lire la pandémie comme un moment de reconfiguration permet aussi de penser, au-delà du seul journalisme, son impact sur l'ensemble de l'écosystème des données, devenu entre 2020 et 2023 un enjeu stratégique et politique. La pandémie est en effet survenue dans des sociétés déjà engagées dans un processus de datification, c'est-à-dire la traduction croissante de dimensions de la vie sociale en données quantifiables, comparables et exploitables à des fins de connaissance et d'action. La crise sanitaire a accéléré cette dynamique de trois manières : par la quantification des comportements et des mobilités, par le recours accru à la modélisation prédictive, et par la diffusion en temps quasi réel d'indicateurs sanitaires tels que les cas, les hospitalisations, les décès, le taux de reproduction ou la couverture vaccinale. Ces transformations ont modifié la temporalité de l'information sanitaire, la perception publique de la crise et le rapport des journalistes comme des institutions aux données.

Cette centralité nouvelle des données a toutefois produit deux dynamiques ambivalentes. D'un côté, elle a pu renforcer une forme de dataïsme, soit la

croyance dans la capacité des données à produire une représentation objective et neutre du réel. La multiplication des indicateurs, des courbes et des tableaux de bord a parfois contribué à naturaliser des choix méthodologiques et à rendre moins visibles les conditions sociales, institutionnelles et techniques de production des chiffres.

De l'autre côté, la pandémie a nourri une montée de la suspicion et des controverses à l'égard des données. Les révisions de séries statistiques, les changements de définition, les divergences entre institutions ou les débats autour des décès « avec » ou « dus » au Covid-19 ont fragilisé la confiance d'une partie des publics. Dans un écosystème informationnel marqué par les plateformes numériques, les données et les visualisations ont aussi circulé hors de leur contexte d'origine, parfois réinterprétées, contestées ou mobilisées dans des contre-discours très divers. La donnée n'a donc pas seulement été un outil de connaissance et de gouvernement ; elle est aussi devenue un objet de lutte symbolique et politique.

Une question demeure : que subsiste-t-il de ces transformations dans le « monde d'après » ? Quels apprentissages institutionnels, professionnels et collectifs ont été consolidés ? La pandémie a-t-elle constitué un point de bascule durable dans notre culture des données, ou plutôt une parenthèse exceptionnelle dont les effets se sont partiellement dissipés ? C'est à partir de ces interrogations que ce rapport analyse la production, la médiatisation et la réception des données en temps de crise sanitaire.

Chapitre 1. Ce que les données de pandémie représentent

Avant d'examiner qui produit les données, comment elles circulent et comment elles sont médiatisées, il faut d'abord revenir sur ce qu'elles représentent. Les chiffres de pandémie ne sont pas de simples unités comptables : ils traduisent des définitions, des choix de mesure, des contraintes techniques et des arbitrages politiques. C'est ce caractère construit des données que le premier chapitre propose d'éclairer.

Une donnée n'est pas un simple chiffre

La fonction première de la donnée statistique est de traduire une propriété ou un phénomène du monde réel en quelque chose de quantifiable et comparable. En contexte de pandémie, cette fonction prend une dimension toute particulière : la donnée ne se contente pas de décrire la situation, elle devient un outil d'action. Comprendre comment elle est construite, par qui et pourquoi, est donc essentiel pour saisir comment la quantification hérite et incarne des systèmes de valeurs, des contraintes techniques et des jeux de pouvoir.

L'expression « flatten the curve » (« aplatis la courbe »), popularisée dès mars 2020 aux États-Unis, est un bon exemple : la formule n'invitait pas seulement à observer la propagation du virus, mais à agir sur la courbe pour transformer le réel. Sa force visuelle a aussi montré une limite : une courbe peut rendre la crise intelligible tout en effaçant la pluralité des conséquences humaines, sociales et économiques qu'elle recouvre.

Les données de pandémie ne sont jamais de simples reflets du réel. Un « cas », un « décès Covid » ou un « lit disponible » semblent être des unités évidentes ; ils reposent pourtant sur des définitions, des conventions de collecte, des choix de publication et des arbitrages techniques qui influencent leur comparabilité, leur interprétation et parfois leur usage politique. La donnée statistique quantitative a quelque chose de rassurant : elle semble propre, « objective », incontestable. Pourtant, derrière un

chiffre aussi simple que « 170 décès aujourd'hui en Suisse » se cachent des choix méthodologiques complexes. Comment centraliser les informations de 26 cantons ? Comment intégrer les décès découverts plusieurs jours après les faits ? Comment traiter les erreurs, les retards ou les changements de définition ? Chaque unité statistique est le résultat d'un processus impliquant des définitions, des conventions et des risques d'erreur.

Dans le débat politique durant la pandémie, des appels à des chiffres « plus objectifs » ont été formulés. Ces demandes pouvaient relever d'une exigence légitime de précision statistique, mais elles traduisaient aussi parfois la recherche d'indicateurs confortant une lecture particulière de la pandémie et des mesures jugées nécessaires ou disproportionnées. Elles montrent surtout que les catégories et les modes de classification font eux-mêmes l'objet de débats. Un exemple emblématique est la controverse autour des décès « dus au Covid-19 » ou « avec le Covid-19 », souvent mobilisée dans des discours contestant la dangerosité du virus ou la proportionnalité des mesures sanitaires. Si cette distinction peut paraître légitime, sa mise en pratique avec une certitude totale est extrêmement difficile : une personne jeune et sans antécédents connus peut présenter des facteurs de comorbidité non diagnostiqués ; à l'inverse, une personne en fin de vie peut voir son état aggravé de manière déterminante par une infection. Vouloir des chiffres « plus objectifs » sur la cause des décès revient donc parfois à exiger une donnée encore plus complexe à produire que celle que l'on conteste.

Les chiffres sont toujours à la fois moins et plus que ce qu'ils mesurent. Moins, parce qu'ils ne restituent jamais toute la complexité du réel ; plus, parce qu'ils deviennent, selon les contextes, des arguments, des proxys ou des instruments d'action publique.

Cas, hospitalisations, lits disponibles en soins intensifs, décès, R_e : tous ces indicateurs ont traduit à

un moment, et à leur manière, la gravité de la situation pandémique. Ces indicateurs ont aussi eu une dimension performative : ils ont contribué à évaluer le risque, à justifier des restrictions, à fermer ou rouvrir des écoles, des restaurants ou des lieux de travail. Leur pouvoir ne tient donc pas seulement à ce qu'ils décrivent, mais également à l'autorité institutionnelle et politique qui leur a été accordée. Dans une démocratie semi-directe comme la Suisse, l'accès compréhensible aux informations (et donc aux données) nécessaires pour se forger une opinion, voter et élire constitue une condition centrale du débat démocratique.

Des indicateurs utiles, mais instables

Les cas et les décès, parfois qualifiés de « météo Covid », ont occupé une place centrale tout au long de la crise. Nos recherches montrent qu'ils font l'objet de 19 % des représentations graphiques produites par les médias et sont très présents dans les textes des articles.

Pour une analyse complète de la couverture médiatique via les données, voir le chapitre 4

S'ils paraissent simples à comprendre, ils sont loin d'être simples à produire de manière fiable et en temps réel. Ainsi, la définition d'un « cas » est moins évidente qu'il n'y paraît. En théorie, un test positif correspond à un cas. En pratique, plusieurs facteurs compliquent ce décompte : capacités de test insuffisantes lors de certaines phases de la pandémie, absence de signalement de nombreux tests réalisés à domicile, doublons possibles lorsqu'une même personne effectue plusieurs tests positifs, ou encore renoncement au test par crainte de l'isolement. Ces difficultés ont probablement conduit à une sous-représentation des infections réelles. Le taux de positivité est alors devenu un indicateur complémentaire important pour estimer l'écart entre les cas signalés et les infections effectives. Plus tard, la mesure du pathogène dans les eaux usées a permis de suivre la circulation du virus en partie indépendamment du recours individuel au test.

Le cadre législatif

Les deux principaux indicateurs que sont les cas et les décès peuvent sembler aller de soi. Leur collecte repose sur un cadre légal précis, la Loi sur les épidémies (LEp), entrée en vigueur en 2016. Celle-ci prévoit un « système d'information » visant à identifier les personnes malades et à mettre en place des mesures (art. 60 LEp). En revanche, ce dispositif n'énonce pas explicitement une obligation générale de publication de ces données à destination de la population suisse.

La fonction structurante du cadre législatif pour la collecte des données sanitaires s'étend également aux données hospitalières. Avec l'Ordonnance 2 Covid-19 du 13 mars 2020, les cantons se voient imposer la communication d'informations sur l'occupation et la disponibilité des lits hospitaliers et des soins intensifs. À partir de ces données primaires sont ensuite élaborés divers indicateurs secondaires, tels que le taux de létalité (qui diverge selon le variant) ou encore le taux de reproduction effectif R_e .

Un autre enjeu lié aux indicateurs est celui de leur comparabilité. Dans une pandémie qui n'a pas de frontière, comparer des régions géographiques distinctes est essentiel. Mais même pour des régions utilisant des normes a priori identiques, des différences administratives, de protocoles ou de méthodes peuvent rendre un même indicateur difficilement comparable.

Certains indicateurs ont connu une visibilité brève. C'est le cas du nombre de guérisons, très présent lors des premières semaines, puis rapidement marginalisé lorsque le volume de nouveaux cas a rendu ce suivi moins lisible et moins utile. L'attention s'est alors déplacée vers d'autres mesures : hospitalisations, occupation des soins intensifs, décès, mais aussi nombre de personnes considérées comme immunisées. Cet exemple montre qu'un indicateur n'a de sens qu'en contexte.

Au début de la pandémie, les personnes guéries étaient souvent considérées comme protégées contre

une nouvelle infection et comme contribuant à freiner la circulation du virus. L'arrivée des vaccins, puis les réinfections liées aux variants, ont progressivement complexifié cette lecture : un même chiffre pouvait mesurer une protection individuelle, signaler une dynamique collective et nourrir un récit de sortie progressive de crise. La capacité des chiffres à désigner plusieurs réalités à la fois est devenue particulièrement visible durant la pandémie. Le nombre de personnes en quarantaine, par exemple, ne renseignait pas seulement sur la circulation du virus. Il indiquait aussi un ralentissement social et économique : des personnes absentes de l'école ou du travail, des foyers isolés, des proches ou voisins mobilisés pour livrer des courses, et plus largement une société contrainte de réorganiser ses activités quotidiennes.

Moins présents dans le récit dominant, des indicateurs non sanitaires ont également été mobilisés durant la pandémie. Les données économiques, comme le taux de chômage ou le PIB, ont parfois été opposées aux indicateurs sanitaires pour discuter la proportionnalité des mesures. Les sondages et les résultats de votations ont, de leur côté, servi à quantifier les attitudes de la population à l'égard de la gestion de crise, notamment lors des trois référendums liés au cadre proposé par le gouvernement. Ces indicateurs sont toutefois restés secondaires par rapport aux données sanitaires et n'ont pas fait l'objet, pour la plupart, d'une collecte aussi systématique par les autorités. Or, lorsque certaines réalités ne sont pas traduites en chiffres officiels, ou lorsque leur mesure est coûteuse et fragmentaire, elles peinent à exister dans le débat public. Le manque de statistiques sociales a ainsi contribué à rendre moins visibles des dimensions pourtant centrales de la pandémie : pauvreté, détresse psychologique, violence domestique, décrochage scolaire ou isolement. Des associations et ONG ont parfois tenté de documenter ces réalités avec des moyens limités, mais sans disposer du même poids institutionnel que les indicateurs sanitaires.

Duc-Quang Nguyen, « La donnée ne peut plus rester exclusivement biologique » (page 40)

Le niveau de précision et de granularité des données (la finesse du niveau de détail) constitue également un enjeu important. Plus une donnée est détaillée (par âge, territoire, nationalité, profession ou statut socio-économique), plus elle permet des analyses fines et, parfois, des mesures ciblées. Mais cette granularité accroît aussi les risques d'interprétation politique, morale ou stigmatisante. Les données de mortalité par âge, par exemple, ont permis de mieux identifier les populations les plus exposées ; elles ont aussi pu être mobilisées pour discuter, voire contester, la proportionnalité de certaines mesures sanitaires (par exemple envers les personnes âgées). De même, croiser des indicateurs sanitaires avec des variables socio-économiques ou migratoires peut éclairer des inégalités d'exposition, mais aussi favoriser des lectures abusives.

La granularité n'est donc utile que si elle s'accompagne d'un travail de contextualisation. Sans cela, elle peut donner une impression de précision tout en masquant les mécanismes sociaux qui expliquent les écarts observés. Elle rappelle aussi que les données quantitatives ne peuvent pas constituer la seule base de compréhension d'une crise.

Ce chapitre invite donc à décentrer le regard : si les données ne sont jamais le reflet direct de la réalité, la pandémie l'a rendu particulièrement visible. En situation de crise, les chiffres ne se contentent pas de mesurer : ils orientent l'attention, structurent les débats et peuvent justifier des décisions collectives. Leur utilité dépend donc des conditions dans lesquelles ils sont produits, définis, comparés, contextualisés et rendus publics. Avant de demander ou produire davantage de données, il faut comprendre ce qu'elles mesurent vraiment, ce qu'elles laissent dans l'ombre et comment elles peuvent être mobilisées dans l'action publique comme dans le débat démocratique.

Chapitre 2. Produire et faire circuler les données dans un écosystème sous tension

Cette construction des données ne relève toutefois pas d'un acteur unique. Elle s'inscrit dans un écosystème composé d'administrations, de cantons, d'institutions de santé, de chercheurs, d'agrégateurs, de médias, d'acteurs privés et de publics. Ce chapitre examine donc la manière dont les données ont été produites, consolidées, partagées, reprises et parfois contestées dans cet écosystème sous tension.

Les sources des données en temps de pandémie

L'écosystème des données constitue un environnement hybride, composé d'acteurs publics,

scientifiques, médiatiques, privés et de la société civile. Pendant la pandémie, aucun de ces acteurs n'a maîtrisé seul la chaîne des données : les informations ont été produites, consolidées, interprétées, visualisées, contestées et relayées dans une série d'interdépendances.

Le tableau ci-dessous repose sur l'analyse des sources mentionnées dans un corpus de 813 datavisualisations interactives consacrées au Covid-19, extraites de 36 724 articles publiés par cinq médias suisses romands entre le 23 janvier 2020 et le 26 décembre 2022, puis normalisées et classées en 11 catégories de sources.

Catégorie de source	Rôle dans la chaîne des données	Exemples	% MT*	% DVU*
Autorités publiques suisses	Elles produisent, centralisent et publient les principaux indicateurs sanitaires, démographiques, administratifs ou économiques. Elles définissent aussi les catégories, les seuils et les modalités de collecte qui structurent les usages ultérieurs des données.	OFSP, OFS, cantons, SECO, Administration fédérale des finances, communes	64,47 %	46,22 %
Intermédiaires et agrégateurs de données	Ils collectent, harmonisent et republient des données produites par d'autres acteurs, notamment lorsque celles-ci sont fragmentées, hétérogènes ou difficilement exploitables.	Our World in Data, Johns Hopkins University, Corona-data.ch, icumonitoring.ch	10,83 %	7,67 %
Entités inter- ou supranationales	Elles produisent ou consolident des données à l'échelle internationale, souvent utilisées pour comparer les situations nationales ou situer la Suisse dans un contexte plus large.	OMS, Commission européenne, OCDE	10,16 %	3,33 %
Instituts de sondage	Ils produisent des données d'opinion, d'attitudes et de perceptions sociales, notamment sur la gestion de la crise, les mesures sanitaires ou les votations.	Sotomo, LeeWas, gfs.bern	5,63 %	16,11 %
Médias et production interne	Ils collectent, retraitent, agrègent ou (co)produisent des données, parfois à partir de sources externes, parfois à travers leurs propres enquêtes ou dispositifs éditoriaux.	Médias qui traitent ou produisent eux-mêmes des données	2,68 %	8,11 %
Recherche académique et	Elle produit des connaissances, des modèles, des résultats quantifiés ou des bases de données	Articles scientifiques, prépublications, bases de données	2,31 %	6,44 %

Catégorie de source	Rôle dans la chaîne des données	Exemples	% MT*	% DVU*
publications scientifiques	spécialisées ; contribue aussi à contextualiser des indicateurs complexes.			
Entreprises privées	Elles produisent des données issues de leurs services, plateformes ou infrastructures, par exemple sur la mobilité, les comportements numériques ou certains effets économiques de la crise.	Google, Apple, opérateurs de mobilité, plateformes numériques	1,61 %	5 %
Autorités publiques étrangères	Elles publient des données nationales mobilisées pour suivre la situation dans d'autres pays ou établir des comparaisons internationales.	Ministères de la santé, agences sanitaires nationales, gouvernements étrangers	1,11 %	3,11 %
Organisations socio-économiques	Elles produisent des données sectorielles sur les effets économiques, professionnels ou sociaux de la crise.	Fédérations professionnelles, syndicats, organisations économiques	0,54 %	1,78 %
Institutions de santé	Elles font remonter ou publient des données cliniques, hospitalières ou de surveillance sanitaire, souvent en lien avec les capacités du système de santé.	Hôpitaux, réseaux hospitaliers, systèmes de surveillance clinique	0,47 %	1,56 %
Société civile	Elle produit, relaie ou recontextualise des données hors des circuits institutionnels classiques, parfois dans une logique de vigilance, de contestation ou de documentation alternative.	Initiatives citoyennes, collectifs, associations, individus	0,36 %	1,20 %

*La part des mentions totales (% MT) tient compte des réutilisations d'une même visualisation dans plusieurs articles ; la part des visualisations uniques (% DVU) ne compte chaque source qu'une seule fois par visualisation.

Tableau 1. 11 catégories de sources de données identifiées dans les datavisualisations Covid-19 publiées par cinq médias de Suisse romande entre janvier 2020 et décembre 2022 (projet PANDA)

Une forte concentration autour des sources officielles

Nos résultats s'inscrivent dans le prolongement de travaux montrant que, durant la pandémie, les médias se sont fortement appuyés sur des sources officielles et gouvernementales. Notre analyse des datavisualisations romandes confirme ce constat, mais permet surtout de mieux comprendre la structure de l'écosystème des données qui s'est imposé pendant la crise. Malgré une apparente diversité de sources (on en compte 175 distinctes), la circulation médiatique des données a reposé sur un nombre restreint d'acteurs, principalement publics (les autorités publiques suisses représentent près de 65 % de toutes les mentions de sources dans le corpus

étudié), capables de produire, centraliser, stabiliser et publier des indicateurs suivis dans le temps.

Cette configuration tient en partie au contexte suisse. Dans un système fédéral où les cantons jouent un rôle important dans la collecte et la transmission des informations sanitaires, la lisibilité publique des données dépend fortement de leur capacité à être consolidées au niveau national, puis rendues accessibles dans des formats réutilisables. L'OFSP s'est ainsi imposé comme un point de passage central, non seulement parce qu'il détenait les principaux indicateurs sanitaires, mais aussi parce que

ses données permettaient aux médias de suivre la crise de manière régulière, comparable et actualisable. Cette centralité a rendu les médias fortement dépendants des modalités de publication retenues par les instances productrices de données : formats, temporalités, niveaux de détail et possibilités de réutilisation.

Le recours à des agrégateurs et à des plateformes intermédiaires montre que les journalistes ont aussi cherché des données déjà traitées, comparables ou directement exploitables, notamment pour les comparaisons internationales ou le suivi de séries hétérogènes. À l'inverse, la faible présence des sources académiques, de la société civile ou des organisations socio-économiques ne signifie pas nécessairement qu'elles étaient absentes de la compréhension de la crise. Elle indique plutôt que leurs données étaient moins souvent disponibles sous une forme directement intégrable dans des datavisualisations journalistiques, moins régulièrement actualisées, ou moins adaptées à une republication continue. L'un des enseignements importants de la pandémie est que ce qui devient visible publiquement n'est pas seulement ce qui est pertinent, mais aussi ce qui est produit, structuré et diffusé dans des conditions permettant sa réutilisation rapide.

Mathieu Rebeaud, « La collaboration entre scientifiques, journalistes et gouvernements est centrale » (page 41)

Journalistes et institutions : dépendance, friction et collaboration

Au-delà de la domination des sources officielles, les entretiens menés auprès de journalistes et de représentants institutionnels, notamment à l'OFSP, permettent de mieux comprendre les relations concrètes entre rédactions et producteurs de données pendant la pandémie. Ces relations n'ont été ni simplement dépendantes, ni totalement conflictuelles. Elles ont oscillé entre coopération, asymétrie d'accès et frictions, selon les moments de

la crise, les types de données concernées et les profils journalistiques.

Les journalistes interrogés ont surtout mis en avant trois types d'obstacles. Le premier concerne l'accès aux données, particulièrement difficile durant les premiers mois de la pandémie : dispersion des sources, absence d'historiques consolidés et modalités de publication instables ont compliqué le travail des rédactions. Le deuxième obstacle tient à la cohérence des données. Avant la coordination assurée par OpenZH (une initiative *open data* portée par le canton de Zurich pour agréger et harmoniser les données Covid-19 publiées par les cantons), puis par l'OFSP, les divergences entre cantons rendaient certaines séries difficilement comparables et ralentissaient leur exploitation médiatique. Le troisième concerne la granularité : plusieurs journalistes, en particulier les spécialistes des données, ont regretté de ne pas disposer d'un niveau de détail suffisant pour mener des analyses plus fines. Plus largement, les relations entre journalistes et autorités ont parfois été marquées par des frictions, parfois interprétées par les journalistes comme une volonté de contrôler les données, de masquer des erreurs ou de minimiser des réalités inconfortables. Les institutions invoquent de leur côté une forme de débordement lié à la soudaineté de la crise, impliquant une importante surcharge de travail mais aussi des contraintes techniques et juridiques liées par exemple à la protection des données.

Ces obstacles ont eu un effet direct sur le travail journalistique. Les journalistes de données ont consacré un temps considérable à collecter, vérifier, nettoyer, standardiser et archiver des données qui auraient idéalement dû être mises à disposition par les institutions productrices elles-mêmes. Cette situation pointe une tension structurelle du journalisme de données en contexte de crise : le temps consacré au traitement technique réduit celui disponible pour l'analyse, l'interprétation et la mise en perspective, qui constituent pourtant la valeur ajoutée journalistique. À l'inverse, cette immersion forcée dans les données a aussi renforcé l'expertise critique de certains journalistes, capables de

comparer les jeux de données, d'en repérer les limites et d'arbitrer entre sources, par exemple en privilégiant progressivement les statistiques de l'OFSP à celles de l'OFSP pour le traitement des décès. Ce sont également ces journalistes plus spécialisés qui ont eu recours à des sources complémentaires comme les agrégateurs (internationaux mais aussi nationaux), les publications scientifiques ou des données non sanitaires (par exemple les statistiques douanières pour vérifier les affirmations sur les pénuries de masques).

Le rapport aux sources officielles a donc varié selon les profils. Les journalistes généralistes, de desk et d'agences ont souvent dû s'appuyer sur les données administratives faute d'alternative, dans un contexte de production sous tension. Les datajournalistes et les journalistes scientifiques ont exprimé davantage de distance critique, jouant parfois un rôle d'alerte au sein des rédactions. Mais même ces profils spécialisés ont dû, à certains moments, faire confiance aux données officielles ou renoncer à certains sujets faute de données suffisamment solides.

Nicolas Willemin, « Les données sont celles que l'autorité veut bien communiquer » (page 42)

Les experts ont enfin joué un rôle central de traduction et de validation. Dans un contexte d'incertitude scientifique, des épidémiologistes, virologues ou spécialistes des données ont aidé les journalistes à interpréter des indicateurs complexes, à identifier les jeux de données pertinents ou à valider certains calculs. Cette collaboration a parfois dépassé le simple recueil de citations pour devenir un travail de fond, certains experts agissant comme interlocuteurs réguliers ou « *sparring partners* » des rédactions. Les réseaux informels ont également compté : médecins cantonaux, responsables hospitaliers, conseillers d'État ou agents des institutions publiques ont parfois permis d'obtenir rapidement des explications, de contextualiser des chiffres ou d'anticiper certaines décisions. Ces circuits parallèles montrent que, même dans une crise très quantifiée, l'accès aux données ne dépendait pas seulement de fichiers et d'API, mais aussi de relations

professionnelles, de confiance et de capacité à identifier les bons interlocuteurs.

Une collaboration inédite autour des données communales

La collaboration entre la NZZ et l'OFSP autour des cas de Covid-19 à l'échelle communale illustre les tensions entre transparence, granularité et protection des données. L'OFSP refusait de transmettre les données brutes des cas de Covid à l'échelle communale en Suisse, en raison du risque d'identification des personnes dans les zones faiblement peuplées. L'OFSP a toutefois accepté une forme de collaboration inédite : plutôt que de transmettre les données, il a fait tourner sur ses propres serveurs un script fourni par le journal suisse alémanique, permettant de représenter les cas par regroupement de niveaux d'incidence et de faire apparaître les différences entre communes, sans exposer les valeurs brutes.

Cet exemple, souvent cité dans nos entretiens, montre qu'une coopération technique entre institutions et médias peut permettre de concilier intérêt public, analyse journalistique et protection des données, même si cette solution n'a pas été généralisée à l'ensemble des rédactions.

Des données publiées aux données exploitables

La pandémie a révélé les limites d'un appareil d'État qui n'était pas préparé à répondre à une demande aussi soudaine et massive de données actualisées, comparables et réutilisables.

Les premiers mois ont été particulièrement révélateurs. Les problèmes liés à la transmission par fax des dénombrements de cas, en partie dus au manque d'équipement numérique de certains cabinets médicaux, n'étaient qu'un aspect du problème : les journalistes ont aussi déploré la publication de données sous forme de PDF non éditables, parfois remplacés chaque jour sans archivage organisé. Plusieurs rédactions ont ainsi constitué des bases de données de substitution. La mise en place progressive de tableaux de bord de

données (*dashboards*) et d'API a marqué une amélioration importante, mais elle est intervenue après une phase de bricolage coûteuse et chronophage pour les rédactions.

Titus Plattner, « Ouvrir les données publiques n'affaiblit pas les institutions » (page 43)

Cette situation rappelle qu'une donnée rendue publique n'est pas nécessairement une donnée exploitable. Les formats ouverts et structurés, comme le CSV, le JSON, le XML ou le HTML correctement balisé, facilitent la vérification, l'automatisation et le croisement des sources. À l'inverse, les fichiers PDF, les tableaux non standardisés ou les formats propriétaires compliquent l'extraction et augmentent le risque d'erreurs. Au-delà du format, la stabilité de la structure des données est tout aussi déterminante : une simple modification de colonne, de définition ou de fréquence de mise à jour peut casser des scripts d'automatisation et rendre une série difficilement exploitable dans le temps.

L'interopérabilité est d'ailleurs une condition centrale du journalisme de données en contexte de crise. Pour analyser la pandémie, les journalistes n'ont pas seulement suivi les cas, les décès ou l'occupation hospitalière ; ils ont croisé ces indicateurs avec des données de mobilité, de vaccination, d'opinion, d'économie ou de politiques publiques. Cette capacité à relier des sources hétérogènes repose sur des formats standards et ouverts, des identifiants stables ainsi qu'une documentation accessible et rigoureuse des données et des outils de production.

L'Open Government Data : se préparer avant la crise

La mise à disposition proactive de données gouvernementales structurées et accessibles s'inscrit dans le prolongement des principes de transparence et d'accès à l'information. L'*Open Government Data* (OGD) en constitue une déclinaison spécifique : il porte sur des données quantitatives et structurées, et suppose une démarche de publication proactive

plutôt qu'une simple réponse à des demandes ponctuelles.

En Suisse, plusieurs dispositifs vont dans ce sens. La LTrans pose le principe d'une administration fondée sur la transparence, tandis que la LMETA renforce l'exigence de publication proactive de données lisibles par machine. Le portail opendata.swiss, créé en 2016, constitue le point d'entrée central des données ouvertes de l'administration publique suisse, tandis que le [Masterplan Open Government Data 2024-2027](#) fixe des objectifs en matière de disponibilité des données, de qualité des métadonnées, d'infrastructures, de synergies avec la recherche et d'échanges entre acteurs. Ces avancées restent toutefois incomplètes : selon le rapport *Open Data Maturity 2025*, la Suisse se situe encore dans la catégorie des « suiveurs », au 21^e rang sur 36 pays évalués. Le fédéralisme constitue également une difficulté spécifique, dans la mesure où les pratiques liées aux données peuvent varier fortement entre Confédération, cantons, communes et entités parapubliques.

La pandémie l'a clairement montré : les infrastructures et l'implémentation de mesures en matière d'OGD doivent être pensées en temps normal, car il est très difficile de les construire en contexte de crise. L'enjeu n'est pas seulement d'ouvrir davantage de données, mais de les rendre effectivement réutilisables : structurées, documentées, historisées, publiées dans des formats ouverts, accompagnées de licences claires et suffisamment stables pour permettre leur vérification, leur croisement et leur mise en contexte. Pour les institutions, cela suppose d'anticiper les usages professionnels des données, notamment par les journalistes, les chercheurs, les acteurs de santé publique ou les organisations de la société civile. Pour les rédactions, cela suppose de disposer des compétences nécessaires pour vérifier, comparer et interpréter ces données sans se limiter à leur reprise.

Olga Baranova, « L'architecture institutionnelle reste largement inchangée depuis la pandémie » (page 44)

La diffusion des données s'est enfin heurtée à une tension récurrente entre transparence publique et obligations légales, notamment en matière de protection des données personnelles. Le principal risque mis en avant par les institutions pour justifier de la rétention de données était celui d'identification : la publication de données granulaires pouvait permettre de remonter à des individus précis. Ce risque, important à des petites échelles, ne disparaissait pas avec des populations plus importantes, car des tiers disposant déjà de bases de données (assureurs, par exemple) pouvaient croiser les informations publiées par l'État pour reconstituer des profils de santé. L'échelle locale a été particulièrement impactée par cette limite. Si les

cantons disposaient des données granulaires à des fins de pilotage opérationnel, ces dernières n'étaient pas pour autant rendues publiques. La législation a aussi empêché le croisement de fichiers : l'OFSP ne pouvait pas coupler les données de vaccination avec les déclarations de cas, ce qui a par exemple limité l'analyse en temps réel de l'efficacité vaccinale.

Ces contraintes ont régulièrement été interprétées par les journalistes comme une volonté de contrôler l'information, signe que les institutions ont parfois manqué de pédagogie sur les risques juridiques et de souplesse dans la recherche de solutions intermédiaires.

Chapitre 3. Médias et journalistes face à l'urgence des données

Après avoir examiné les conditions de production et de circulation des données, ce chapitre s'intéresse à leur traitement au sein des rédactions. La pandémie a placé les journalistes face à une double contrainte : répondre à une demande d'information immédiate et massive, tout en travaillant avec des données instables, parfois incomplètes et techniquement difficiles à exploiter. Elle a ainsi fait du journalisme de données un espace central de la couverture médiatique, mais aussi un lieu de tensions, d'apprentissages et de transformations professionnelles.

La pandémie, catalyseur du journalisme de données

Avant d'examiner la manière dont les données ont transformé les pratiques journalistiques, il faut rappeler que la pandémie a profondément bouleversé les conditions ordinaires de production de l'information. Les journalistes ont travaillé dans un contexte d'urgence, d'incertitude scientifique, de réorganisation des rédactions, de télétravail contraint, de pression économique mais aussi de forte exposition émotionnelle. Plusieurs travaux internationaux ont montré que la couverture du Covid-19 a renforcé les rôles de service, d'éducation et de facilitation entre experts, autorités et publics, parfois au détriment d'une posture plus critique de « chien de garde » vis-à-vis des institutions.

Les entretiens menés dans le cadre de notre projet auprès de journalistes de plusieurs rédactions romandes convergent sur un point : la pandémie a agi comme un puissant catalyseur pour le journalisme de données, et plus largement pour l'usage des données dans les rédactions, moment que certains qualifient même d'« âge d'or ». La donnée est passée du statut d'illustration secondaire, ou de matériau réservé à des enquêtes spécifiques, à celui de matière première indispensable pour couvrir la crise, y compris en temps réel. De manière rapide, voire soudaine, les rédactions ont dû se réorganiser, souvent de manière

informelle, pour faire face à cette « datification » de la pandémie.

Ce déplacement a modifié la place des datajournalistes. Dans plusieurs grandes rédactions, ces profils spécialisés existaient déjà, mais ils travaillaient souvent dans de petites équipes situées en périphérie de la rédaction principale. La pandémie les a rapprochés du cœur de la production éditoriale. Ils ont contribué à valider des graphiques, à initier des sujets fondés sur les données, à former leurs collègues, à identifier des sources et à alerter sur les limites de certains indicateurs. Dans les rédactions locales ou de plus petite taille, où il n'existait pas toujours de poste dédié, le travail avec les données a davantage reposé sur des compétences transversales, parfois acquises dans l'urgence et sur le tas.

Ce déplacement vers le centre des rédactions et des productions médiatiques s'explique aussi par la demande du public. Les rédactions en chef ont rapidement constaté l'intérêt des lecteurs pour les graphiques, les cartes et les tableaux de bord. Les visualisations et pages de suivi sont devenues des contenus très consultés, parfois même des produits d'appel. Cette visibilité nouvelle a renforcé la reconnaissance des compétences « data », mais elle a aussi accru la pression sur les journalistes chargés de produire, vérifier et actualiser ces contenus, tout particulièrement lorsqu'ils étaient seuls à ce poste.

Des rédactions réorganisées autour des données

La pandémie a aussi transformé l'organisation du travail. Le journalisme de données, souvent décrit comme une pratique relativement solitaire, est devenu plus collectif. Plusieurs rédactions ont multiplié les articles en binôme ou en trinôme, associant datajournalistes et d'autres profils. Les collaborations avec les rubriques scientifiques ont été particulièrement importantes : elles permettaient de combiner la maîtrise des données avec la

compréhension des mécanismes épidémiques et l'accès à des experts capables d'interpréter les indicateurs.

Ces collaborations n'ont pas supprimé les frictions. Des tensions sont parfois apparues, notamment entre rubriques politiques et scientifiques, par exemple dans l'évaluation de la gravité de la situation ou de la portée des mesures sanitaires. Les premières tendaient à s'appuyer davantage sur les discours officiels et politiques, tandis que les secondes adoptaient une posture plus prudente, nourrie par les alertes des épidémiologistes. Les rédactions en chef ont souvent dû arbitrer ces divergences ou les faire cohabiter dans les espaces éditoriaux.

D'autres tensions relevaient des supports et des métiers. Le travail visuel des données n'était pas toujours réalisé par les datajournalistes, notamment dans les déclinaisons imprimées ou télévisuelles, souvent confiées à des graphistes. Cette division du travail pouvait produire des écarts entre logique graphique, lisibilité des données et contraintes propres au numérique. Elle révélait aussi la frontière encore marquée, dans certaines rédactions, entre cultures *print* et web. Plus largement, les choix de visualisation pouvaient être discutés au nom de leur lisibilité, mais aussi de leurs effets possibles ou de leur cadrage éditorial : un journaliste de données rapporte par exemple qu'une carte montrant la propagation rapide du virus en Chine, au début de la pandémie, a été refusée parce qu'elle « faisait peur ».

Colin Porlezza, « Les rédactions devraient s'engager activement dans le développement et la mise en œuvre d'outils collaboratifs et *open source* » (page 45)

Malgré leur centralité nouvelle, une grande partie du travail des journalistes de données est restée invisible pour le public. Collecter, vérifier, nettoyer, automatiser, surveiller des API, réparer des scripts ou mettre à jour des tableaux de bord sont des tâches cruciales, mais rarement valorisées comme production journalistique en tant que telle. Elles n'aboutissent pas toujours à des articles signés, ou

servent de base à des contenus publiés par d'autres. À certaines périodes, l'essentiel du travail de certains journalistes de données a consisté à maintenir des dispositifs de suivi, dans des flux quasi continus d'actualisation.

Cette visibilité accrue a aussi exposé les journalistes de données à la critique, notamment de la part de publics sceptiques ou hostiles aux mesures sanitaires. Plusieurs ont dû justifier leurs méthodes, expliquer leurs choix de sources ou répondre à des accusations de manipulation. En parallèle, ils ont participé à des échanges informels avec leurs confrères, notamment sur les réseaux sociaux, formant une sorte de « méta-rédaction » où se discutaient publiquement méthodes, sources et choix techniques.

Dashboards et outils du journalisme de données

Le journalisme de données fonctionne souvent sur des temporalités longues : identifier ou collecter des données, les vérifier, les nettoyer, les croiser, puis les interpréter et les mettre en récit. La pandémie a profondément bousculé ce rythme. Dès les premières semaines, les rédactions ont dû suivre des chiffres changeants, souvent incomplets, corrigés ou publiés avec retard. Le travail sur les données s'est ainsi trouvé pris dans une logique d'immédiateté habituellement associée aux formats d'actualité continue.

Les tableaux de bord sont devenus l'un des formats emblématiques de cette nouvelle temporalité. Le site corona-data.ch, créé par Daniel Probst, alors doctorant, a brièvement joué un rôle pionnier en Suisse, avant d'être rejoint par les *dashboards* de plusieurs médias puis par la plateforme fédérale covid19.admin.ch, mise en ligne en novembre 2020 et progressivement enrichie. Ces dispositifs permettaient de suivre l'évolution de la crise, de comparer les territoires et de donner au public un accès direct à des indicateurs actualisés.

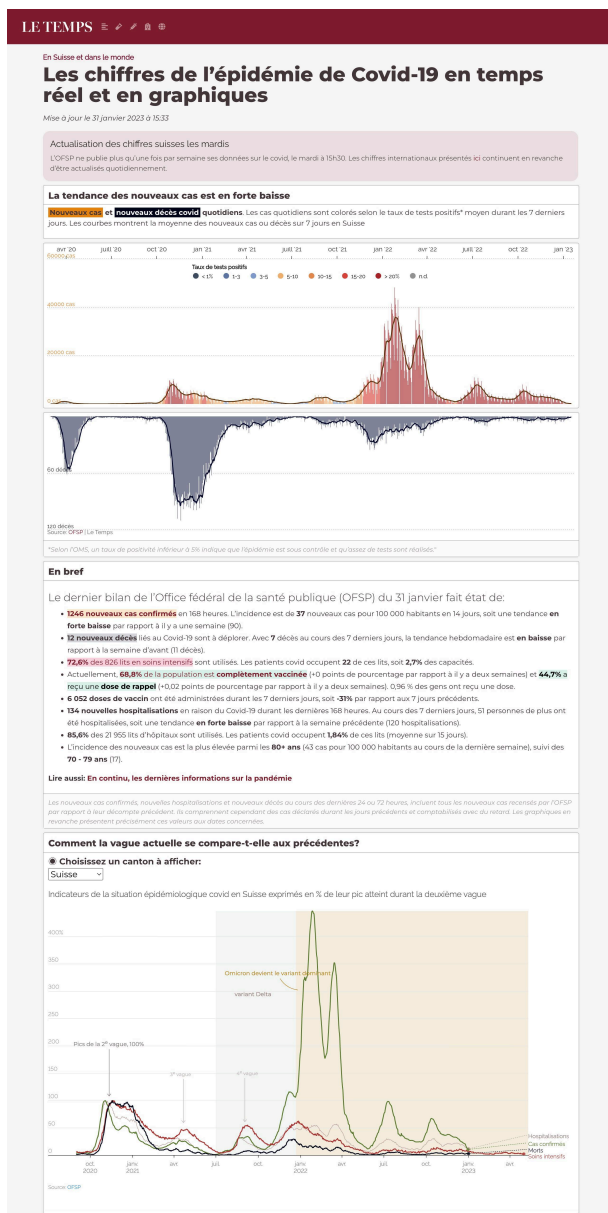


Figure 1. Extrait du *dashboard* Covid du Temps

Ils soulèvent néanmoins plusieurs questions éditoriales. Un suivi en temps quasi réel favorise-t-il la compréhension de la crise, ou entretient-il une forme d'attention anxieuse aux derniers chiffres ? Quelle place reste-t-il pour les indicateurs sociaux, économiques ou qualitatifs lorsque l'attention se concentre sur les courbes de cas et de décès ? Plusieurs rédactions indiquent que leurs tableaux de bord figuraient parmi les pages les plus consultées de leurs sites, ce qui a renforcé leur place dans l'offre éditoriale. Mais cette centralité ne risquait-elle pas d'orienter la perception de la crise vers ce qui était le

plus facilement mesurable, actualisable et visualisable ?

Gaël Hürlimann, « Les médias doivent s'affranchir du travail ingrat de récolte et de mise au propre des données » (page 46)

La question de l'éditorialisation des *dashboards* est à cet égard centrale. Les administrations publiques ont généralement privilégié une présentation des chiffres relativement peu commentée, répondant surtout au « combien ». Certains journalistes de données ont défendu une approche similaire, laissant les lecteurs explorer les données par eux-mêmes. D'autres ont insisté sur le rôle éditorial et d'accompagnement du journalisme : sélectionner les indicateurs, situer les valeurs dans des tendances, annoter les événements importants, expliquer les ruptures méthodologiques ou relier les chiffres aux décisions publiques. Cette tension entre documentation et interprétation traverse l'ensemble du journalisme de données.

La création de *dashboards* s'inscrit dans le passage progressif d'un travail artisanal, dominant au début de la pandémie (lorsque les données étaient dispersées, instables ou peu exploitables), à des formes plus automatisées de journalisme de données. L'automatisation, fondée sur des indicateurs récurrents et des publications régulièrement mises à jour, s'est appuyée sur des tableaux, des langages de programmation comme Python ou R, des API, ainsi que sur des plateformes de visualisation. Ces outils ont permis de produire rapidement des graphiques interactifs, intégrés aux articles ou aux *dashboards*, parfois alimentés par des données externes actualisées automatiquement.

Cette évolution a constitué un gain évident en efficacité, mais elle a aussi déplacé les enjeux de vérification. Lorsque la collecte, le traitement et la publication sont automatisés, les étapes où les erreurs, incohérences ou ruptures de série étaient auparavant repérées manuellement peuvent disparaître. L'automatisation ne supprime donc pas le besoin de contrôle ; elle exige au contraire de

nouvelles routines de vérification, de documentation et de surveillance des flux de données.

Les outils de visualisation (comme Datawrapper, Flourish ou Tableau) ont enfin joué un rôle structurant. Par leurs formats disponibles, leurs champs contraints, leurs options de titrage, de légende, d'annotation, de téléchargement des données ou d'interactivité, ils orientent ce qui peut être représenté facilement, et selon quelles formes. Ils peuvent encourager certaines bonnes pratiques, mais aussi imposer des contraintes, notamment lorsque l'actualisation automatique d'une visualisation de données se répercute sur ses usages antérieurs, au risque de décontextualiser d'anciens articles.

La période pandémique a ainsi été une période d'apprentissage accéléré. Des journalistes de données ont bricolé des solutions, testé de nouveaux outils, formé des collègues moins spécialisés et même suggéré des fonctionnalités aux outils de visualisation.

Antonin Descampe, « Le datajournaliste devient un opérateur, capable de combiner des instruments hétérogènes » (page 47)

Quid du rapport aux données à l'heure de l'IA ?

Depuis la fin de la pandémie, l'arrivée d'outils d'IA générative a rouvert la question de l'accès aux compétences techniques en journalisme de données. Ces outils peuvent faciliter certaines tâches : identifier des jeux de données, formuler une requête, trouver une fonction statistique, générer ou corriger du code, explorer de grands corpus documentaires, notamment des PDF administratifs. Ils pourraient donc abaisser certaines barrières d'entrée et permettre à des journalistes non spécialistes de traiter plus facilement des données.

Mais cette promesse doit être abordée avec prudence. Le risque principal est celui d'un usage insuffisamment conscientisé : erreurs de calcul, hallucinations, interprétation erronée ou orientée des variables, choix méthodologiques opaques ou transformations non documentées. L'IA peut assister le travail de collecte, de nettoyage ou d'exploration, mais elle ne peut pas remplacer la vérification, la compréhension statistique, la connaissance du contexte et la responsabilité éditoriale. Pour le journalisme de données, l'enjeu n'est donc pas de déléguer l'analyse à ces outils, mais de les intégrer dans des chaînes de production maîtrisées.

Chapitre 4. Couvrir la pandémie par les données : ce que les médias ont rendu visible

Ce chapitre se concentre sur les contenus eux-mêmes, en observant ce que les médias ont effectivement donné à voir : quels chiffres, quels indicateurs, quels formats graphiques et quels cadrages ont structuré la couverture de la pandémie en Suisse romande.

Une couverture médiatique fortement quantifiée

L'analyse menée dans le cadre du projet PANDA permet d'observer plus précisément la place des données dans la production médiatique suisse romande lors de cet événement. Elle repose sur l'ensemble des articles publiés en ligne par cinq médias entre janvier 2020 et décembre 2022 : un média audiovisuel de service public, deux quotidiens régionaux, un journal supra-régional et un média en ligne spécialisé dans la santé et la science. Sur les 177 324 articles collectés, 36 724 ont été identifiés comme portant sur le Covid-19 (voir l'encadré qui suit pour l'explication de la méthode).

Cette couverture a été marquée par une forte densité de chiffres, de graphiques et d'indicateurs. Les articles « Covid-19 » ne se sont pas contentés de raconter la crise : ils l'ont suivie, mesurée et rendue visible à travers des données sanitaires, économiques, sociales et politiques. Les cas, les décès, les hospitalisations, le taux de positivité, la couverture vaccinale, les résultats de sondages ou encore les indicateurs économiques ont structuré une partie importante du récit médiatique de la pandémie. Dans les articles Covid-19, nous comptons 9,23 chiffres tous les 3 500 signes, soit 18 % de plus que dans le corpus hors Covid. Les datavisualisations interactives liées à la pandémie sont, elles, au nombre de 813. Certaines ont été publiées plusieurs fois par le média (incluant des mises à jour), pour un total de 4 322 occurrences (« réutilisations » incluses).

	Total
Corpus	
Nombre d'articles total collectés	177 324
Articles du corpus « Covid-19 »	36 724
Datavisualisations	
Datavisualisations interactives uniques (hors comptabilisations des réutilisations)	813
Datavisualisations statiques (images png ou jpeg)	1 087
Nombre total de datavisualisations publiées (réutilisations incluses)	4 322
Nombre d'occurrences de la datavisualisation la plus utilisée	401
Pourcentage de datavisualisations interactives à usage unique	82 %
Présence des chiffres	
Taux de chiffres bruts dans les textes des articles pour 3 500 signes	8,25
- Corpus Covid-19	9,23
- Corpus hors Covid-19	7,86
Indicateurs statistiques (corpus Covid-19)	
Taux d'indicateurs pour 3 500 signes	2,37
- Catégorie « Sanitaire »	1,45
- Sous-catégorie « Décès »	0,52
- Sous-catégorie « Cas »	0,49
- Sous-catégorie « Hôpital/soins intensifs »	0,09
- Catégorie « Socioéconomique »	0,88
- Catégorie « Divers »	0,04

Tableau 2. Statistiques sur les corpus du projet PANDA

Cette quantification ne s'est toutefois pas répartie de manière homogène. Les indicateurs sanitaires représentent 61 % des indicateurs identifiés, et près de 50 % des datavisualisations interactives ont pour variable statistique principale un indicateur sanitaire, en particulier durant les premiers mois de la pandémie. Cette présence massive a contribué à installer une « météo Covid » (un suivi régulier,

parfois quotidien, de la situation épidémique à travers quelques indicateurs récurrents), rendue possible notamment par l'automatisation du traitement de ces données.

Méthodes d'analyse des corpus du projet PANDA

La partie du projet consacrée à la production médiatique repose sur des méthodes computationnelles, afin de traiter un corpus volumineux d'articles et de visualisations. À partir des données fournies par les cinq médias partenaires, nous avons constitué une base uniformisée de l'ensemble des articles publiés en ligne entre janvier 2020 et décembre 2022, puis isolé un sous-corpus Covid-19 à partir d'un critère large, à savoir au moins deux occurrences de termes liés à la pandémie, afin de limiter le risque d'exclure des articles pertinents.

Les datavisualisations interactives ont été repérées à partir des codes *iframe* dans les articles, puis vérifiées manuellement afin d'écarter les éléments non pertinents. Les métadonnées des visualisations ont été collectées automatiquement, puis complétées par un codage manuel de 26 variables, portant notamment sur les titres, sous-titres, propriétés interactives, indicateurs représentés, types d'encodage, annotations et sources. Ce corpus comprend 813 datavisualisations interactives uniques, dont certaines apparaissent dans plusieurs articles. Les datavisualisations statiques ont été identifiées à partir des images intégrées aux articles et codées de manière simplifiée par catégorie thématique.

Les indicateurs statistiques présents dans les textes ont été repérés à partir du contexte immédiat des chiffres présents dans les articles. Cette méthode permet d'observer des tendances dans l'usage des indicateurs, notamment selon leur date de publication, plutôt que de produire un décompte parfaitement exhaustif.

Chronologies, cartes, tableaux : les formes graphiques de la crise

Les datavisualisations interactives du corpus présentent une grande variété de formes graphiques, mais quelques formats dominent nettement. Les graphiques chronologiques, les barres/colonnes, les cartes et les tableaux jouent chacun un rôle distinct dans la couverture de la pandémie : suivre l'évolution temporelle, comparer des catégories, situer géographiquement la crise ou agréger plusieurs indicateurs dans un même espace. Leur fréquence et leur réutilisation renseignent aussi sur leur compatibilité avec une couverture actualisée et automatisée.

Les graphiques à barres et colonnes constituent la forme la plus fréquente parmi les visualisations uniques. Ils servent moins à suivre une évolution qu'à comparer des catégories : cantons, groupes de population, types de vaccins, secteurs économiques ou résultats de sondages. Leur polyvalence explique leur forte présence dans le corpus, mais leur moindre réutilisation suggère qu'ils se prêtent moins bien à une mise à jour continue. Ils semblent donc avoir surtout servi à produire des graphiques ponctuels, liés à un article ou à un angle précis.

Les graphiques chronologiques occupent également une place centrale, ce qui confirme le poids de la temporalité dans la couverture médiatique de la pandémie. La crise a été racontée comme une succession de vagues, de seuils, de mesures et d'assouplissements. Les questions auxquelles ces graphiques répondaient étaient d'abord temporelles : la situation s'améliore-t-elle ? La courbe repart-elle à la hausse ? Les mesures produisent-elles un effet ? À quel moment peut-on espérer une sortie de crise ? Leur part reste élevée lorsqu'on tient compte des réutilisations, signe qu'ils se prêtaient bien à un suivi actualisé de la situation sanitaire.

Les cartes présentent un profil inverse à celui des barres et colonnes : elles sont moins nombreuses parmi les visualisations uniques, mais certaines sont fortement réutilisées. Leur efficacité apparente tient

à leur capacité à ancrer la pandémie dans l'espace, en comparant des territoires, notamment les cantons suisses ou les pays. Cette puissance visuelle peut toutefois masquer des biais de perception. Une carte choroplèthe (où chaque territoire est coloré selon la valeur d'un indicateur) donne beaucoup de poids visuel aux territoires les plus vastes. En Suisse, un canton comme les Grisons, très étendu mais relativement peu peuplé, peut ainsi donner l'impression d'une situation globale plus préoccupante qu'un canton urbain plus petit comme Bâle. Les cartes anamorphes (qui déforment la taille des territoires en fonction d'une variable) ou les tilemaps (des cartes schématiques où chaque territoire est représenté par une tuile de taille identique) réduisent une partie de ce problème, mais introduisent d'autres contraintes de lecture. Elles rappellent qu'aucune forme graphique n'est neutre : chaque choix facilite certaines comparaisons et en rend d'autres plus difficiles.

Les tableaux occupent une place particulière. Leur inclusion dans un corpus de datavisualisations peut

sembler discutable, car un tableau simple ne repose pas toujours sur un encodage graphique élaboré. Elle se justifie toutefois dans le contexte de la pandémie : les outils de visualisation proposent des tableaux enrichis, combinant chiffres, couleurs, petits graphiques, filtres ou tris, et ces formats ont joué un rôle important dans le suivi médiatique de la crise. La part limitée des tableaux parmi les visualisations uniques contraste avec leur poids très élevé dans le nombre de mentions totales (% MT), ce qui indique une forte réutilisation. Dans le corpus, la visualisation la plus réutilisée est un tableau simple du *Temps* consacré aux cas annoncés quotidiennement, publié 401 fois. Les tableaux enrichis, plus présents à partir de l'automne 2020, témoignent d'une évolution rapide des pratiques. Ils permettent de suivre plusieurs indicateurs simultanément, se prêtent bien aux mises à jour automatiques via API et évitent certains biais de perception associés aux cartes. Leur essor montre aussi le rôle structurant des plateformes comme Datawrapper, dont les fonctionnalités définissent en partie ce qui devient facile ou difficile à représenter.

Type de visualisation	% DVU % MT*	Usages principaux pendant la pandémie	Intérêt principal	Limites ou risques
Graphique à barres et à colonnes	40,7 % 15,1 %	Usages très variés	Compare un petit nombre d'entités ou de populations à un instant T Facile à lire, intuitif	Ne convient pas à un nombre élevé d'entités de population (p. ex. cantons)
Graphique chronologique (timeline)	32,5 % 28,6 %	Documenter l'évolution de la situation sanitaire, projeter sa possible évolution future	Ancrage de la pandémie dans le temps Rend visible l'effet des mesures	Peut favoriser un focus sur le court terme Surinterprétation Effets anxiogènes
Tableau	10,3 % 31,8 %	Afficher un nombre élevé de variables pour une population donnée	Riche et dense en informations Permet l'exploration interactive	Plus complexes à lire Message dispersé
Carte	8,1 % 20,7 %	Comparer des situations sanitaires, des opinions ou votations	Ancrage de la pandémie dans le territoire	Biais de perception Stigmatisation de régions ou cantons
Camembert	5,5 % 2,2 %	Comparer des sous-entités au sein d'un tout	Très intuitif et facile à lire	Importants biais lorsque nombre élevé de parts

* La part des visualisations uniques (% DVU) ne compte le type de visualisation qu'une seule fois par visualisation ; la part des mentions totales (% MT) tient compte des réutilisations d'une même visualisation dans plusieurs articles.

Tableau 3. Répartition des principaux types de datavisualisation, hormis la catégorie « autres » comptant pour 2,9 % des DVU et 1,7 % des MT (données du projet PANDA)

Les cadrages de la pandémie via les indicateurs

Au-delà des visualisations de données, nous nous sommes intéressés à l'usage des indicateurs dans les textes des articles : comment certains chiffres et catégories d'indicateurs ont-ils contribué à cadrer le discours médiatique au fil de la pandémie ?

Étudier les indicateurs et leurs cadrages

Identifier des indicateurs statistiques et leur cadrage dans un texte à l'aide de méthodes computationnelles pose un défi méthodologique : tous les chiffres ne sont pas des indicateurs, tous ne concernent pas la pandémie, et leur restitution doit rester suffisamment contextualisée.

Notre méthode a d'abord consisté à repérer l'ensemble des chiffres présents dans les articles, y compris ceux qui ne constituent pas des indicateurs statistiques au sens strict, comme les dates ou les résultats sportifs. Ce premier repérage permet une comparaison générale de la présence des chiffres entre les articles « Covid-19 » et les autres.

Nous avons ensuite identifié et classé les indicateurs à partir d'une approche déductive, fondée sur les indicateurs préalablement repérés dans les visualisations de données. Pour chaque chiffre, nous avons analysé son contexte immédiat, avant et après son apparition dans le texte, afin de repérer la présence éventuelle de mots-clés associés à ces indicateurs. Cette méthode comporte des angles morts, mais elle permet d'étudier efficacement les indicateurs les plus récurrents.

Enfin, les résultats ont été normalisés afin de tenir compte des différences de longueur des articles. Nous avons calculé une densité de chiffres ou d'indicateurs pour 3 500 signes, soit l'équivalent approximatif d'un article d'actualité court. Les taux peuvent ainsi être lus comme un nombre moyen d'occurrences rapporté à un format journalistique standard.

Parmi les indicateurs identifiés dans les textes, les cas et les décès dominent largement, sur l'ensemble de la période et quel que soit le média. La vaccination rejoint ce premier ensemble à partir de la fin 2020, devenant à son tour un indicateur fortement discuté. On observe toutefois une baisse progressive de la présence des indicateurs sanitaires dans les articles, particulièrement nette après la levée quasi totale des restrictions au printemps 2022.

La présence des cas et des décès augmente lors des deux premières vagues, mais leur évolution ne suffit pas, à elle seule, à identifier clairement les différentes vagues épidémiques. Cela invite à nuancer leur rôle dans les articles. Dans de nombreuses productions journalistiques, ces chiffres semblent fonctionner comme un contexte général, souvent mentionné de manière routinière, sans nécessairement faire l'objet d'une analyse approfondie. Il est plausible qu'une partie de ce suivi se soit progressivement déplacée du texte vers les datavisualisations, notamment les tableaux et chronologies automatisés, à mesure que les graphiques réutilisables se sont imposés dans les articles, en particulier après la mise à disposition de l'API de l'OFSP à l'automne 2020.

Les indicateurs liés aux hospitalisations et à l'occupation des soins intensifs sont moins présents en volume, mais leurs hausses sont plus nettement associées aux vagues, en particulier lors de la deuxième. Ils ont contribué à recadrer la pandémie autour de la capacité de résistance du système de santé : le risque de décès dépendait aussi de la capacité des hôpitaux à absorber l'afflux de patients. Leur force tient aussi à leur dimension prospective : ils ne décrivent pas seulement la situation présente, mais signalent un risque à venir, celui d'une saturation hospitalière susceptible d'entraîner une hausse des décès.

Le taux de reproduction effectif (ou R_e) apparaît beaucoup plus faiblement, mais il joue un rôle ponctuellement important entre la fin 2020 et le début 2021. Difficile à comprendre et à expliquer, comme l'ont souligné plusieurs journalistes en entretien, il semble avoir surtout gagné en visibilité

lorsqu'il est devenu un indicateur directement lié aux décisions politiques. Dès lors que le gouvernement suisse a associé, par voie d'ordonnance, le renforcement ou l'allègement de certaines restrictions à des seuils du taux de reproduction, cet indicateur technique est devenu un repère médiatique et administratif.

Les indicateurs économiques et sociaux sont plus diversifiés et proviennent de sources plus variées que les indicateurs sanitaires. Leur évolution ne permet toutefois pas de confirmer clairement l'hypothèse de rebonds médiatiques (sur le thème de l'impact économique et social de la pandémie) en dehors des pics épidémiques ou après de longues périodes de restrictions. En fin de période, leur présence se

maintient alors que les indicateurs sanitaires reculent, mais une lecture qualitative suggère plutôt un traitement rétrospectif de thèmes sociaux et économiques dans lesquels la pandémie constitue un contexte explicatif important.

Ces résultats posent une question plus générale : à quoi aurait ressemblé la couverture médiatique si des indicateurs sociaux, psychologiques et économiques de qualité avaient été mis à disposition par les autorités dans des formats accessibles, régulièrement mis à jour et facilement réutilisables ? Cette question dépasse l'analyse des contenus publiés et renvoie à l'organisation même de l'écosystème des données en temps de crise sur laquelle nous reviendrons dans le chapitre 6 consacré aux perspectives.

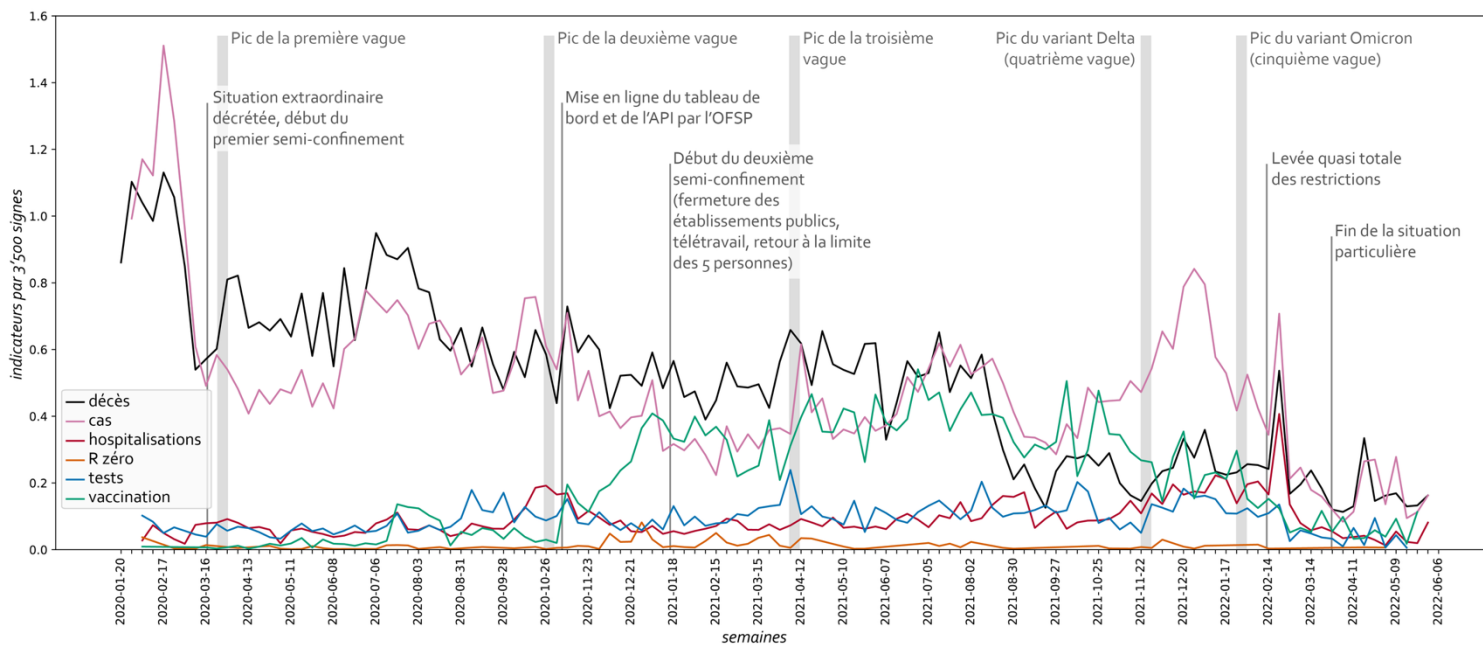


Figure 2. Taux de présence hebdomadaire des sous-catégories d'indicateurs sanitaires dans les articles du corpus Covid-19 (par 3 500 signes)

Chapitre 5. La compréhension des données par les publics en contexte d'incertitude

Après avoir analysé la production et la mise en forme médiatique des données, ce chapitre se concentre sur leur réception : comment les données sont-elles comprises par des publics hétérogènes ? Quels facteurs influencent leur interprétation, leur crédibilité perçue et leur appropriation ?

S'informer en pandémie : entre besoin de repères et fatigue informationnelle

La pandémie a constitué une situation inédite non seulement sur le plan sanitaire, mais aussi dans les pratiques d'information. Dès les premières semaines, les publics ont dû composer avec une forte incertitude : évaluer les risques du quotidien, comprendre les mesures en vigueur, anticiper les restrictions et ajuster leurs comportements (prendre les transports, voir des proches, travailler, etc.). Dans ce contexte, l'information devenait un outil d'orientation pratique.

Les premiers mois de la crise ont été marqués par un intérêt marqué pour les sources institutionnelles et les médias traditionnels. Plusieurs travaux ont décrit ce moment comme relevant d'un effet de ralliement autour des autorités : en situation de crise, les clivages ordinaires tendent provisoirement à s'atténuer et la confiance envers l'exécutif, les experts et les relais médiatiques augmente. En Suisse, les audiences radio et télévision ont progressé au début de la pandémie, alors même qu'elles étaient structurellement orientées à la baisse. Plus largement, le rapport du Reuters Institute de 2021 a observé un effet Covid sur la consommation d'information, notamment pour la télévision et les abonnements payants, ainsi qu'une hausse de la confiance déclarée envers l'actualité.

Ce regain de confiance n'a toutefois été ni uniforme ni durable. Il a surtout bénéficié aux acteurs perçus comme les plus légitimes dans le traitement de la

crise, notamment les médias de service public. Il a aussi été rapidement mis à l'épreuve par la densification de l'écosystème informationnel numérique. La pandémie a été accompagnée d'une infodémie, au sens d'une surabondance d'informations de qualité très variable : données officielles, articles journalistiques, prises de parole d'experts, conseils pratiques, mais aussi interprétations alternatives, contenus militants et désinformation. Pour une partie des publics, le problème n'était donc pas seulement de trouver de l'information, mais de trier, hiérarchiser et interpréter des informations trop nombreuses, parfois contradictoires ou difficiles à évaluer.

Cette saturation a nourri une fatigue informationnelle. Suivre les chiffres quotidiens, les annonces politiques, les débats scientifiques et les controverses sanitaires a pu nourrir un sentiment d'anxiété, puis de lassitude ou d'évitement (la « fatigue informationnelle »). Certains publics prennent leur distance avec le suivi de l'actualité au risque de devenir plus exposés vis-à-vis d'informations erronées voire manipulatoires. La communication des données devait donc répondre à une tension : fournir suffisamment d'informations pour permettre de comprendre et d'agir, sans ajouter à la confusion ni à la surcharge.

Dans ce contexte, les données et les visualisations ont joué un rôle particulier. Elles ont permis de transformer une menace diffuse et difficile à saisir en phénomènes mesurables : courbes de cas, hospitalisations, décès, taux de reproduction, vaccinations, comparaisons cantonales ou internationales. Plusieurs recherches sur la réception des visualisations de données pendant la pandémie montrent que les chiffres ont pu aider les citoyens à gérer l'incertitude, et donner un sentiment de maîtrise. Les tableaux de bord et graphiques ont ainsi fonctionné comme des repères pratiques : un moyen

de suivre l'évolution de la situation et d'ajuster ses comportements. L'enjeu, pour les institutions et les médias, est donc double : aider des publics hétérogènes à s'orienter dans l'incertitude sans les submerger, et éviter de réduire la pandémie à quelques indicateurs récurrents au détriment de dimensions moins facilement quantifiables.

Littératie des données : tous les publics ne lisent pas les chiffres de la même manière

La pandémie a rappelé que l'accès aux données ne garantit pas leur compréhension. Lire une courbe, comparer des taux, interpréter une moyenne, distinguer une corrélation d'une causalité ou comprendre un intervalle d'incertitude suppose des compétences qui ne sont pas également distribuées dans la population. La littératie des données ne se réduit pas à la capacité de lire des chiffres : elle désigne un ensemble de compétences permettant de comprendre comment les données sont produites, ce qu'elles mesurent, ce qu'elles ne mesurent pas, et dans quelles limites elles peuvent être interprétées.

Cette question est centrale en contexte de crise. Les publics ne recevaient pas les indicateurs pandémiques comme des observateurs neutres. Ils les interprétaient à partir de leurs connaissances préalables, de leur confiance envers les institutions et les médias, de leur expérience personnelle de la crise, de leurs inquiétudes, mais aussi de leurs convictions politiques ou sanitaires. Une même visualisation pouvait donc être lue comme une information utile, comme une preuve de sérieux, comme un signal anxigène ou comme un objet suspect. Les différences de réception ne tiennent pas seulement à la clarté du graphique ; elles dépendent aussi du rapport préalable des publics aux sources, aux chiffres et à la crise elle-même.

Les compétences statistiques jouent un rôle important dans cette réception. Un public familier des données comprendra plus facilement qu'un indicateur puisse être révisé, qu'une série soit affectée par un changement de méthode, qu'un taux

doive être rapporté à une population de référence, ou qu'une marge d'erreur ne constitue pas une faiblesse mais une information sur la précision de la mesure. À l'inverse, pour des publics moins familiers de ces conventions, les révisions, incertitudes ou divergences entre sources peuvent être interprétées comme des contradictions, voire comme des signes d'incompétence ou de manipulation.

Pour les médias et les institutions, cette hétérogénéité impose de ne pas concevoir les visualisations pour un lecteur abstrait, supposé attentif, compétent et disponible. Elle invite au contraire à proposer plusieurs niveaux de lecture : un message principal clair, des repères visuels stables, des définitions accessibles, puis des compléments méthodologiques pour les lecteurs qui souhaitent aller plus loin. La pédagogie ne consiste pas à simplifier jusqu'à effacer la complexité, mais à organiser cette complexité pour qu'elle puisse être (progressivement) comprise.

La littératie des données ne relève donc pas seulement d'une responsabilité individuelle des publics. Elle dépend aussi des infrastructures d'information mises à leur disposition : définitions accessibles, notes méthodologiques, sources partagées et explications des limites. Une communication responsable des données doit ainsi viser un double objectif : renforcer les compétences collectives face aux chiffres, tout en concevant des dispositifs qui n'exigent pas des publics un niveau d'expertise trop élevé pour comprendre l'essentiel.

Monique Lehky Hagen et Diego Kuonen, « La littératie des données exige un changement de culture sociétal fondamental, profond et continu » (page 48)

Représenter l'incertitude sans affaiblir la confiance

En situation de crise sanitaire, les données évoluent rapidement : révisions, changements de définition, retards de *reporting* et corrections successives font de l'incertitude une composante normale de

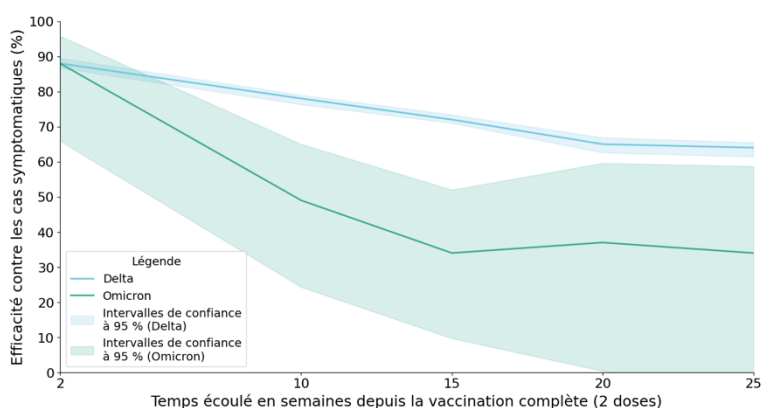
l'information plutôt qu'une exception. Communiquer cette incertitude crée toutefois une tension : afficher l'incertitude peut signaler transparence et rigueur, mais rendre le message plus complexe à traiter, ce qui est particulièrement délicat en contexte de crise où le public doit prendre des décisions coûteuses (vaccination, comportements de prévention) à partir d'informations nécessairement imparfaites.

L'un des objectifs du projet PANDA était précisément d'examiner les effets de cette incertitude lorsqu'elle est intégrée aux visualisations de données. Faut-il, par exemple, afficher des intervalles de confiance ou bandes d'incertitude ? Ces éléments améliorent-ils la compréhension et la confiance, ou rendent-ils le message plus difficile à saisir ? Deux mécanismes ont été testés : un mécanisme de coût, selon lequel l'incertitude réduit la fluidité de traitement et diminue la mise à jour mesurée des croyances en lien avec le message transmis (par exemple : indiquer qu'un vaccin est efficace dans une fourchette de 91 % à 97 % réduit-il la croyance en son efficacité par rapport à la simple mention d'une efficacité de 94 % ?) ; et un mécanisme de bénéfice, selon lequel elle peut au contraire renforcer la confiance, en particulier chez les personnes disposant d'une plus grande littératie statistique.

Les résultats convergent vers une conclusion nuancée : l'affichage de l'incertitude produit des effets mixtes. Le résultat le plus robuste concerne la difficulté de traitement : l'affichage de l'incertitude réduit la fluidité de traitement. Cette baisse de fluidité peut ensuite freiner la mise à jour des croyances. Certaines expériences montrent que les publics les plus à l'aise avec les chiffres peuvent percevoir l'incertitude comme un signe de rigueur méthodologique, renforçant la confiance et facilitant l'intégration du message. L'expérience menée en Suisse confirme cette logique de « coût de traitement », tout en montrant que ce coût passe surtout par la complexité perçue de la visualisation. Le problème n'est donc pas seulement « faut-il montrer l'incertitude ? », mais surtout « comment la représenter sans surcharger le lecteur ? »

Ces résultats suggèrent qu'il n'existe pas de réponse universelle. L'effet dépend d'un arbitrage entre transparence et coût cognitif, et d'une anticipation de l'hétérogénéité des publics. L'enjeu n'est pas de « montrer plus d'information », mais de rendre l'information plus traitable sans renoncer à la transparence.

Efficacité du Pfizer contre les Covid-19 symptomatiques au fil des semaines



Données britanniques.
Source des données : UK Health Security Agency (UKHSA).

Figure 3. Datavisualisation avec intervalles de confiance, utilisée dans le cadre de l'expérience sur les effets de la visualisation de l'incertitude

Trois principes opérationnels se dégagent.

Réduire la complexité perçue. La priorité est de préserver la lisibilité de la visualisation. Cela peut suggérer de privilégier des encodages d'incertitude faciles à interpréter (par exemple une bande de confiance discrète) plutôt que des dispositifs visuellement chargés, de limiter les objets graphiques concurrents (séries, couches, annotations simultanées), d'introduire l'information par étapes, et d'intégrer une micro-explication au point de lecture (titre, sous-titre ou légende). Des conventions graphiques stables évitent au public de réapprendre le code visuel à chaque visualisation.

Présenter l'incertitude de manière sélective.

L'affichage doit être ajusté au public et à l'objectif de communication. Pour un large public, si l'incertitude n'est pas indispensable à la compréhension immédiate, il peut être préférable de ne pas la représenter directement dans le graphique principal et de la déplacer vers une note ou un complément. Pour des publics plus avertis, comme le suggèrent certaines expériences, son affichage peut au contraire renforcer la confiance dans la visualisation, à condition d'être accompagné d'un guidage textuel.

Adopter une logique de divulgation progressive.

Plutôt que de tout afficher d'emblée, il s'agit de proposer un niveau d'information principal simple et lisible, puis de donner accès à des détails supplémentaires via une note, un lien, un encadré ou un dispositif interactif. Cette approche concilie rapidité de lecture et exigences de transparence, et permet de sortir de l'alternative entre une certitude simplificatrice (facile à lire mais potentiellement trompeuse) et une incertitude illisible (transparente mais difficile à exploiter). Concrètement : d'abord un message principal simple (tendance, ordre de grandeur), ensuite l'incertitude accompagnée d'une courte explication, puis un accès optionnel aux détails (méthode, sources, limites, révisions). Cette « transparence graduée » pourrait aider à préserver la confiance tout en limitant les coûts de lisibilité, ce qui est crucial dans la communication de crise.

Favoriser la transparence et l'accompagnement de la réception des données

La transparence constitue l'un des idéaux du journalisme de données. Les journalistes de données interrogés dans le cadre du projet PANDA y reviennent presque unanimement lorsqu'il s'agit d'indiquer d'où viennent les données. Sur ce point, les pratiques observées dans les productions analysées vont globalement dans le bon sens : la mention des sources est presque toujours présente ; seules 31 visualisations sur 813 ne contiennent aucune identification de la source, ni dans le graphique lui-même ni dans le texte qui l'accompagne.

Cette présence quasi systématique des sources ne signifie toutefois pas que la transparence soit complète. Dans beaucoup de cas, la source est mentionnée, mais le lien vers le jeu de données original n'est pas fourni, même si la majorité des visualisations donnent accès aux données brutes sous forme de tableau, notamment grâce aux fonctionnalités proposées par défaut par Datawrapper, outil très utilisé durant la pandémie. C'est un progrès important par rapport à des travaux antérieurs montrant que la mise à disposition des jeux de données restait rare dans les productions de datajournalisme.

Toutefois, donner accès aux données ne suffit souvent pas : encore faut-il aider le lecteur à comprendre ce qu'elles contiennent, comment elles ont été produites, quels choix de traitement ont été faits et quelles limites elles comportent. Sur ce point, les pratiques restent plus inégales. Plusieurs journalistes interrogés défendent l'ajout de notes méthodologiques, d'encadrés explicatifs, voire la mise à disposition de feuilles de calcul ou de scripts permettant de reproduire les traitements. Dans les productions analysées, ces éléments demeurent pourtant peu visibles. Près de 70 % des visualisations de données ne contiennent aucun autre élément que le titre, un éventuel sous-titre et la mention de la source. Environ 30 % proposent des informations complémentaires, souvent dans un champ « note »,

qui apportent un contexte, une précision méthodologique ou une limite d'interprétation.

Cette tension est bien documentée par la littérature scientifique. Trop d'explications techniques peuvent alourdir la lecture, détourner l'attention du message principal ou décourager des publics peu familiers des données. Mais l'absence d'accompagnement peut aussi nourrir une forme de mystification des chiffres : le graphique apparaît alors comme une preuve autosuffisante, sans que le lecteur puisse saisir les conditions de production, les choix méthodologiques ou les incertitudes qui l'entourent.

Le texte qui accompagne les visualisations joue ici un rôle décisif. Il peut rappeler ce que mesure exactement un indicateur, signaler un changement de définition, expliciter une rupture de série, préciser une source ou distinguer une observation d'une projection. Le vocabulaire employé est également important. Des formulations comme « selon les données disponibles », « à ce stade », « il semble que », « cette estimation suggère » ou « les résultats indiquent » permettent de signaler le degré de certitude associé à une information. Cette modalité épistémique est particulièrement importante en contexte de crise, où les connaissances évoluent et où les données sont souvent provisoires.

Dans cette perspective, certains chercheurs proposent de compléter le fact-checking par une forme d'*uncertainty-checking* : il ne s'agit pas seulement de vérifier si une affirmation est vraie ou fausse, mais d'évaluer son degré de certitude au regard des preuves disponibles. Pour les journalistes comme pour les institutions, cette approche invite à mieux expliciter ce que l'on sait, ce que l'on ignore encore, ce qui est probable et ce qui reste incertain.

Charles-Henry Groult, « Notre responsabilité est de faire preuve d'une extrême prudence, voire d'assumer une part d'incertitude » (page 49)

Décrire ou interpréter : ce que montrent les titres des datavisualisations

L'analyse des titres des datavisualisations interactives du corpus PANDA montre que les médias ne cadrent pas tous les graphiques de la même manière. Certains titres se contentent de décrire l'indicateur représenté, par exemple « Taux de vaccination au sein de la population ». D'autres proposent une lecture plus interprétative, comme « Les cantons les moins vaccinés dénombrent plus de cas » : ils restent factuels, mais sélectionnent un élément saillant et orientent davantage l'interprétation.

Les pratiques varient fortement selon les médias, la plupart optant néanmoins pour des titres descriptifs. Ces écarts peuvent tenir aux lignes éditoriales, aux formats d'articles, aux règles propres au service public, mais aussi aux cultures professionnelles et aux trajectoires individuelles des journalistes de données. Ils rappellent surtout que le titre d'un graphique n'est pas un simple élément d'habillage : il participe à la médiation des données. Un titre descriptif laisse davantage le lecteur interpréter ; un titre interprétatif facilite la compréhension d'un message, mais engage plus fortement la responsabilité éditoriale du média.

Concevoir des visualisations adaptées

Une visualisation n'est jamais un simple contenant neutre. Sa structure, son degré de lisibilité, la manière dont elle hiérarchise l'information, la place accordée au texte ou encore les choix d'échelle, de couleur et d'annotation contribuent à orienter la perception du public. Pour les médias, les administrations et les acteurs impliqués dans la communication de données, les choix graphiques ne relèvent donc pas seulement de l'esthétique : ils participent directement à la compréhension, à la crédibilité perçue et à la disposition des publics à poursuivre la lecture, à faire confiance au message ou à le relayer.

Nos travaux, menés à partir de visualisations réelles publiées dans la presse suisse romande, montrent que les publics ne réagissent pas seulement aux données elles-mêmes, mais aussi à leur mise en

forme. Le résultat le plus robuste concerne la complexité visuelle : lorsqu'un graphique paraît trop chargé, il réduit la clarté et l'utilisabilité perçues, affaiblit l'attrait esthétique et peut même augmenter l'impression que la visualisation cherche à manipuler. La réception est moins favorable : le public manifeste moins d'intérêt, évalue le graphique plus négativement, se montre moins enclin à le partager et moins disposé à lire l'article qui l'accompagne. La densité d'information produit des effets similaires, mais plus ciblés : elle réduit surtout la clarté, l'informativité perçue, l'utilisabilité et l'intérêt, tout en diminuant également la propension à lire l'article associé.

Ces résultats plaident pour une règle simple : la lisibilité immédiate doit rester l'objectif par défaut. Plusieurs études convergent en ce sens : les usages de l'information sont souvent rapides, fragmentés et réalisés sur mobile, ce qui favorise des visualisations capables de donner une vue d'ensemble et de dégager une tendance plutôt que d'exposer trop de détails d'emblée. Les publics semblent apprécier ces formes comme des « raccourcis informationnels », tandis que le texte peut prendre le relais pour préciser les indicateurs, les méthodes ou les limites. Une visualisation trop chargée risque au contraire de détourner le regard plutôt que de favoriser l'exploration autonome.

Valentin Tombez, « Une lecture brute des chiffres nourrit des interprétations trompeuses » (page 50)

Nos résultats montrent que trois dimensions améliorent certains aspects de la réception, en particulier la compréhension et la qualité perçue. La première est l'accessibilité informationnelle : un titre explicite, une structure lisible, des repères visuels clairs et des libellés compréhensibles améliorent la clarté et l'informativité perçues, ainsi que les jugements de compétence, de bienveillance de l'émetteur, et d'attrait esthétique. La deuxième est l'hétérogénéité visuelle maîtrisée : différencier suffisamment les éléments pour guider le regard, sans multiplier les couleurs, les formes ou les niveaux

d'information au point de créer une surcharge. La troisième est l'articulation précise entre texte et visuel : le titre, le sous-titre, les notes et les annotations doivent aider à comprendre ce qu'il faut regarder et avec quel degré de prudence. Ces approches améliorent la compréhension, même si elles n'augmentent pas automatiquement l'appréciation du graphique, la plausibilité perçue du message ou l'intérêt porté au contenu.

L'interactivité observée dans les datavisualisations PANDA

Dans notre corpus de 813 datavisualisations interactives, l'interactivité apparaît surtout comme une aide ponctuelle à la lecture plutôt que comme un véritable dispositif d'exploration. 19 % des visualisations ne proposent aucune interactivité. Parmi celles qui en intègrent, la fonction dominante est la saillance : dans 66 % des cas, l'interaction sert principalement à mettre en évidence une valeur, une série, une variable ou une unité territoriale, par exemple au survol ou au clic.

Les formes d'interactivité plus exploratoires restent minoritaires. Elles concernent 12 % des visualisations et prennent surtout deux formes : alterner entre plusieurs graphiques ou indicateurs dans un même espace, ou trier les lignes d'un tableau selon différentes variables. Enfin, dans 3 % des cas, l'interactivité a été jugée sans plus-value claire pour la compréhension.

Ces résultats invitent à relativiser l'idée selon laquelle le numérique rendrait automatiquement les visualisations plus riches ou plus intelligibles. L'interactivité est utile lorsqu'elle clarifie, compare, filtre ou approfondit ; elle l'est beaucoup moins lorsqu'elle ajoute une couche de manipulation sans bénéfice réel pour le lecteur.

L'interactivité doit être pensée dans cette même logique. Elle n'est utile que si elle améliore réellement l'usage : afficher un détail à la demande, filtrer une série, comparer des catégories ou faire apparaître des définitions sans encombrer le graphique principal.

Elle devient problématique lorsqu'elle masque l'information essentielle, oblige à multiplier les manipulations ou complique l'accès au message, en particulier sur mobile. Un bon graphique interactif doit donc rester compréhensible avant même que l'on commence à cliquer.

Le texte qui accompagne la visualisation fait partie intégrante du dispositif. Titre, sous-titre, notes, annotations, légende, source et brève explication méthodologique jouent un rôle décisif dans la manière dont le graphique est lu. Ces éléments agissent comme un contrat de lecture : ils indiquent ce qu'il faut regarder, comment interpréter le graphique et avec quel niveau de prudence. Une phrase simple peut suffire à lever une ambiguïté, à rappeler ce que mesure un indicateur ou à expliciter une convention visuelle, travail de guidage particulièrement important dans des contextes marqués par l'incertitude.

Laurence Senn, « Il est essentiel de transmettre des chiffres avec l'analyse nécessaire à leur interprétation » (page 51)

Ainsi, pour des visualisations destinées à un large public, nos résultats invitent à cinq priorités de design :

- Éviter la complexité « par défaut »
- Rendre le message principal immédiatement lisible
- Guider le regard sans surcharger
- Expliciter les conventions de lecture (unités, échelles, périodes de référence, etc.)
- Assurer la cohérence entre texte et visuel

Ce qui distingue une visualisation utile n'est pas sa richesse ou sa sophistication, mais sa capacité à rendre le message principal immédiatement lisible : elle donne accès à des détails lorsque c'est utile, signale ses sources, évite les effets gratuits et respecte la capacité d'attention du public. Pour les médias comme pour les institutions publiques, cette exigence conditionne la possibilité même d'une information compréhensible, crédible et partageable.

Bien concevoir une visualisation, ce n'est donc pas seulement « bien montrer des données » : c'est trouver l'équilibre entre un contenu complexe, un dispositif graphique et des publics qui doivent pouvoir comprendre sans être découragés ou submergés.

Vers des expériences de lecture adaptatives

Les sections précédentes montrent qu'un même graphique ne sera pas compris de la même manière par tous les publics. Les lecteurs diffèrent par leur intérêt pour le sujet, leur familiarité avec les données, leur confiance dans les sources, leur disponibilité attentionnelle et leurs attentes face à l'information. Une piste consiste donc à concevoir des expériences de lecture plus adaptatives, capables de guider les publics sans imposer à tous le même niveau de détail ni la même progression.

Dans le cadre du projet PANDA, nous avons exploré cette piste à travers deux dispositifs interactifs.

Le premier (Vizualink) visait à favoriser des liens dynamiques entre le texte d'un article et les représentations graphiques. Le principe est simple : en survolant un segment de texte mis en évidence, le lecteur voit simultanément s'accroître la partie correspondante dans le graphique associé.

Cette approche vise à guider le regard et à faciliter la compréhension sans effort supplémentaire. Testée auprès de 360 participants, cette intégration texte-graphique améliore significativement la qualité perçue de l'expérience de lecture : les utilisateurs perçoivent la lecture comme plus fluide, ce qui favorise une réflexion plus approfondie.

Since the last survey in March, the vaccination campaign has made significant progress. In total, 60% of respondents have now received at least one dose. As a result, only 2% of Swiss people are now ready to be vaccinated as soon as possible, while 12% prefer to wait and 25,5% are convinced they do not want to receive an injection.

"Would you like to get vaccinated against Covid-19?"

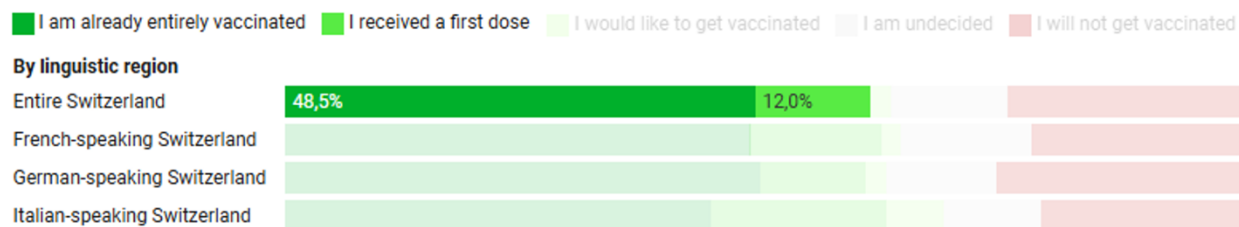


Figure 4. Capture d'écran du dispositif interactif Vizualink élaboré dans le cadre du projet PANDA

L'outil profite aussi bien aux experts du sujet, qui y voient un gain de fluidité, qu'aux novices, pour qui il explicite les liens logiques entre données et récit. Des limites ont néanmoins été identifiées : sur petits écrans, il devient difficile de maintenir une vue d'ensemble du graphique tout en interagissant avec le texte, et l'utilité du dispositif diminue face à des graphiques jugés trop simples.

Une seconde piste explorée concerne l'usage de l'IA générative comme assistant de lecture. Des techniques comme le *scrollytelling* (où le visuel évolue avec le défilement de la page) apportent une dimension immersive, mais reposent souvent sur des structures semi-statiques qui ne répondent pas aux interrogations précises des lecteurs. Pour pallier ce manque d'adaptabilité, nous avons développé NewsChat, un assistant de lecture basé sur GPT-40, conçu pour être « conscient du contenu » : il peut fournir des résumés sur mesure, expliquer des termes complexes ou clarifier des points de données précis selon les besoins de l'utilisateur.

En guidant NewsChat par le biais de *prompt-tuning* (l'ajustement des consignes données au modèle pour orienter le type de réponses attendues), nous avons fait en sorte qu'il agisse comme un médiateur pédagogique, capable de recentrer les questions hors sujet et de fournir des explications concises basées exclusivement sur les données de l'article.

Les premiers résultats sont encourageants, mais ambivalents. NewsChat semble particulièrement bénéfique pour les lecteurs ayant un faible intérêt de départ pour le sujet traité, l'interactivité avec l'IA semblant compenser le désintérêt initial et stimuler l'engagement. Cette piste est importante pour les médias, car elle suggère que des dispositifs adaptatifs pourraient aider à maintenir l'attention de publics qui auraient autrement quitté l'article rapidement. Mais ces bénéfices s'accompagnent de réserves.

L'introduction de cet assistant conversationnel peut réduire légèrement la crédibilité perçue de l'article chez certains lecteurs, en particulier chez ceux qui entretiennent déjà des attitudes négatives envers les *chatbots*. Les participants ont par ailleurs exprimé des inquiétudes concernant les hallucinations, les biais algorithmiques, le risque de dépendance excessive à l'outil ou la possibilité que l'assistant oriente la lecture au lieu de l'accompagner. Ces résultats exploratoires doivent bien sûr être replacés dans le contexte d'une technologie et des rapports à elle qui évoluent extrêmement rapidement.

Au-delà de leur caractère exploratoire, ces résultats sont des pistes intéressantes pour les médias et les institutions produisant des données. En améliorant la qualité et la fluidité de l'expérience de lecture, on peut favoriser l'engagement des publics à poursuivre la lecture, mieux comprendre le contenu, recommander l'article, revenir vers le média voire payer pour le contenu. Leur intérêt ne réside donc pas

seulement dans l'innovation formelle, mais dans leur capacité à soutenir une expérience de lecture plus accessible et plus active. Il convient néanmoins de rester prudent quant aux réserves et inquiétudes que génère l'usage de ces outils.

Leur déploiement appelle donc une certaine prudence. Un assistant de lecture ne devrait pas se substituer au travail journalistique, mais aider à l'éclairer : répondre à des questions précises, renvoyer aux passages pertinents de l'article, aider à

la compréhension des données, pallier un manque de littératie des données, et éviter de proposer une lecture trop fermée ou trop personnalisée. L'enjeu est moins de produire un parcours entièrement individualisé que d'offrir plusieurs portes d'entrée dans un même récit de données, sans affaiblir la rigueur du travail éditorial.

Antoine Flahault, « L'IA pourrait éclairer de manière significative les décisions dans le cadre d'une future pandémie » (page 52)

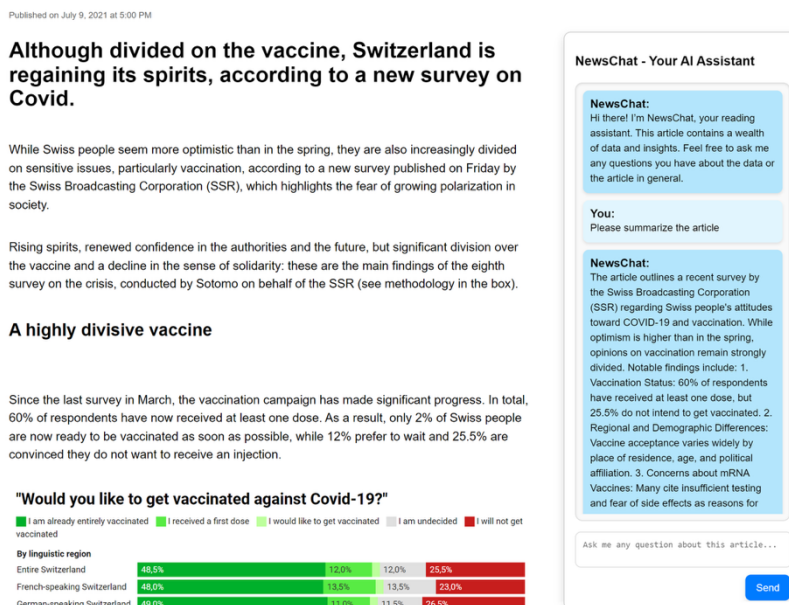


Figure 5. Capture d'écran du dispositif de *chatbot* NewsChat élaboré dans le cadre du projet PANDA

Chapitre 6. Perspectives pour une future crise sanitaire

Passer de résultats de recherche, parfois exploratoires et nécessairement partiels, à des recommandations opérationnelles est un exercice délicat. Il le serait d'autant plus s'il s'agissait de proposer aux autorités et aux médias une liste fermée de mesures à appliquer lors d'une future crise sanitaire, alors même que l'écosystème informationnel, le cadre légal et les technologies de production et de diffusion des données évoluent rapidement.

Nous proposons donc plutôt une série de perspectives : des pistes d'action fondées sur les résultats du projet PANDA, mises en dialogue avec la littérature scientifique et grise, ainsi qu'avec les entretiens et expériences menés auprès des différents acteurs impliqués dans la production, la médiatisation et la réception des données. Ces perspectives visent moins à prescrire une solution unique qu'à identifier les conditions d'un usage d'intérêt public des données en temps de crise.

Mettre en place un système de *monitoring* public multithématique

L'une des principales perspectives issues du projet PANDA est la nécessité de préparer, en amont de la crise, un dispositif public de *monitoring* mobilisable rapidement en cas de pandémie. La mise en ligne du tableau de bord Covid-19 de l'OFSP, accompagné d'une API permettant aux médias d'automatiser la collecte des données, a marqué un tournant important durant la pandémie. Le portail d'information sur les maladies transmissibles de l'OFSP constitue aujourd'hui une ressource riche, structurée et régulièrement mise à jour. Plus largement, opendata.swiss poursuit le travail de mise à disposition des données d'intérêt public. Ces évolutions constituent des bases importantes pour préparer de futurs dispositifs de suivi en temps de crise.

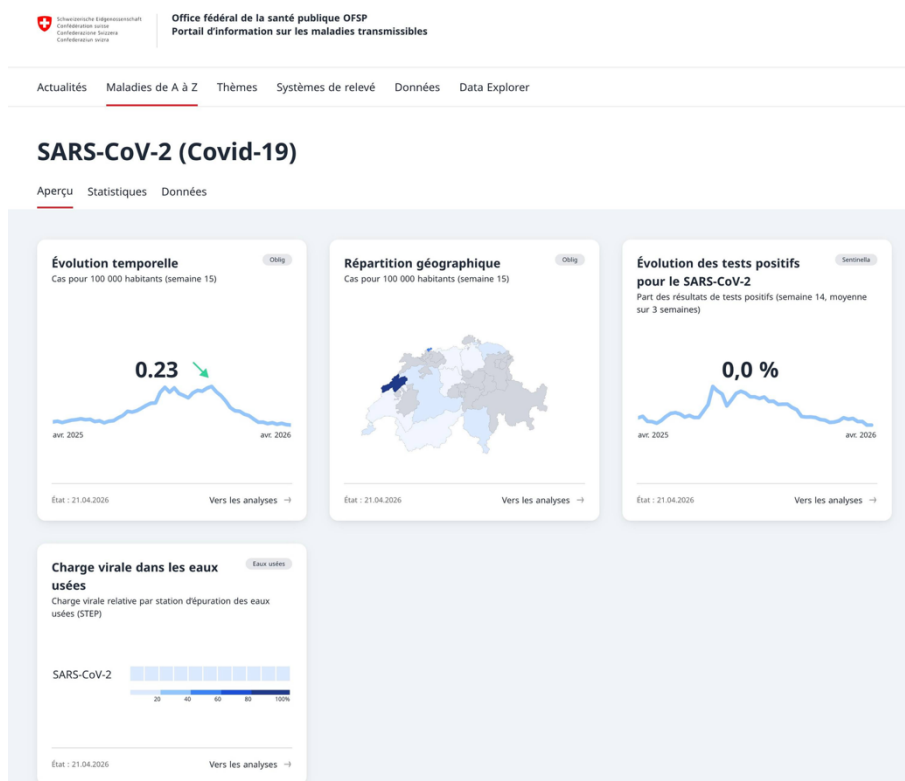


Figure 6. Copie d'écran du tableau de bord Covid-19 proposé sur le site de l'OFSP

Elles ne répondent toutefois pas à l'ensemble des besoins révélés par la crise du Covid-19. En contexte pandémique, l'information utile aux citoyens, aux médias et aux décideurs ne peut pas se limiter aux seuls indicateurs épidémiologiques. Comprendre la situation, en débattre, se forger une opinion, agir ou s'exprimer par les urnes suppose de disposer d'une vue d'ensemble qui va au-delà des seules données sanitaires, intégrant aussi les effets sociaux, psychologiques et économiques de la crise. Une telle vue d'ensemble graphique n'a pas véritablement existé durant la pandémie, ni du côté des autorités ni du côté des médias. Elle devrait inclure les principales données sanitaires, mais aussi des indicateurs relatifs à la pauvreté, à la précarité, à l'isolement, à la santé mentale, au chômage, au PIB, à l'activité économique ou aux faillites, parmi d'autres dimensions possibles. Des projets de l'OFS montrent déjà une volonté d'offrir de tels aperçus statistiques transversaux.

Plusieurs défis techniques doivent être traités. Le premier consiste à définir des temporalités de mise à jour adaptées : suffisamment rapides pour permettre un suivi utile de la crise et de ses effets, mais suffisamment robustes pour préserver la fiabilité des données. Un second défi consiste à identifier, en amont, les indicateurs clés susceptibles d'alimenter une telle plateforme, en dépassant une approche trop cloisonnée. Certains indicateurs existent déjà dans des formats adaptés, à commencer par les principaux indicateurs sanitaires mobilisés dans le tableau de bord de l'OFSP. D'autres sont collectés, mais selon des procédures, des niveaux de granularité ou des temporalités peu compatibles avec le suivi d'une crise. D'autres encore devraient faire l'objet d'une stratégie de collecte spécifique, permanente pour certains domaines, activable uniquement en période de crise pour d'autres.

Le numérique ouvre aussi de nouvelles possibilités pour documenter des dimensions plus difficiles à saisir par les seuls indicateurs statistiques classiques. Des sondages en ligne, des analyses de sentiment ou des outils d'analyse des cadrages médiatiques pourraient, avec les précautions méthodologiques nécessaires, compléter les données sanitaires,

sociales et économiques. L'objectif ne serait pas de multiplier les indicateurs, mais d'offrir aux décideurs, aux médias et aux citoyens une vue plus complète de la crise et de ses effets.

Un tel dispositif devrait être pensé comme une infrastructure de données, et non comme un simple tableau de bord. Il devrait proposer une API, des métadonnées robustes, des historiques, des définitions stables, des indications sur les limites des indicateurs et des visualisations conformes aux bonnes pratiques de lisibilité et de transparence.

Adapter la communication des données aux publics et aux objectifs

Les visualisations de données devraient être conçues à partir d'un objectif explicite : expliquer une situation, faciliter une comparaison, renforcer la confiance, corriger une croyance erronée, soutenir une décision pratique ou encourager l'engagement. Cet objectif doit guider les choix graphiques, le niveau de détail, le vocabulaire utilisé et la place accordée aux sources, aux limites et à l'incertitude.

Il est nécessaire de prévoir plusieurs niveaux de lecture. Le premier doit permettre une compréhension rapide du message principal ; les niveaux suivants peuvent donner accès aux détails méthodologiques, aux définitions, aux marges d'erreur, aux données brutes ou aux limites de l'indicateur. Cette logique de transparence graduée évite deux écueils : des graphiques trop simplifiés, qui donnent une impression trompeuse de certitude, et des visualisations trop chargées, qui découragent la compréhension.

La communication doit aussi tenir compte de l'hétérogénéité des publics. Tous ne disposent pas du même intérêt pour le sujet, du même degré de littératie statistique ni de la même confiance envers les sources. Les médias, administrations et acteurs publics devraient donc privilégier des visualisations lisibles par défaut, avec des repères stables, des libellés compréhensibles, des sources visibles et un accompagnement textuel clair. Lorsque cela est

pertinent, des dispositifs interactifs ou adaptatifs peuvent être utilisés pour approfondir la lecture, mais seulement s'ils servent réellement la compréhension.

L'enjeu n'est pas de produire des visualisations plus sophistiquées, mais mieux orientées : un bon graphique de crise doit aider le public à comprendre ce que les données permettent d'affirmer, ce qu'elles ne permettent pas d'affirmer et avec quel degré de prudence.

Structurer les interfaces entre institutions, médias et experts

La pandémie a montré l'importance des échanges entre autorités fédérales et cantonales, journalistes, chercheurs et experts. Une partie de ces échanges s'est construite dans l'urgence, souvent de manière informelle, grâce à des contacts personnels, des réseaux professionnels ou des collaborations ponctuelles. Ces circuits ont parfois permis de résoudre rapidement des problèmes de compréhension, de vérification ou d'accès aux données. Mais ils ont aussi produit des frustrations : interlocuteurs difficiles à identifier, réponses tardives, absence d'explication sur certains changements méthodologiques, demandes journalistiques mal comprises ou impossibles à traiter dans les délais.

Il serait donc utile de préparer, en temps normal, une interface de crise entre producteurs et utilisateurs professionnels de données. Celle-ci pourrait rester légère, mais devrait être clairement identifiée et rapidement activable. Elle aurait pour fonction de faciliter les échanges techniques, d'expliciter les choix méthodologiques, de signaler les changements dans les jeux de données, de recueillir les besoins prioritaires des utilisateurs et de faire circuler les informations utiles entre administrations, médias, chercheurs et experts.

Concrètement, cette interface pourrait prendre plusieurs formes : briefings techniques réguliers, listes de diffusion spécialisées, interlocuteurs de référence par domaine, documentation des métadonnées, journal des modifications des flux de

données, espace de questions-réponses, partage de scripts ou de modèles de données, voire co-construction de certains formats de publication. L'objectif n'est pas de donner aux médias un accès exclusif aux données, mais de garantir que les données publiques puissent être comprises, réutilisées et questionnées correctement par celles et ceux qui les médiatisent ou les analysent.

Cette structuration devrait aussi reconnaître le rôle des experts comme médiateurs entre données, institutions et publics. Des profils capables d'expliquer les indicateurs, leurs limites, leurs incertitudes et leurs usages possibles pourraient être mobilisés pour accompagner les journalistes et les administrations, sans se substituer à leur responsabilité respective. Une telle interface permettrait de réduire la dépendance aux relations informelles, tout en conservant la souplesse indispensable en période de crise.

Renforcer la littératie des données chez les professionnels et les publics

La pandémie a montré que l'accès aux données ne suffit pas à garantir leur compréhension. Interpréter un taux d'incidence, comparer des séries, comprendre une marge d'erreur, distinguer corrélation et causalité, ou saisir les limites d'un indicateur suppose des compétences qui ne sont pas également distribuées. La littératie des données doit donc être considérée comme une compétence clé, utile aussi bien aux citoyens qu'aux professionnels amenés à produire, commenter ou relayer des chiffres.

Cette littératie ne devrait pas être pensée uniquement comme une responsabilité individuelle des publics. Elle dépend aussi de la qualité des dispositifs d'information mis à leur disposition. Les institutions et les médias doivent donc travailler sur deux fronts : renforcer les compétences collectives face aux données (de la scolarité obligatoire aux formations professionnelles et universitaires, dans les dimensions statistiques, analytiques et critiques), mais aussi concevoir des contenus qui restent compréhensibles sans exiger un niveau d'expertise

trop élevé. Former les publics ne dispense pas de produire des indicateurs clairs, des visualisations lisibles, des définitions accessibles et des explications sur les limites des chiffres.

Concrètement, cette perspective invite à développer des ressources pédagogiques partagées : glossaires d'indicateurs, guides de lecture des graphiques, explications simples sur les marges d'erreur, les

changements de méthode, les révisions de séries ou les comparaisons entre territoires. Elle suppose aussi de renforcer la formation des journalistes généralistes (les journalistes de données disposant généralement de compétences plus pointues en la matière), des porte-parole institutionnels, des responsables politiques et des administrations à la lecture critique des données.

Conclusion

La pandémie de Covid-19 a montré que les données sont indispensables à la compréhension et à la gouvernance des crises sanitaires. Mais elle a aussi révélé leurs fragilités : définitions instables, formats peu exploitables, dépendance aux sources officielles, difficultés de coordination, surcharge informationnelle, inégalités de compréhension et tensions autour de l'incertitude.

Le principal enseignement de ce rapport est que la donnée ne devient utile publiquement qu'à certaines conditions. Elle doit être produite de manière fiable, accompagnée de métadonnées, vérifiée avec méthode, mise en récit avec prudence, visualisée de façon lisible et comprise par des publics aux compétences diverses. Autrement dit, l'enjeu n'est

pas seulement de disposer de plus de données, mais de construire les conditions sociales, techniques et démocratiques de leur intelligibilité.

Préparer une future crise suppose donc d'agir en amont : structurer les infrastructures de données, clarifier les interfaces entre institutions et médias, intégrer les dimensions sociales et économiques aux systèmes de *monitoring*, renforcer la littératie des données et concevoir des visualisations adaptées aux publics comme aux situations d'incertitude. C'est à cette condition que les données pourront pleinement jouer leur rôle dans le contexte suisse : soutenir (mais aussi monitorer voire questionner) la décision publique, nourrir le débat collectif, faciliter la coordination entre niveaux institutionnels et contribuer au bon fonctionnement démocratique.

Chapitre 7. Et si une pandémie arrivait aujourd'hui, que ferait-on différemment ?

Pour cette dernière partie, nous avons sollicité des contributrices et contributeurs externes au projet, dont l'expérience et l'expertise, pendant la pandémie, nous ont paru particulièrement pertinentes. À l'époque, ils et elles étaient (data)journalistes, rédacteur en chef, membre de la *task force*, chercheurs, médecin cheffe, etc. Il convient de signaler qu'un certain nombre de personnes contactées, représentant notamment des administrations et institutions publiques, ont décliné, pour diverses raisons, la proposition.

Nous remercions très chaleureusement les personnes qui ont accepté de se plier à l'exercice, en respectant le format imposé : un court texte orienté à la fois sur leur expérience de la pandémie en lien avec les données, dans leur fonction de l'époque, tout en se projetant sur l'avenir, en tentant de répondre à cette complexe question : « Et si une pandémie arrivait aujourd'hui, que ferait-on différemment ? ».

Les contributions sont présentées dans leur ordre d'apparition, au fil du rapport :

- Duc-Quang Nguyen, « La donnée ne peut plus rester exclusivement biologique »
- Mathieu Rebeaud, « La collaboration entre scientifiques, journalistes et gouvernements est centrale »
- Nicolas Willemin, « Les données sont celles que l'autorité veut bien communiquer »
- Titus Plattner, « Ouvrir les données publiques n'affaiblit pas les institutions »
- Olga Baranova, « L'architecture institutionnelle reste largement inchangée depuis la pandémie »
- Colin Porlezza, « Les rédactions devraient s'engager dans les outils collaboratifs et *open source* »
- Gaël Hürlimann, « Les médias doivent s'affranchir du travail ingrat de récolte et de mise au propre des données »
- Antonin Descampe, « Le datajournaliste devient un opérateur, capable de combiner des instruments hétérogènes »
- Monique Lehky Hagen et Diego Kuonen, « La littératie des données exige un changement de culture sociétal fondamental, profond et continu »
- Charles-Henry Groult, « Notre responsabilité est de faire preuve d'une extrême prudence, voire d'assumer une part d'incertitude »
- Valentin Tombez, « Une lecture brute des chiffres nourrit des interprétations trompeuses »
- Laurence Senn, « Il est essentiel de transmettre des chiffres avec l'analyse nécessaire à leur interprétation »
- Antoine Flahault, « L'IA pourrait éclairer de manière significative les décisions dans le cadre d'une future pandémie »

« La donnée ne peut plus rester exclusivement biologique »

Par Duc-Quang Nguyen, Docteur en biologie computationnelle de l'Université d'Oxford, journaliste de données au Temps (et au sein de Tamedia au début de la pandémie et jusqu'à l'été 2021).

L'expérience de la pandémie de Covid-19 en Suisse a mis en lumière certaines fragilités structurelles dans l'accès et le traitement des données de santé. On peut supposer que l'enjeu d'une future crise résiderait moins dans la multiplication des graphiques que dans la fiabilité technique et la pertinence des indicateurs choisis.

Un premier enseignement porte sur la fin nécessaire d'un certain « fédéralisme artisanal ». En 2020, l'absence de standards a contraint les rédactions à une agrégation manuelle de fichiers PDF. Si le portail des maladies transmissibles ([IDD](#)) et ses API constituent une avancée réelle, leur flexibilité face à un éventuel agent pathogène inédit reste encore à éprouver. La réactivité des médias dépendra sans doute de la capacité des administrations à adapter leurs flux en temps réel pour éviter que les journalistes ne redeviennent de simples collecteurs de données primaires.

Un second point de réflexion concerne la compréhension des chiffres par le public. J'ai été frappé par la fréquence des mésinterprétations des tableaux de bord Covid par les lecteurs, malgré la présence de notes méthodologiques détaillées. Cette expérience suggère que les précisions en bas de page ne suffisent pas toujours ; l'UX/UI d'un graphique pourrait, en soi, expliciter davantage les incertitudes et les limites méthodologiques. L'enjeu serait de mieux communiquer sur le non-dit des chiffres pour

lever toute ambiguïté, ce qui est essentiel pour une couverture rigoureuse d'une pandémie.

Cette rigueur gagnerait sans doute à s'accompagner d'indicateurs plus prospectifs. Les mesures classiques - cas, hospitalisations ou décès - sont par nature rétrospectives et accusent un retard systématique sur la réalité des contaminations. Une vision plus prédictive pourrait reposer sur l'analyse des eaux usées croisée avec les données de mobilité. Ces sources permettraient d'identifier des foyers émergents avant l'explosion des courbes épidémiques, agissant comme un outil de sensibilisation plus direct.

Par ailleurs, il semble que la donnée ne puisse plus rester exclusivement biologique. L'exemple du port du masque ou de la vaccination a montré que l'adhésion aux mesures sanitaires suit des logiques très clivantes. Une future couverture médiatique gagnerait à quantifier et à creuser les raisons de ces freins comportementaux. Documenter ces réalités sociologiques permettrait de potentiellement expliquer les disparités régionales de l'incidence plutôt que de simplement les constater, tout en reflétant la diversité des sensibilités face aux mesures instaurées.

Enfin, la pertinence de l'information semble également liée à sa granularité. Descendre à l'échelle de la commune ou du quartier pourrait rendre la donnée plus concrète pour le citoyen. À ce niveau, pour autant que la protection de la sphère privée demeure garantie, l'indicateur perdrait de sa froideur statistique pour devenir un outil d'évaluation du risque individuel plus tangible.

« La collaboration entre scientifiques, journalistes et gouvernements est centrale »

Par Mathieu Rebeaud, doctorant en biochimie pendant la pandémie, qui a participé à un réseau de chercheurs travaillant à évaluer la recherche sur le SARS-CoV-2 et à dénoncer la désinformation scientifique, notamment sur les réseaux sociaux.

Dans le contexte du COVID-19, qui s'est accompagné d'une infodémie et de la propagation de fake news, le filtre de la relecture scientifique et du temps assez long qui l'accompagne s'est avéré insuffisant. En concurrence avec les voix des experts scientifiques compétents et des journalistes scientifiques, nous avons assisté à la montée d'un nouveau type de clientélisme ou d'opportunisme qui s'est nourri de l'ignorance et/ou de l'incompréhension du grand public sur les données, comme comparer des résultats préliminaires avec des résultats confirmés.

Même pour un scientifique qui a l'habitude de lire de la littérature scientifique, la surabondance de nouvelles, parfois contradictoires, sur les décisions et la marche à suivre pouvait laisser perplexe. Par exemple, les premières informations sur la transmission du virus avec l'opposition entre gel hydroalcoolique et port du masque n'ont pas aidé à une compréhension des mesures par le grand public.

Rappelons-nous des premiers résultats présentés, relatifs à l'hydroxychloroquine (cf. [cette méta-analyse](#) à laquelle j'ai contribué), présentés dès mars 2020. Trop beaux pour être vrais, ils ont pourtant été repris partout, notamment par la presse grand public. Il était sans doute difficile de ne pas le faire, l'auteur de l'étude étant un professeur renommé et connu et dont les travaux pour la plupart avaient l'air sérieux.

Étant impliqué sur les réseaux sociaux (feu Twitter, dorénavant Bluesky) pour faire de la vulgarisation ou

parler de chats, j'ai dès le début de la pandémie perçu ce que l'accès aux données pouvait apporter. J'ai commencé en partageant des données de l'épidémie en Chine, mais surtout en Corée du Sud, ayant un ami journaliste habitant à Séoul. J'ai assez vite vu l'arrivée massive de publications ([plus de 8 000 dès fin avril 2020](#)), de données sur les malades, voire de propositions de traitements plus ou moins validés.

Une importante discussion a eu lieu aussi sur le rôle des autorités comme l'OMS qui avait créé une section sur son site pour combattre la désinformation et donner des liens utiles. Avec quelques collègues scientifiques, graphistes et journalistes, nous avons créé [Kezacovid](#) pour parler des nouvelles études, des données les plus pertinentes et partager des informations de qualité à destination du grand public. Avec une question centrale : Comment mieux partager les résultats et les données en temps de pandémie et lesquelles choisir ou nuancer ?

Ce que j'en retire aujourd'hui : il faudrait une meilleure collaboration entre scientifiques et journalistes, nuancer les propos et mettre en perspective l'avis d'une personne face au consensus si les résultats sont étonnants, se méfier des données qui ont l'air trop belles pour être vraies. Possiblement ne pas donner d'écho à des données de mauvaise qualité ou qui ont l'air d'affoler les réseaux sociaux, vu que souvent la caisse de résonance est amplifiée dès que reprise par les médias dits traditionnels. La collaboration et la vérification entre scientifiques, journalistes et gouvernements (démocratiques) seront donc centrales, malgré le temps court de l'information dans ce genre de situation. D'où l'importance d'un service public en bonne santé lui aussi.

« Les données sont celles que l'autorité veut bien communiquer »

Par **Nicolas Willemin**, journaliste à la rubrique neuchâteloise du journal *ArcInfo*, après avoir été rédacteur en chef de *L'Impartial/L'Express* pendant dix ans.

Le vendredi 13 mars 2020, les Neuchâtelois découvrent que le coronavirus dont on parle depuis le début de l'année va provoquer la fermeture, pour de longues semaines, de la plupart des activités sociales. À la rédaction d'*ArcInfo*, une poignée de journalistes travaillent sur ce sujet depuis le mois de janvier. Ils tentent d'avoir des informations et des données sur la progression de ce virus, particulièrement dans la région couverte par notre quotidien, soit le canton de Neuchâtel et les régions avoisinantes. Pour un quotidien régional comme le nôtre, le défi est de gérer l'afflux d'informations et, petit à petit, de données chiffrées que nous recevons pour informer correctement nos lecteurs. Sans minimiser ce qui se passe, mais sans amplifier non plus la crise qui arrive. Nous nous concentrons sur ce qui se passe chez nous, sur le nombre de cas : celui des infections puis celui des malades et enfin celui des morts.

Par la suite viendront les simulations chiffrées pour anticiper l'évolution de la pandémie, son degré de propagation. On ira jusqu'à se plonger dans les eaux usées pour sortir des chiffres. Et ces chiffres arrivent assez vite. Mais une fois publiés, comment les analyser, comment les mettre en perspective, comment les décortiquer pour arriver à fournir une information pertinente ? Le Canton publie des chiffres, la Confédération aussi, et parfois, il y a des petites différences.

La plupart des journalistes que nous sommes ne sont pas forcément des spécialistes en traitement des données. On découvre ainsi les fameux R_0 et R_e , soit le nombre de nouveaux cas qu'une seule personne infectée et contagieuse générerait en moyenne, et sa version effective calculée a posteriori. Quel est le R_e

dans le canton de Neuchâtel ? Comment va-t-il évoluer ? Comment le comparer avec des cantons voisins ? L'Office fédéral de la santé publique (OFSP) met à disposition des fichiers très détaillés, mais difficilement compréhensibles pour des journalistes qui ne sont pas des « data scientists ». Et pourtant les lecteurs attendent des explications régionales ou locales.

Et surtout, nous journalistes sentons bien que, derrière tous ces chiffres donnés, il y a un discours officiel des autorités qui tentent de montrer qu'elles contrôlent la situation. Et que nous les médias, nous devons nous baser sur les chiffres que l'on met à notre disposition. Et qui sont ceux qu'on veut bien nous donner. Car il est difficile, voire impossible, d'en avoir d'autres.

Il est clair qu'aujourd'hui, avec les outils offerts par l'intelligence artificielle, il serait plus facile, surtout pour nous journalistes généralistes régionaux, de décortiquer tous les fichiers de données qui étaient alors mis à notre disposition. Mais il ne faut pas se faire d'illusions : ce ne sont que des outils : les données fournies resteront celles que l'autorité veut bien communiquer.

Quelles leçons tirer de ce qui s'est passé au printemps 2020 avec l'arrivée d'une nouvelle catastrophe sanitaire ? Probablement aurons-nous un regard plus critique sur le flux de données transmis par les autorités. Peut-être pourrions-nous trouver d'autres sources de données venant d'acteurs extérieurs. Mais il est clair que les journalistes devront maîtriser de plus en plus le traitement des données, apprendre à utiliser efficacement les outils d'intelligence artificielle et garder un esprit critique sur les informations transmises. Tout en restant fidèle à une règle de base de notre profession : savoir raconter une histoire avec des êtres humains.

« Ouvrir les données publiques n'affaiblit pas les institutions »

Par Titus Plattner, journaliste et chef de projets d'innovation au sein de l'AI Lab de Tamedia. Il a dirigé l'équipe en charge des formats interactifs durant la pandémie et a coécrit *La Première Vague* (Slatkine, 2021), un livre sur la gestion de la pandémie en Suisse.

En 2020, des équipes hospitalières consultaient nos tableaux de bord journalistiques faute de données publiques fiables. Cinq ans plus tard, la Suisse est-elle prête ?

Dès le 25 février 2020, dans les rédactions de Tamedia, nous étions une petite équipe à recenser les cas de COVID-19 en Suisse, rafraîchissant les sites cantonaux plusieurs fois par jour. Nos données reflétaient parfois le nombre de tests positifs avec 48 heures d'avance sur l'OFSP. Des équipes hospitalières consultaient nos tableaux pour leurs points quotidiens. Deux épidémiologistes nous ont même demandé l'accès à nos données brutes pour affiner leurs projections.

Au sein des administrations, les informations transitaient encore via des chaînes de traitement dignes d'une autre époque. Durant la crise, des résultats de tests ont réellement été transmis par fax entre laboratoires et autorités cantonales, avant d'être ressaisi manuellement dans des tableurs. En plus, la primeur de ces données, voire le contrôle d'une partie d'entre elles, étaient jalousement réservés aux décideurs et décideuses politiques.

« Database Hugging Disorder », aurait diagnostiqué le statisticien suédois Hans Rosling : ouvrir les données publiques n'affaiblit pas les institutions. Au contraire, cela éclaire les décisions de tout le monde. La preuve était faite en temps réel.

En période de crise sanitaire, les données devraient être une infrastructure, pas un instrument de pouvoir.

Donner une information tardive ou fragmentée revient à piloter dans la tempête avec la météo d'avant-hier. Stopper les visites en EMS deux jours plus tôt, interdire les rassemblements plus vite, prononcer un confinement avec quelques jours d'avance : tout cela aurait sauvé des vies.

Alors, a-t-on progressé ? En partie. La Suisse dispose désormais d'opendata.swiss et la LMETA, entrée en vigueur en 2024, pousse vers une administration « priorité au numérique ». Mais une prochaine crise remettra à nu les mêmes points de friction : accès, formats, coordination entre cantons et Confédération, tentation de réserver l'information à un cercle restreint.

Que faire différemment ? Se préparer avant la crise. Définir un jeu de données minimal indispensable (nomenclature stable, métadonnées, calendrier, versionnage) publié en continu via des API documentées, pas à coups de PDF. Le tester : exercices réguliers, simulations de montée en charge, vérification des chaînes de remontée numérique. Clarifier la gouvernance en amont : en 2020, l'OFSP n'a obtenu le soutien de l'Office fédéral de la statistique que des mois après le début de la pandémie... alors qu'ils sont chapeautés par le même département. Si la coordination échoue sous un même toit, elle n'a aucune chance entre 26 cantons sans structure préétablie. Enfin, standardiser un accès sécurisé (API + connecteurs de type MCP) pour brancher rapidement des outils d'analyse, y compris des grands modèles de langage.

Et côté médias ? Nous devons aussi faire mieux : en 2020, nous avons parfois diffusé des chiffres sans suffisamment en documenter les limites. Exiger la transparence des autres suppose de se l'appliquer d'abord.

« L'architecture institutionnelle reste largement inchangée depuis la pandémie »

Par **Olga Baranova**, cofondatrice de CH++, association créée en janvier 2021 avec notamment l'épidémiologiste Marcel Salathé et l'entrepreneur Hannes Gassert, qui travaille sur des sujets de politique numérique, d'information et, plus largement, de sécurité.

La pandémie de Covid-19 a rappelé une évidence : la gestion d'une crise sanitaire dépend autant de la solidité des infrastructures que de la rapidité des processus décisionnels. En Suisse, la crise a mis en lumière des forces indéniables - la qualité de la recherche scientifique et la robustesse des institutions - mais aussi des faiblesses structurelles. Les systèmes d'information se sont révélés fragmentés et inégalement matures, l'accès aux données parfois trop lent, et la communication institutionnelle souvent en décalage avec la vitesse des médias et des réseaux sociaux.

Depuis lors, certaines leçons ont été tirées. Le projet d'e-ID étatique constitue un premier élément d'une infrastructure numérique de base. Une telle identité numérique pourrait faciliter la mise en place rapide de services publics numériques sûrs et respectueux de la protection des données. L'expérience du contact tracing numérique l'a montré : l'efficacité de ces outils dépend largement de l'intégration et de l'automatisation des processus. Pendant la pandémie, la nécessité d'interventions humaines - par exemple pour l'émission et la transmission des codes de notification - a entraîné des retards considérables et limité l'impact du système.

Autre évolution importante : la loi sur l'utilisation des moyens électroniques pour l'exécution des tâches des autorités (LMETA). Elle consacre notamment le

principe de l'*open government data*, selon lequel les données produites par l'administration fédérale doivent être accessibles au public dans des formats ouverts et réutilisables. En période de crise, l'ouverture des données est une condition pour permettre aux chercheurs, journalistes, entreprises ou organisations de la société civile d'analyser la situation, de développer des outils d'information et de contribuer à l'effort collectif. Le principal bémol reste toutefois l'échelle : pour l'instant, ces obligations ne concernent que la Confédération.

Ces avancées sont réelles. Mais elles ne doivent pas donner l'illusion que la Suisse est désormais pleinement prête pour la prochaine crise. Sur le plan institutionnel, l'architecture institutionnelle reste largement inchangée depuis la pandémie. Les logiques de silo persistent et les dynamiques entre la Confédération et les cantons peuvent encore ralentir les flux décisionnels. À cela s'ajoute un contexte budgétaire plus restrictif, qui pèse notamment sur les investissements dans le numérique et dans le domaine de la formation, de la recherche et de l'innovation.

Surtout, la prochaine pandémie n'arrivera certainement pas dans un environnement stable. Elle pourrait survenir dans un contexte de tensions géopolitiques, de cybermenaces accrues et de pression croissante sur les infrastructures critiques. Dans ce paysage de multicrises, la résilience numérique et institutionnelle devient un enjeu stratégique. La pandémie de Covid-19 a été un test. La prochaine crise dira si les leçons ont réellement été tirées.

« Les rédactions devraient s’engager dans les outils collaboratifs et open source »

Par Colin Porlezza, professeur associé en journalisme numérique et directeur de l’Institute of Media and Journalism à l’Università della Svizzera italiana (USI). Il est également membre de la Commission fédérale des médias (COFEM).

La pandémie a constitué un « moment critique » pour le journalisme. Les rédactions ont dû réorganiser rapidement leurs processus de travail autour de la production à distance, tout en développant le journalisme de données et la couverture en direct. La crise a également accru la valeur des informations fiables, renforçant le rôle du journalisme en matière de vérification et d’explication. Dans l’ensemble, la pandémie n’a pas simplement chamboulé les routines journalistiques, elle a mis en évidence le rôle central que jouent des rédactions résilientes et dignes de confiance dans les sociétés démocratiques.

Dès lors, comment les rédactions devraient-elles faire face aux futures pandémies ? Une réponse possible émerge du volet de recherche consacré à la technologie d’intérêt public (*Public Interest Technology*, PIT) désignant le développement et la gouvernance de technologies numériques orientées vers la responsabilité, la transparence et la participation démocratique, plutôt que vers la seule efficacité ou le profit. Appliquée aux médias, la PIT implique le développement et l’utilisation d’infrastructures technologiques (outils open source, systèmes algorithmiques, plateformes de données) qui soutiennent les fonctions essentielles du journalisme, notamment la vérification, l’indépendance et l’intérêt public.

Dans cette perspective, les rédactions ne se contentent pas d’utiliser des technologies : elles peuvent aussi agir comme des laboratoires technologiques publics, expérimentant et façonnant des outils qui reflètent des valeurs d’ouverture, de transparence et de collaboration. Cela est

particulièrement pertinent en contexte de crise, où les flux d’informations s’accroissent, l’incertitude est forte et les risques de désinformation augmentent, d’autant que les données pertinentes proviennent souvent d’institutions telles que les administrations publiques ou des organisations comme Our World in Data.

Les rédactions devraient ainsi s’engager activement dans le développement et la mise en œuvre collaborative d’outils d’investigation open source, permettant une vérification au-delà des frontières et réduisant la dépendance à des plateformes propriétaires. Les outils de suivi des données épidémiques ou d’analyse des politiques de santé publique pourraient être partagés et améliorés collectivement, renforçant à la fois l’exactitude journalistique et la responsabilité institutionnelle. Cela vaut également pour les outils algorithmiques qui doivent rester soumis à un contrôle éditorial et alignés sur des principes de responsabilité.

Les rédactions devraient donc investir dans une culture de l’IA et de collaboration interdisciplinaire, en travaillant avec des *data scientists*, des experts en santé publique et des technologues civiques pour coproduire des infrastructures de connaissances au service du public qui garantissent des informations précises et fiables. Cette perspective nous rapprocherait d’un journalisme comme forme de production de connaissances ancrée au sein de publics et d’acteurs en réseau. En se positionnant comme des participants actifs au développement de technologies d’intérêt public, les médias pourront renforcer leur rôle démocratique et contrebalancer l’influence déjà massive des écosystèmes d’information pilotés par les plateformes. Une telle approche consoliderait non seulement la capacité du journalisme à faire face aux crises, mais contribuerait également à une sphère publique plus inclusive.

« Les médias doivent s’affranchir du travail ingrat de récolte et de mise au propre des données »

Par **Gaël Hürlimann**, co-rédacteur-en-chef du quotidien *Le Temps* de 2015 à 2021, aujourd’hui directeur de la communication à l’EPFL (École polytechnique fédérale de Lausanne).

Au *Temps*, nous avons l’habitude de cette boutade pour réveiller notre auditoire, qu’il s’agisse des journalistes réunis pour des bilans réguliers, de lecteurs lors de nos rencontres en rédaction ou d’annonceurs en rendez-vous d’affaires : « Savez-vous quel est l’article le plus lu de l’année 2020 ? » Et la réponse : « Il s’agit, et de loin, du [tableau de bord L’expansion du Covid en Suisse et dans le monde](#) mis en place par le datajournaliste Paul Ronga ». À ce succès d’audience pour un format qui était alors en train de perdre de son aura, s’ajoute une reconnaissance par les pairs : la première place de la catégorie « coup de cœur » du prix Suva des médias cet « ovni » journalistique n’entrant pas dans les autres catégories, « presse » ou « audiovisuel ».

Pendant de longs mois, la récolte de données fiables a constitué la plus grande difficulté pour garantir la pertinence de cette page. Entre les multiples sources (Office fédéral de la statistique, cantons, Université Johns Hopkins, unité OpenZH, ...), les changements de mode de comptabilisation en cours d’année, les fréquences de mise à jour aléatoires et les multiples formats techniques, notre journaliste, aidé du développeur Ivo Marques, a dû effectuer une veille permanente et des changements dans la manière de récolter, nettoyer, « conformer » et parfois corriger les données pour produire un tableau de bord fiable.

Mais nous n’en sommes pourtant pas encore au journalisme : cet effort sur la data a pris une place considérable même si la petite communauté des

datajournalistes romands a le plus souvent joué le jeu du partage d’information : données en open source, identification de fournisseurs fiables et échanges réguliers. Mais cela a pourtant pris le pas sur l’autre volet, essentiel pour la profession : l’application des critères classiques du journalisme à un jeu de données. Je me souviens d’échanges nourris avec d’autres journalistes de données qui estimaient devoir mettre à disposition des internautes le plus de chiffres, de critères de tri, d’options d’affichage, de tableaux de chiffres bruts, etc. Heureusement, ce n’était pas la position de Paul. Car le travail journalistique nécessite de faire des choix : quelles données sont pertinentes ? Quel impact sur la compréhension des chiffres et leur interprétation lorsqu’on choisit une échelle logarithmique plutôt que linéaire ? Qu’est-ce qui n’a pas d’importance (et ne mérite pas d’être montré) ? Qu’est-ce que les chiffres et les graphiques ne disent pas (et qu’on exprimera plutôt dans un paragraphe de texte) ?

Si une nouvelle crise majeure devait survenir (épidémie, crise énergétique, enjeux climatiques, ...), je plaide pour que la profession et les fournisseurs de données aient tiré les leçons et s’entendent sur la mise à disposition de données fiables. Les médias ne perdraient pas de temps et d’énergie sur une partie du travail qui constitue certes le socle nécessaire au travail journalistique mais qui n’offre pratiquement pas de valeur ajoutée. Une collaboration entre tous, une « centrale » réunissant par hypothèse médias, Office fédéral de la statistique et pourquoi pas hautes écoles permettrait à chaque média de développer sa spécificité : quelle ligne éditoriale, quelle grille de lecture et par là quelle valeur ajoutée spécifique appliquer aux données fiables mises en commun ?

« Le datajournaliste devient un opérateur, capable de combiner des instruments hétérogènes »

Par Antonin Descampe, ingénieur, professeur en journalisme à l'UCLouvain et responsable du *Media Innovation and Intelligibility Lab*.

Les mutations techniques, économiques et sociales du journalisme renforcent le rôle stratégique du datajournalisme, désormais reconnu mais en transformation accélérée. Les enquêtes récentes décrivent des professionnels autodidactes et polyvalents, travaillant sous forte contrainte de temps, de ressources et d'accès à des données fiables, et dont la légitimité et la visibilité ont été accrues par la pandémie de COVID-19. Parallèlement, la généralisation de l'intelligence artificielle et de l'automatisation dans les processus éditoriaux impose aux journalistes d'en comprendre les usages et d'en questionner les effets sur la production et la narration de l'information. Dans ce contexte, il est intéressant de s'interroger sur ce qu'une formation en datajournalisme se doit aujourd'hui de transmettre à de futurs journalistes. Certaines compétences restent centrales (culture statistique de base, rigueur méthodologique, transposition des valeurs journalistiques) tandis que d'autres émergent. Parmi celles-ci, une littératie de l'intelligence artificielle, notamment générative, devient indispensable. Elle suppose de dépasser l'idée qu'il suffirait d'apprendre à « prompter » pour être en mesure de collecter, nettoyer ou identifier des tendances ou des motifs dans de grands ensembles de données. Ainsi, une compréhension minimale des modèles de langage permet d'en saisir les limites : leur caractère fondamentalement non-déterministe les rend impropres à remplacer la précision et la reproductibilité d'un code en Python ou d'une formule de tableur.

Ces modèles n'en sont pas moins des outils complémentaires pertinents pour le datajournaliste. Trois usages principaux peuvent être distingués.

D'abord, un support de réflexion et d'exploration, utile pour formuler des hypothèses, identifier des angles ou questionner une démarche. Ensuite, un outil d'analyse de contenus textuels ou visuels (extraction d'entités nommées, détection de sentiments, reconnaissance d'objets), pertinent lorsque ce type de données est mobilisé. Enfin, un levier d'assistance au codage, qui abaisse fortement les barrières à l'entrée pour produire des scripts réalisant des opérations complexes. Ce dernier usage ne dispense toutefois pas d'une compréhension des logiques algorithmiques et d'une pensée computationnelle. Piloter efficacement ces outils implique de savoir décomposer un problème en étapes, et de maîtriser des notions fondamentales (variables, boucles, conditions) afin d'évaluer la validité du code généré. L'enjeu n'est plus de maîtriser la syntaxe, mais de comprendre ce qui peut être automatisé, comment, et avec quelles limites. Au-delà de ces usages, une « méta-compétence » se dessine : la capacité à orchestrer un écosystème d'outils. Tableurs, modèles de langage, environnements de programmation et outils de visualisation doivent être mobilisés de manière complémentaire. L'apprentissage porte dès lors moins sur chaque outil pris isolément que sur leur articulation : savoir quand mobiliser tel dispositif, pour quel objectif, et comment faire circuler les données entre les différentes étapes du traitement. Le datajournaliste devient ainsi un opérateur, capable de combiner des instruments hétérogènes pour produire une information fondée sur les données. Cette évolution ne réduit pas l'exigence critique : elle la déplace. Comprendre les logiques, les biais, les écueils, et les finalités des outils mobilisés demeure essentiel. Mais elle permet aussi de s'affranchir d'une partie des contraintes techniques, pour en réalité recentrer la formation sur les enjeux fondamentaux du datajournalisme : enquêter, interpréter et mettre en récit des données.

« La littératie des données exige un changement de culture sociétal fondamental, profond et continu »

Par Dr. med. Monique Lehky Hagen eMBA healthcare et Prof. Dr. ès sc CStat PStat Diego Kuonen, respectivement médecin généraliste et professeur en sciences de données à l'Université de Genève. Ils ont lancé en juillet 2020 un appel à une campagne de littératie des données.

La pandémie de COVID-19 a mis en évidence les effets d'un déficit de culture de littératie des données dans la population, mais aussi au sein des instances politiques, médiatiques et scientifiques. La vitesse inédite d'acquisition de connaissances provisoires et parfois (apparemment) contradictoires a amplifié ces difficultés. Pour contribuer à une prise de conscience collective, nous avons lancé en juillet 2020 un appel à une [campagne nationale en matière de littératie des données](#), qui a abouti à l'élaboration d'une [Charte suisse de littératie des données](#) avec les Académies suisses des sciences, publiée en 2024. La littératie des données désigne les compétences permettant de collecter, gérer, évaluer et utiliser les données de manière critique et rigoureuse, dans leur contexte, en respectant les principes éthiques et de protection des données. Elle suppose un apprentissage tout au long de la vie et une culture de feed-back continu entre producteurs et consommateurs de données. Dans une société où cette culture est ancrée, les données ne sont plus perçues comme vraies ou fausses de manière absolue : leur valeur dépend du contexte, qui lui-même évolue. Cela exige de l'inclusion (au niveau des données et des acteurs), mais aussi de la transparence sur les méthodes, une mise en contexte systématique et une discussion critique, autant de pratiques qui permettent de lutter contre la désinformation. Afin d'évoluer vers une société plus résiliente et éviter une répétition des polarisations vécues, nous appelons vivement les instances politiques et professionnelles à mettre rapidement en route, de façon collaborative, les mesures nécessaires pour :

- 1) Développer la littératie des données dès l'école enfantine puis tout au long de la vie, avec une culture du questionnement : Quelle est la véracité probable de l'information ? Est-elle cohérente ? Quelles sont ses sources et sont-elles crédibles ? Quels biais sont possibles ? Les données sont-elles de qualité ? Sur quelle méthodologie sont-elles fondées ? Permettent-elles vraiment les conclusions qu'on en tire ?
- 2) Conduire les décideurs politiques et les journalistes à communiquer sur des données incertaines ou évolutives, de manière éthique et dans le respect du cadre juridique : Qu'est-ce que je peux / dois faire avec ces données afin de les mettre au profit de la société ? Comment dois-je publier les données sans porter atteinte aux intérêts personnels et éthiques des personnes ou sujets concernés, mais aussi aux règles existantes ou à la protection des données ?
- 3) Développer des systèmes de collecte de données flexibles et activables rapidement en cas de crise, impliquant patients et professionnels du domaine ambulatoire, pour accélérer la production de connaissances utiles avant que les situations ne deviennent critiques (par exemple l'hospitalisation des patients avec des symptômes nécessitant le traitement aux soins intensifs).
- 4) Établir des lignes de bonne pratique partagées entre institutions, médias et société civile, fondées sur les principes de la Charte suisse de littératie des données, et les rendre accessibles au grand public.

La littératie des données exige un changement de culture sociétal fondamental, profond et continu, d'autant plus à l'heure de l'intelligence artificielle. La pandémie de Covid-19 a montré que les données pouvaient faire l'objet d'utilisation manipulatoire, inadéquate et arbitraire. Il est donc urgent de promouvoir une culture partagée de littératie des données, sans laisser personne de côté, pour favoriser un fonctionnement inclusif, participatif et autodéterminé de nos sociétés, bien au-delà de la seule préparation à une prochaine pandémie.

« Notre responsabilité est de faire preuve d'une extrême prudence, voire d'assumer une part d'incertitude »

Par Charles-Henry Groult, responsable du service vidéo du quotidien français Le Monde pendant la pandémie et actuellement directeur adjoint de la rédaction du même journal.

« 300 000 ». Quand une journaliste scientifique du Monde a partagé, en conférence de rédaction, une première projection du nombre de morts pour la France en cas d'absence de mesures de prévention, je me souviens avoir été saisi d'un double vertige : à quoi ressembleront notre pays, nos villes, nos familles dans quelques mois ? Et quelle responsabilité du Monde face à ce séisme ?

Il devait être midi, heure de notre point quotidien, le 12 mars 2020. Le 17, la France était confinée. Aux questions ont vite succédé deux réponses presque aussi vertigineuses. La réponse de nos lecteurs, de notre audience, d'abord. Massive, historique. Jamais autant de personnes ne se sont connectées, inscrites ou abonnées au Monde qu'en mars 2020 : 315 millions de visites. J'y ai vu un marqueur fondamental : quand la crédibilité de l'information devient un enjeu réellement vital (avoir les bonnes informations peut avoir un impact sur ma vie, celle de mes proches), le public se tourne massivement vers les sources traditionnelles d'information reconnues pour leurs standards de qualité, leurs sources et leur quête d'impartialité.

Deuxième réponse, celle de la rédaction. La crise Covid est survenue pile au moment du déménagement de notre rédaction dans un autre quartier de Paris. Avant même de rejoindre nos nouveaux locaux, nos ordinateurs ont donc tous été livrés en 48 h... à nos domiciles. La crise Covid a servi d'accélérateur de certaines évolutions de la rédaction : Slack comme outil de communication prioritaire, le mail n'étant plus utilisé que pour de rares communiqués internes, le partage de lexiques et de bonnes pratiques via des Google Doc ressources

(« précisons si possible qu'il s'agit de chiffres « déclarés » ou « officiels », évitons le terme de « record », privilégions « hausse inédite », « nouveau bond », ne confondons pas chloroquine et hydroxychloroquine, » etc.). Le choix fut fait, courant avril je crois, de donner les chiffres des contaminations avec prudence et de ne plus en faire des titres à eux seuls, car ils dépendent des politiques de dépistage. Les entrées et sorties en réanimation, les hospitalisations et les décès devaient être privilégiés en titraille. En plus d'écrire leurs propres sujets, les journalistes des services Sciences et Planète reliaient les productions des autres services, dont les scripts du service vidéo, pour s'assurer que tous et toutes partageaient les mêmes sources et mêmes données. À un instant T, Le Monde devait offrir une photographie homogène de la situation, quel que soit le format ou le support. Le service Décodeurs, qui réunit l'essentiel de nos datajournalistes, fut un levier déterminant pour rendre lisible des jeux de données massifs et nous aider à arbitrer dans le choix des sources. Devant des enjeux aussi graves, la responsabilité du Monde est aussi de faire preuve d'une extrême prudence, voire d'assumer une part d'incertitude : dans nos vidéos « À quand la fin du Covid », les plus regardées, le choix fut constamment fait de toujours partager plusieurs scénarios, tantôt optimistes (un virus progressivement comparable au rhume), tantôt moins (un virus « fuyant », plus proche de la grippe, et sans vaccin efficace), appuyés sur une poignée de publications scientifiques ultra respectées : Science, Nature, bases classiques (Scopus, PubMed), grandes universités, structures publiques européennes et américaines.

Responsabilité, adaptation et coordination, voici peut-être les trois mots que je retiens de cette séquence et qui restent encore aujourd'hui des défis pour nous.

« Une lecture brute des chiffres nourrit des interprétations trompeuses »

Par **Valentin Tombez**, journaliste de données à la RTS, durant la pandémie et encore aujourd’hui.

La pandémie de Covid-19 l’a montré : les chiffres alimentent le débat public, orientent les décisions politiques, mais peuvent aussi accroître l’anxiété d’une partie de la population. Dans un tel climat, la nécessité de contextualiser les données, avec rigueur et rapidité, devient centrale. Pourtant, cette contextualisation n’a pas toujours été réalisée, ou pas immédiatement.

À la fin de l’automne 2021, la Suisse est frappée par une cinquième vague, alors que la campagne de vaccination bat son plein. Une part importante de la population est vaccinée, tandis qu’une autre hésite encore. Le nombre d’hospitalisations repart à la hausse. Mais, cette fois, les données distinguent les malades selon leur statut vaccinal. À première vue, le nombre de personnes vaccinées hospitalisées paraît presque aussi élevé que celui des personnes non vaccinées. Cette lecture brute des chiffres nourrit immédiatement des interprétations trompeuses, largement relayées chez les opposants à la vaccination.

Ces données avaient juste besoin d’être rapportées à la taille des populations concernées. Or la population vaccinée était alors bien plus nombreuse que la population non vaccinée. Une normalisation révèle, selon les groupes d’âges, des taux d’hospitalisation 10 à 20 fois plus élevés pour les non-vaccinés. Un point que nous avons expliqué et mis en évidence, entre autres, dans notre traitement graphique. Ce principe est élémentaire en statistique. Pourtant, durant une pandémie, les réflexes les plus fondamentaux peuvent s’émousser. D’où l’importance de revenir sans cesse aux bases de l’interprétation des données.

Les très nombreux retours et partages suscités par ces publications témoignent, à mon sens, de l’intérêt

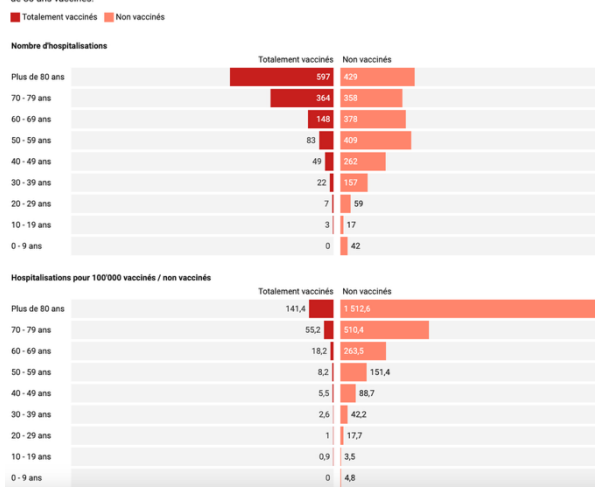
- voire du besoin - d’un large public pour des données rigoureusement contextualisées. Il convient cependant de ne pas surinterpréter ces chiffres qui ne permettent pas, à eux seuls, de conclure à l’efficacité d’un vaccin, mais établissent un état des lieux plus honnête et plus intelligible de la situation.

D’autres exemples confirment ce besoin constant de contextualisation. Les retards administratifs, d’abord, faisaient parfois remonter certaines données avec plusieurs jours de décalage, ce qui rendait les chiffres les plus récents fragiles, voire trompeurs, car incomplets. Les décisions politiques, ensuite, avaient un effet direct sur les indicateurs suivis : la gratuité, puis la fin de la gratuité des tests, a profondément modifié le volume des dépistages et rendaient certaines comparaisons temporelles trompeuses.

Ce travail de contextualisation ne se limite évidemment pas aux pandémies. Mais dans un climat à la fois anxiogène, conflictuel et fortement polarisé, son importance semble démultipliée. Dans l’hypothèse d’une future crise sanitaire, il resterait donc essentiel de ne pas seulement publier les données, mais aussi de systématiquement les expliquer, de les situer, et de les rendre lisibles pour le plus grand nombre.

Covid-19, 5e vague: hospitalisations en Suisse selon le statut vaccinal

Admissions à l’hôpital liée au Covid-19 en Suisse entre le 11 octobre et le 12 décembre 2021. Dans le second graphique, les hospitalisations sont comparées à population égale, par exemple les hospitalisations des plus de 80 ans vaccinés sont rapportées à la population des plus de 80 ans vaccinés.



« Il est essentiel de transmettre des chiffres avec l'analyse nécessaire à leur interprétation »

.....
Par Laurence Senn, médecin cheffe responsable de l'unité d'hygiène, prévention et contrôle de l'infection, service des maladies infectieuses du CHUV à Lausanne, pendant la pandémie, et encore aujourd'hui.

Lors d'une pandémie liée à un pathogène hautement virulent comme le SARS-CoV-2, les hôpitaux sont rapidement mis à contribution pour la prise en charge des cas les plus sévères. L'évolution du nombre d'admissions en lien avec l'infection, les caractéristiques des cas et le taux de décès, ainsi que le taux d'occupation des unités de soins intensifs et des unités standard, et le report d'interventions électives, sont des indicateurs clés pour évaluer la morbi-mortalité de l'infection, mesurer son impact sur le système de soins stationnaires et guider les mesures de prévention et contrôle de la propagation de l'agent pathogène.

Si la production de certains de ces indicateurs issus des bases de données administratives et du laboratoire a pu être automatisée, elle reste manuelle pour d'autres et nécessite des ressources ; cela concerne notamment certaines caractéristiques des cas et les interventions électives reportées. Ces indicateurs sont transmis régulièrement à la direction de l'hôpital qui les fait suivre au département de la santé du canton et à l'OFSP. Certains d'entre eux sont également communiqués aux médias, par devoir de transparence et d'information.

La pression des médias avec des délais rédactionnels souvent très courts et la forte sollicitation des experts impliqués dans la prise en charge clinique et les

différentes cellules de conduite de crise ont pu amener les hôpitaux à transmettre des chiffres sans les accompagner de l'analyse qui aurait été nécessaire à leur interprétation et à une meilleure compréhension des mesures de prévention et contrôle mises en place dans les hôpitaux et dans la communauté.

Dans les premières semaines d'une pandémie due à un virus émergent, les experts et l'ensemble des professionnels impliqués en première ligne dans la prise en charge des cas doivent composer avec de nombreuses inconnues. Dans ces situations hautement stressantes et déstabilisantes, il est important de pouvoir s'appuyer sur les principes de base de la prévention de la transmission des pathogènes, dans le cas présent les virus respiratoires, qui restent globalement les mêmes quel que soit le virus et qui sont appliqués chaque année lors de l'épidémie de grippe saisonnière : identification dès l'admission des cas suspects, mise en isolement dans un secteur dédié, application stricte de l'hygiène des mains et port d'équipements de protection individuelle par le personnel (masque +/- lunettes de protection et surblouse selon le type de soins).

Communiquer plus clairement et rapidement ce type d'information aux médias aurait pu contribuer à faire passer un message rassurant dans un contexte très anxiogène.

« L'IA pourrait éclairer de manière significative les décisions dans le cadre d'une future pandémie »

Par **Antoine Flahault**, médecin épidémiologiste, professeur honoraire à l'UNIGE et professeur à l'Université Paris Cité. Il était directeur de l'Institut de santé globale de l'Université de Genève et membre de la *task force* scientifique de la Confédération pendant la pandémie.

La crise sanitaire liée à la pandémie de Covid-19 a été sans précédent, par la mortalité qu'elle a entraînée : 7 millions de décès rapportés, mais 20 à 35 millions estimés, malgré l'ampleur de la riposte, l'arrivée de vaccins moins d'un an après son début, et la mobilisation internationale. Or, ce ne fut pas une crise seulement sanitaire, mais aussi une crise économique (avec des pertes économiques mondiales estimées à 11 000 milliards de dollars), sociale et politique.

Depuis la pandémie de Covid-19, l'intelligence artificielle (IA) a fait des progrès spectaculaires qui ne montrent aucun ralentissement. Il est donc probable qu'en cas de future crise sanitaire d'ampleur proche, les ressources produites, analysées et interprétées par l'IA seront d'un grand secours pour l'expertise sur l'interprétation des données. L'IA pourrait éclairer de manière significative les décisions à ce sujet, en scrutant les données régionales et internationales portant - au début d'un processus pandémique - sur l'émergence du virus dans de nouveaux endroits de la planète, puis sur la progression de l'épidémie, son impact et sa sévérité en effectuant des comparaisons appropriées. L'IA correctement pilotée pourrait alors proposer des mesures de riposte, discuter leurs avantages et inconvénients au regard des données probantes à leur sujet, qu'elles soient publiées dans la littérature scientifique internationale ou en cours d'expérimentation sur un territoire voisin.

Ce dont nous aurons le plus besoin à l'avenir pour gérer ce type de situations, ce seront des ingénieurs, spécialistes en IA, qui connaissent et savent faire

fonctionner les différents moteurs existants. La convergence des analyses de plusieurs IA différentes viendra conforter les décideurs dans leur prise de décision. Par la suite, il conviendra d'installer une couche d'expertise venant s'appuyer sur les résultats argumentés et sourcés produits par l'IA (ou, mieux par les IA), afin de recommander aux décideurs politiques les différentes options qui s'offrent à eux en matière de réponse, en leur présentant les avantages et les inconvénients respectifs de ces options ou scénarios. Cette dernière couche d'expertise pré-décisionnelle devra être indépendante du pouvoir politique, transparente en termes de conflits d'intérêts potentiels, plurielle en matière de disciplines académiques et professionnelles mobilisées, paritaire en termes de genre, comportant des experts adultes jeunes autant que séniors.

L'un des problèmes que rencontreront les décideurs, comme les experts, sera le risque permanent de désinformation, amplifié par l'accès aux réseaux sociaux, mais aussi de travestissement de l'information par les outils d'IA. Une qualité essentielle requise des experts (tout comme les décideurs) sera la capacité de communiquer sur leurs raisonnements interprétatifs à propos de l'évolution de la crise, des prévisions à court terme, des incertitudes à moyen et long terme, et sur les mesures préconisées. Cette communication, fondée sur les données les plus récentes de la science, devra être fluide, non entravée par le pouvoir politique, continue et présente sur tous les canaux de communication utilisés par la population.

Quelques lectures pour aller plus loin

Cairo, A. (2019). *How charts lie: Getting smarter about visual information*. W. W. Norton & Company

Un ouvrage accessible et très utile pour rappeler qu'un graphique peut éclairer autant qu'induire en erreur. Il cadre parfaitement avec l'idée de visualisation responsable et avec les enjeux de crédibilité, de cadrage et de confiance abordés dans le chapitre.

Cellule enquête Tamedia. (2020). *La première vague : Enquête au cœur de la crise du coronavirus en Suisse*. Éditions Slatkine

Un récit-enquête raconté depuis les coulisses d'un média qui revient sur la gestion des premières semaines de la pandémie. On découvre une administration publique et des experts tiraillés entre alarmisme et minimisation alors que les premiers cas et décès arrivent en Suisse.

Nussbaumer Knaflic, C. (2015). *Storytelling with data: A data visualization guide for business professionals*. Wiley

Un classique très opérationnel pour réduire le bruit visuel, guider le regard du lecteur et structurer un message autour d'un point central. Très utile pour appuyer vos passages sur l'accessibilité informationnelle et la simplicité visuelle.

Parasie, S. (2022). *Computing the News: Data Journalism and the Search for Objectivity*. Columbia University Press. <http://www.jstor.org/stable/10.7312/para19976>

La synthèse d'un important travail de recherche sur les enjeux du datajournalisme. Par une approche sociotechnique, l'auteur montre à l'aide d'exemples historiques et contemporains, comment les chiffres construisent des réalités, mettant à mal le mythe de données parfaitement objectives.

Tong, J. (Ed.). (2024). *Data journalism and the COVID-19 disruption*. Routledge. <https://doi.org/10.4324/9781003428923>

Des pratiques professionnelles à la question des publics et de leur réception, ce recueil en 15 chapitres explore les multiples enjeux du traitement médiatique de la pandémie par les chiffres.

Wilke, C. O. (2019). *Fundamentals of data visualization: A primer on making informative and compelling figures*. O'Reilly Media

Un guide très complet pour concevoir des graphiques lisibles, honnêtes et professionnels. Particulièrement utile ici pour sa réflexion sur la hiérarchie visuelle et la visualisation de l'incertitude.