

## Archives web du coronavirus (COVID-19)

**Type(s) de fonds** : archives du web

**Type de collecte** : collecte d'urgence / actualité éphémère

**Période des documents** :

01 février 2020 - 31 juillet 2020 (date de sélection du contenu), soit des premiers cas en France jusqu'à la fin de l'état d'urgence sanitaire.

**Langue(s)** : français, ponctuellement anglais

**Volume total du fonds** – Environ 15 téraoctets (dont collecte vidéos estimée à 1 To), en données compressées.

**Autres données de volumétrie** – collecte réalisée à partir d'une sélection de près de 5000 URLs et complétée par les collectes courantes pour la période considérée. L'ensemble représente 1014 jobs opérés par les robots de la BnF, 402 TLD, 274 784 516 urls collectées, 2683 types MIME, 15504 fichiers WARC.

### Description du contenu

Les archives web du coronavirus (covid-19) désignent les contenus collectés entre début février et fin juillet 2020 lors de la crise sanitaire.

La période définie couvre les différentes phases de l'épidémie en France depuis l'apparition des premiers cas jusqu'à la fin de l'état d'urgence sanitaire.

Les pages préservées rendent compte du caractère global de la crise, de sa couverture médiatique, des prises de position des différents acteurs institutionnels mais aussi de simples citoyens, et plus largement des actions mises en œuvre à travers le web pour endiguer la pandémie, la documenter et la comprendre.

Le cœur de la collection comprend deux ensembles documentaires spécifiques à l'évènement. Le premier correspond aux sélections menées par les bibliothécaires de la BnF et le réseau des correspondants régionaux. Ces sélections couvrent l'ensemble du territoire national et une grande variété de média : sites des grandes institutions, de presse quotidienne, de presse professionnelle, réseaux sociaux et sites personnels. Pour une même institution, les captures à intervalle régulier des pages dédiées au covid-19 permettent de conserver plusieurs états et rendre compte de l'évolution du site au cours du temps.

Le second ensemble documentaire a été produit à partir de la liste des noms de domaine enregistrés auprès de l'AFNIC et a permis d'archiver une part importante des nombreux sites créés spécifiquement

à l'occasion de la crise sanitaire (plate-forme d'entraide, site proposant un diagnostic en ligne, site marchand...).

En dehors de ces deux ensembles spécifiques, on pourra se reporter également aux archives web des élections municipales 2020, aux collections de presse entrées par dépôt numérique et aux collectes courantes menées régulièrement et dont l'objectif est de donner une vision représentative du web français et dont les archives permettent de suivre sites web et comptes de réseaux sociaux dans la durée.

Une collecte vidéo axée sur des chaînes présentant un fort lien avec le covid-19 a été menée en juillet 2020.

**Accès :** consultation sur place (espace recherche) et dans les établissements partenaires de la BnF sur poste dédié. Recherche par URL. Plusieurs actions de valorisation sont prévues pour faciliter l'accès au contenu, en particulier une recherche par mot et un parcours guidé.

*Possibilités de prêt : voir infra Possibilités de reproduction*

**Possibilités de reproduction :** Sur demande pour un usage privé. Pour des besoins de recherche, possibilité de fournir les fichiers de conservation au format WARC sur une infrastructure de stockage sécurisée au sein du DataLab de la BnF.

**Inventaire(s) :** la liste des URL de départ et les paramètres de collecte sont communicables et seront versés sur le site [api.bnf.fr](https://api.bnf.fr) ; le jeu de données intègre également des indicateurs pour chaque URL crawlée (première et dernière date de capture, nombre d'occurrences dans la collection). La liste des URLs qui n'ont pas pu être collectée pour des raisons techniques ou autres (pas d'autorisation du producteur, contenu disparu au moment du crawl) accompagne ce jeu de données.

**Historique et constitution :** Archives constituées entre février et juillet 2020 à partir des sélections faites par le réseau des correspondants internes BnF et le réseau des correspondants régionaux (15 établissements partenaires en région - BDLI). Le nombre de correspondants ayant directement participé à la collecte est de 52. Les captures issues des collectes courantes, en premier lieu de la presse en ligne, complètent la collection pour la période définie (février-juillet).

**Etat d'accroissement :** fonds clos au 31 juillet 2020. Les archives continuent toutefois de s'accroître pour couvrir l'évolution de la situation, notamment dans les outre-mer, et pour archiver l'actualité du web durant la période de déconfinement. A partir de septembre 2020, des sélections de sites ont pu être réactivées pour couvrir le regain d'activité de la pandémie et ses conséquences sanitaires, sociales et économiques. Cette politique de réactivation s'est accentuée à partir du second confinement.

### **Type(s) de documents**

Ventilation des sélections réalisées par les correspondants en lien avec le coronavirus entre février et juillet :

- 3260 sélections pour la catégorie « Sites et pages web », dont médias en ligne, blogs, sites hôtes

- de grande institution (type salle de presse).
- 1329 sélections pour la catégorie « Websocial »
- 249 chaînes Vidéos (Youtube)

Données chiffrées sur les fichiers produits :

- 15 504 fichiers WARC
- 402 TLD
- 2683 types MIME

**Informations complémentaires :** les archives sont conservées au format WARC. Pour des raisons techniques, la BnF ne peut garantir l'exhaustivité des archives pour l'ensemble des sites et pages collectées. Les données techniques sur la collecte peuvent être communiquées.

**Nom du responsable du fonds :** BnF - DDL - DLN - Alexandre Chautemps

**Référence bibliographique :**

Comment les archivistes de la BnF sauvegardent la mémoire du confinement sur Internet

[https://www.lemonde.fr/pixels/article/2020/05/15/a-la-bnf-les-archivistes-du-web-sauvegardent-l-inter-net-francais-du-confinement\\_6039704\\_4408996.html](https://www.lemonde.fr/pixels/article/2020/05/15/a-la-bnf-les-archivistes-du-web-sauvegardent-l-inter-net-francais-du-confinement_6039704_4408996.html)

RFI, La BnF conserve la mémoire du coronavirus sur le Web.

<http://www.rfi.fr/fr/culture/20200519-publi-mercredi-matinla-bnf-conserve-la-m%C3%A9moire-coronavirus-le-web>

Les archives web du Coronavirus : une entreprise collective <https://webcorpora.hypotheses.org/856>

The French coronavirus (COVID-19) web archive collection: focus on collaborative networks

<https://netpreserveblog.wordpress.com/2020/05/27/the-french-coronavirus-covid-19-web-archive-collection/>

**Exposition :** néant

**Thèmes principaux**

Crise du coronavirus (covid-19)

Aspects médicaux et scientifiques

Aspects sociaux

Aspects économiques

Aspects politiques

Médecine

Epidémie

Communication scientifique

Confinement

Déconfinement

**Document(s) remarquable(s) / représentatif(s)**

*#OuSontLesMasques*

<https://www.pasteur.fr/fr/espace-presse/coronavirus-toute-actualite-institut-pasteur-covid-19>

<https://dc-covid.site.ined.fr>

<https://twitter.com/inserm>

[https://twitter.com/raoult\\_didier](https://twitter.com/raoult_didier)

<http://covid-documentation.aphp.fr>

<http://academie-francaise.fr/le-covid-19-ou-la-covid-19>

<http://oreilletendue.com/2020/04/15/bref-lexique-du-confinement/>

<https://kezacovid19.wordpress.com/>

<http://ethique-pandemie.com/>