

# GUIDE DES ARCHIVES DE L'INTERNET

## Remonter le temps

Trois types de recherche sont possibles pour découvrir les collections :

### PARCOURS GUIDÉS

Les parcours guidés permettent d'aborder les collections de façon thématique. Les sujets sont choisis par des bibliothécaires ou des chercheurs partenaires et illustrés par des sélections de sites archivés.

De nouveaux parcours guidés paraissent chaque année, sous forme navigable dans l'application **Archives de l'internet** ou sous forme de bibliographie à télécharger sur le site [bnf.fr](http://bnf.fr) à l'adresse [www.bnf.fr/fr/parcours-guides-archives-de-linternet](http://www.bnf.fr/fr/parcours-guides-archives-de-linternet)

Quelques exemples de parcours guidés :

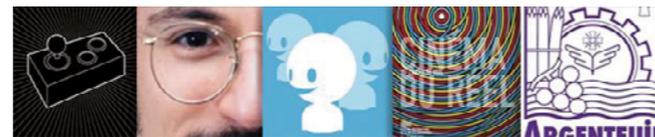
- L'intelligence artificielle sur le web



- Cliquer, voter : l'internet électoral



- Des vidéos à la chaîne : une sélection de chaînes Dailymotion archivées entre 2007 et 2013 et de chaînes YouTube archivées depuis 2017.



- Presse et actualités : une centaine de sites d'actualités (dont des versions abonnés) et une cinquantaine de journaux de presse régionale avec environ 300 éditions locales, collectés quotidiennement.



### RECHERCHE PAR ADRESSE INTERNET (URL)

La recherche par URL est le seul mode de recherche qui porte sur la totalité des archives (soit plus de 55 milliards de pages). Elle permet de retrouver l'archive d'un site, d'une page ou d'un fichier en indiquant son adresse URL. Par exemple : <http://www.bnf.fr>

Des options de recherche avancée permettent de limiter la recherche à une année ou à une collection particulière.

Utiliser la troncature (ex : <http://www.bnf.fr/>\*) permet de lister l'ensemble des pages collectées pour un site.

Un calendrier montre ensuite les dates de capture de l'URL.

Recherche par URL  
 • URL : <http://www.bnf.fr>  
 • Date : du 1 jan. 1996 au 28 août 2024  
 5 042 résultats

2024 : 184 résultats

2023 : 829 résultats

2022 : 555 résultats

jan.	fév.	mar.	avr.	mai.	juin.	juil.	août.	sep.	oct.	nov.	déc.
01	01	04	01	01	01	01	01	01	01	01	01
03	04	07	04	02	02	04	08	10	03	02	05
10	07	14	11	09	03	06	15	12	10	03	12
17	14	21	12	10	06	07	22	19	12	04	19
24	16	23	13	10	13	11	29	24	13	05	26
31	17	24	14	13	20	16	31	24	14	06	30
	18	25	15	14	27	18		25	15	07	31
	19	26	16	15	30	19		26	16	08	
	20	27	17	16	31	20		27	17	09	
	21	28	18	17		21		28	18	10	
	22	29	19	18		22		29	19	11	
	23	30	20	19		23		30	20	12	

La date, l'heure d'archivage et le permalien de chaque page consultée sont affichés dans le bandeau d'information orange en haut de page.

Page archivée le : 02 mai 2024 à 11:59 GMT | [Éléments de la page](#)  
 Permalien : <http://archivesinternet.bnf.fr/20240502115937/http://www.bnf.fr>

Le permalien permet de citer un document de façon pérenne.

### RECHERCHE PAR MOT

L'application **Archives de l'internet Labs** donne accès aux collections Incunables du web français 1996-2000, Attentats 2015, Actualités depuis 2010, Épidémie Covid-19, Web littéraire 2022 et Élections 2002-2010 grâce à la recherche plein texte.



Quatre modes de recherche par mot sont disponibles : simple, avancé, expert et ngram.

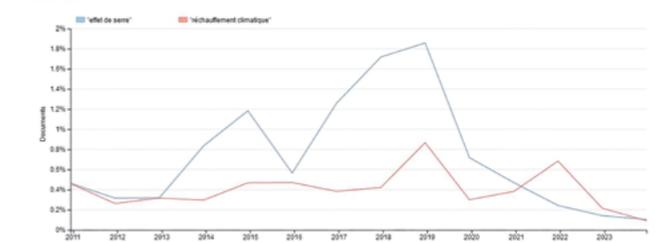
Le mode avancé permet de chercher des expressions et de préciser d'autres critères (période, format de fichier...).

Le mode expert donne la possibilité d'interroger directement les différents champs du document indexé et de construire des équations de recherche complexes.

Grâce à un système de facettes, la liste des résultats obtenue peut être affinée selon la langue, l'année de collecte, etc.

La recherche ngram permet de visualiser l'évolution de la fréquence de mots ou d'expressions à travers le temps dans les documents archivés.

5 337 124 résultats pour "effet de serre", 3 613 211 résultats pour "réchauffement climatique" sur 888 367 562 documents



## BON À SAVOIR

### COMMENT RETROUVER L'ADRESSE D'UN SITE ?

Compte tenu du nombre de sites collectés chaque année (plus de 6 millions), il n'existe pas de catalogue exhaustif des archives de l'internet. Cependant, les sites ayant fait l'objet d'une collecte ciblée sont signalés dans les fiches thématiques de [data.bnf.fr](https://data.bnf.fr). La liste de ces sites à partir de 2010 ainsi que celle des sites du web électoral depuis 2002 sont également téléchargeables sur [data.gouv.fr](https://data.gouv.fr) et [api.bnf.fr](https://api.bnf.fr).

Il est possible d'utiliser une adresse URL issue de l'internet vivant pour une recherche dans les archives.

### RETROUVE-T-ON DANS LES ARCHIVES LE MÊME CONTENU QUE DANS LE WEB VIVANT ?

Les fichiers collectés sont datés et indexés pour être restitués dans leur contexte de publication original, ce qui permet de naviguer dans les archives comme sur l'internet vivant, en cliquant de lien en lien. Certains sites ne sont pas archivés dans leur intégralité. De même, les liens d'un site vers une page qui n'aurait pas été collectée par la BnF ne seront pas actifs. Les contenus qui ont pu poser des difficultés de collecte (contenus payants ou soumis à authentification, flux audio et vidéo, animations interactives, contenus accessibles par formulaire...) ne peuvent être reconstitués à l'identique.

### EST-IL POSSIBLE DE REPRODUIRE DES ARCHIVES ?

Dans le respect du Code de la propriété intellectuelle, de courtes citations textuelles peuvent être utilisées pour illustrer une publication. Toute autre réutilisation des archives nécessite l'autorisation des ayants droit.

Un service payant d'extraction de sites web archivés est proposé par la BnF.

### COMMENT ABORDER CETTE SOURCE SI JE SUIS CHERCHEUR ?

Il existe différents dispositifs d'accueil pour les projets de recherche et une offre de services dédiés pour explorer les données et les collections.

### BESOIN D'AIDE ?

Une aide en ligne est proposée dans l'application.

Les bibliothécaires en salle peuvent également apporter tous les conseils et explications nécessaires.

## Les collections du dépôt légal de l'internet

### CADRE JURIDIQUE

Le dépôt légal, institué au <sup>xvi</sup>e siècle, a pour objectif de conserver la mémoire de toute la production éditoriale française (production scientifique, artistique, de loisir, etc.). Il s'est adapté à l'évolution des supports et permet la constitution d'une collection patrimoniale unique en son genre et irremplaçable.

Depuis le 1<sup>er</sup> août 2006, la Bibliothèque nationale de France a pour mission de collecter, conserver et communiquer les sites internet du domaine français au titre du dépôt légal (Code du patrimoine, articles L131-1 à L133-1 et R131-1 à R133-1). Cette mission s'effectue dans le respect du droit de la propriété intellectuelle et de la protection des données personnelles.

Entrent dans le champ du dépôt légal tous les signes, signaux, écrits, images, sons ou messages communiqués au public par voie électronique. Les réseaux sociaux numériques sont soumis au dépôt légal dans leur partie publique uniquement, les pages et messages privés ne sont pas concernés.

La BnF collecte des sites :

- en **.fr**
- hors **.fr** (**.com**, **.eu**, **.paris**, etc.) mais dont les producteurs sont domiciliés en France ou dont les contenus sont produits en France.

L'Institut national de l'audiovisuel (INA) est responsable de la collecte des sites web des chaînes de télévision et de radio

### CONSTITUTION DES COLLECTIONS

Blogs d'écrivains ou d'historiens d'amateurs, réseaux sociaux de candidats aux élections, sites de laboratoires de recherche, sites de presse et d'information, etc. Les collections couvrent l'ensemble des thématiques disciplinaires représentées à la BnF (ex : la littérature, l'économie) avec des focus sur des sujets spécifiques, des événements récurrents ou des objets techniques particuliers.

Les collectes sont réalisées à l'aide de robots logiciels qui copient les pages web et les éléments qui les composent (images, animations, fichiers audio et vidéo, scripts, feuilles de style, etc.). Compte tenu de la masse d'informations disponibles sur le web, tous les sites et toutes les pages des sites ne peuvent être archivés.

La BnF constitue des échantillons représentatifs de l'internet français en combinant deux modes d'entrée :

- des **collectes larges** à grande échelle qui permettent de constituer des échantillons de plusieurs millions de sites, principalement en **.fr**. Ces collectes sont réalisées une fois par an depuis 2004 ;
- des **collectes ciblées** qui permettent de constituer des archives plus complètes et plus fréquentes de plusieurs milliers de sites. Ces sites sont choisis en raison de leur thème (ex : la littérature) ou de leur rapport à un événement (ex : les élections ou les Jeux olympiques).

Les archives les plus anciennes datent de 1996.

### À NOTER

Si un auteur ou un éditeur souhaite que son site web soit ajouté à l'une des prochaines collectes, il peut envoyer une demande à l'adresse : [depot.legal.web@bnf.fr](mailto:depot.legal.web@bnf.fr) ou **via le formulaire en ligne** [www.bnf.fr/fr/signaler-un-site-web-la-bnf](https://www.bnf.fr/fr/signaler-un-site-web-la-bnf)

## Modalités d'accès

Les archives de l'internet sont consultables à la Bibliothèque nationale de France et dans des établissements partenaires en région.

Elles sont ouvertes à toute personne justifiant du besoin d'une recherche académique, professionnelle ou personnelle.

Par respect du droit de la propriété intellectuelle, **les collections ne sont pas accessibles en ligne.**

### À LA BnF

Il est nécessaire d'être titulaire d'un Pass Recherche. Ce titre donne accès aux espaces de Recherche des différents sites de la BnF : à Paris, en Rez-de-Jardin sur le site François-Mitterrand et dans les salles de lecture des sites Richelieu-Louvois, Opéra et Arsenal ; à Avignon, à la Maison Jean-Vilar.

Les archives de l'internet sont consultables soit sur les postes informatiques en accès libre et sans réservation, soit avec un ordinateur personnel connecté par le biais du dispositif **AVEC** ([avec.bnf.fr](https://avec.bnf.fr)).

Des outils d'exploration enrichie et d'aide à la fouille de texte et de données sont également proposés au sein du BnF DataLab (<https://www.bnf.fr/fr/bnf-datalab>).

### DANS LES ÉTABLISSEMENTS PARTENAIRES

Un arrêté ministériel du 16 septembre 2014 établit la liste des 26 établissements habilités à mettre en place un accès à distance aux collections du dépôt légal de l'internet. Ce service est déjà ouvert dans plusieurs établissements et s'étend progressivement. Les conditions de consultation sont disponibles auprès des établissements concernés.

### EN SAVOIR PLUS

<https://www.bnf.fr/fr/cooperer-autour-de-larchivage-du-web>

Suivez-nous sur :  @DLwebBnF | [webcorpora.hypotheses.org](https://webcorpora.hypotheses.org)



[www.bnf.fr/fr/depot-legal-du-web/](https://www.bnf.fr/fr/depot-legal-du-web/)