

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 pour des raisons d'études, professionnelles ou personnelles. 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, qui peut être obtenue auprès du Service d'accréditation. Ce titre donne accès aux espaces de Recherche des différents sites de la Bibliothèque nationale de France : à Paris, en Rez-de-Jardin sur le site François-Mitterrand et dans les salles de lecture des sites Richelieu-Louvois, Opéra et Arsenal ; en 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 (<http://avec.bnf.fr>).

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, ouvert dans certains établissements, s'étend progressivement. Les conditions de consultation sont disponibles auprès des établissements concernés.

EN SAVOIR PLUS

<http://www.bnf.fr> > Pour les professionnels > Coopération nationale > Pôles associés de dépôt légal

Suivez-nous sur  @DLWebBnF

Les collections du dépôt légal de l'internet

CADRE JURIDIQUE

Depuis le 1^{er} 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 (Facebook, Twitter...) sont soumis au dépôt légal dans leur partie publique uniquement, les pages et messages privés ne sont pas concernés.

Le dépôt légal, institué au XVI^e siècle, a pour objectif de conserver la mémoire de toute la production éditoriale française, quel que soit le public visé (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.

La BnF collecte des sites :

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

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

CONSTITUTION DES COLLECTIONS

Certaines publications numériques font l'objet de collectes spécifiques : c'est notamment le cas de la presse. Ainsi en 2016, la BnF a collecté quotidiennement 30 titres de presse et 250 éditions régionales au format PDF. La liste et les liens vers les fichiers sont disponibles dans le parcours guidé Presse et actualités.

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 l'internet, 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 (la littérature, le développement durable...) ou de leur rapport à un événement (comme les élections ou les Jeux olympiques).

En janvier 2017, les archives de l'internet de la BnF contiennent 28 milliards de fichiers, soit un volume de 793 téraoctets. Les documents les plus anciens remontent à 1996. Les archives s'accroissent progressivement au rythme des collectes.

À NOTER

Si un producteur, 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

Guide des archives de l'internet



Fonctionnalités

La page d'accueil met en valeur une sélection de sites archivés sur différents sujets (littérature, sciences, vie quotidienne, ...) et représentatifs de la diversité des collections du dépôt légal de l'internet. Cette sélection est vivante et renouvelée régulièrement.



I. PARCOURS GUIDÉS

Les parcours guidés permettent de se familiariser avec les archives et de découvrir des contenus sur des sujets choisis par des bibliothécaires ou des chercheurs partenaires. Ils se déclinent en thèmes et sous-thèmes, illustrés avec des captures des archives. De nouveaux parcours paraissent chaque année.

Parmi les parcours déjà publiés, on peut citer :

- Le web gourmand
- La ronde des festivals sur le web
- Le marché de l'art
- Images amateurs, amateurs d'images
- Le web vert : les politiques du développement durable
- (S')écrire en ligne : journaux personnels et littéraires
- Cliquer, voter : l'internet électoral

Le parcours Presse et actualité recense tous les titres et sites de presse collectés quotidiennement par la BnF, il est régulièrement mis à jour.

2. RECHERCHE PAR ADRESSE INTERNET (URL)

La recherche par URL est le seul mode de recherche qui porte sur la totalité des archives. Elle permet de retrouver l'archive d'un site, d'une page ou d'un fichier en indiquant son adresse internet (son 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.

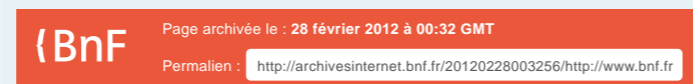
Utiliser la troncature (ex : <http://www.bnf.fr/>*) permet de lister les pages d'un site collectées.

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



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

Le permalien permet de citer un document de façon pérenne.



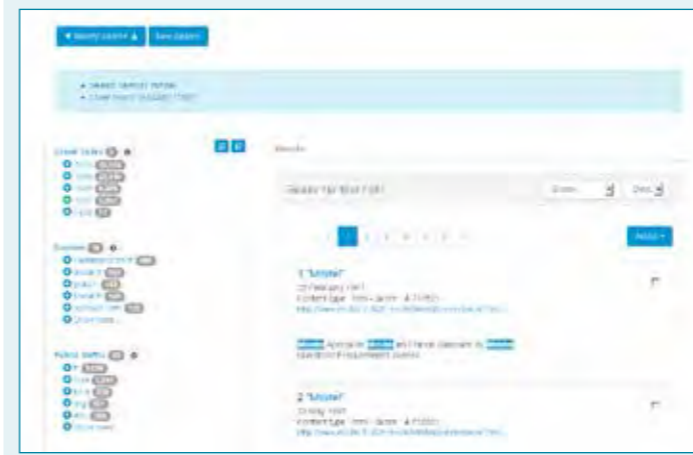
3. RECHERCHE PAR MOT

Le prototype de recherche par mot permet d'explorer la collection *Les incunables du web français 1996-2000* grâce à une indexation plein texte. Cet outil est amené à évoluer et s'enrichir de nouvelles collections.

Deux modes de recherche par mot sont disponibles, simple ou avancé. Le mode avancé permet de chercher des associations et des expressions en jouant sur la proximité des termes ou d'exclure certains termes de la recherche. Des filtres supplémentaires peuvent être appliqués pour limiter la recherche à une période, un site, etc.



Grâce à un système de facettes, la liste de résultats peut ensuite être affinée en fonction de critères tels que l'année de collecte, la langue, etc.



Astuces de navigation

NAVIGATION

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.

RETROUVER L'ADRESSE D'UN SITE

Il n'existe pas de liste complète des sites accessibles dans les archives de l'internet. Cependant, les fiches thématiques de data.bnf.fr signalent ceux qui ont été sélectionnés dans le cadre des collectes ciblées. Les listes des sites du « web électoral » et des sites officiels sont également disponibles sur data.gouv.fr. Il est possible d'utiliser une adresse URL issue de l'internet vivant pour une recherche dans les archives.

AFFICHAGE DES ARCHIVES

La mise en page d'un site archivé ne correspond pas toujours à la mise en page du site d'origine. Certains sites archivés ne présentent que la page d'accueil ou un petit nombre de pages. De même, certains liens vers d'autres sites ne sont pas actifs. C'est notamment le cas de pages ou sites qui n'ont pas été sélectionnés par la BnF mais ont pu être collectés par rebond à partir d'autres sites. Certains contenus ont pu poser des difficultés de collecte : contenus payants ou soumis à mot de passe, flux audio et vidéo, animations interactives, contenus accessibles par formulaire (moteurs de recherche, catalogues).

REPRODUCTION

Dans le respect du Code de la propriété intellectuelle, de courtes citations peuvent être réalisées pour illustrer une publication et des captures d'écran peuvent être utilisées à des fins exclusives d'illustration dans le cadre de l'enseignement et de la recherche. Il est interdit de télécharger les fichiers archivés (textes, images, vidéos, etc.). Des restrictions techniques sur les postes de la BnF ou ceux des établissements partenaires peuvent empêcher la copie de textes ou les captures d'écran.

BESOIN D'AIDE ?

Une aide en ligne est proposée dans l'application. Les bibliothécaires peuvent apporter tous les conseils et explications nécessaires.