

# Archives de l'internet de la BnF : découverte des collections et des outils



Peter Stirling

Bibliothèque nationale de France

22 novembre 2016

# **Les collections du dépôt légal de l'internet à la BnF**

# De quoi les archives sont-elles constituées ?

- Un grand principe : des captures chronologiques de sites internet
- Chaque capture montre l'état du site au moment de sa collecte

# Contenu des archives

- Définition du dépôt légal dans le Code du patrimoine
- Des sites définis juridiquement comme « français », c'est-à-dire :
  - en .fr
  - hors .fr dont les auteurs sont domiciliés en France
  - produits en France
- Mission partagée entre l'INA et la BnF

# Collectes larges et collectes ciblées

- Des documents qui entrent par deux grandes filières :
  - Collectes larges : des échantillons annuels d'un très grand nombre de sites
  - Collecte ciblées : des captures plus complètes, plus régulières d'un nombre limité de sites, sélectionnés par des bibliothécaires et faisant l'objet d'une politique documentaire concertée
  - Des acquisitions rétrospectives d'avril 1996 à 2005
- Des collections dont l'accroissement est quotidien

# Comment les archives sont-elles constituées ?

- Par l'intermédiaire d'un logiciel de collecte :  
Heritrix
  - Un robot crawler, soit le principe d'un « aspirateur » ou d'un « moissonneur » de sites
  - Il est lancé sur une liste d'adresses URL de départ dites « graines »
  - Il en extrait les liens dans le code des pages, en les suivant tel un internaute automatisé
  - Il copie les éléments qu'il trouve et qui font partie du périmètre de la collecte

# Une grande variété de documents

- Des captures de sites web, dans la continuité des différents documents de la BnF...
- ... des équivalents des publications sur d'autres types de supports...
- ... et les nouvelles formes de publications propres au medium web.

- Des sites officiels (Gouvernement, ministères...)
- Des sites de référence (CNRS, laboratoires de recherche...)
- Des sites thématiques (art, littérature, musique)
- Des sites d'actualité (presse nationale, régionale, spécialisée...)
- Des journaux en PDF (titres de presse quotidienne régionale)
- Des livres numériques en EPUB
- Contenus vidéo et audio
- Réseaux sociaux (dans leur partie publique)
- Toute la diversité du dépôt légal (jeux en ligne, sites d'annonces, catalogues...)

**Pour quels usages ? Quelles recherches ?**

- Etudier des sites disparus (par exemple des sites de campagne électorale)...
- ...ou des sites au contenu renouvelé
- Etudier le traitement d'un événement à travers des sites différents

# Citer l'archive comme source

- Utiliser le permalien
- Il permet de citer un document de façon pérenne
- [http://archivesinternet.bnf.fr/20090304210422/http://www.latribunedelart.com/Etudes\\_2005/Froment-Meurice\\_Madeleine.htm](http://archivesinternet.bnf.fr/20090304210422/http://www.latribunedelart.com/Etudes_2005/Froment-Meurice_Madeleine.htm)

# Usages académiques

- **Travaux de recherche récents :**
  - Cartographie des sites internet sur la Grande guerre et analyse d'un forum de discussion (Télécom ParisTech, dans le cadre du Labex *les Passés dans le présent*) :  
<http://dh2016.adho.org/abstracts/395>
  - Etude de ressources web sur l'émigration maghrébine en France par une chercheuse associée à la BnF :  
[https://www.cairn.info/resume.php?ID\\_ARTICLE=LCN\\_123\\_0115](https://www.cairn.info/resume.php?ID_ARTICLE=LCN_123_0115)
  - Web90, projet ISCC/CNRS avec financement ANR sur « Patrimoine, mémoires et histoire du web dans les années 1990 » <http://web90.hypotheses.org/>
  - ASAP (Archives Sauvegarde Attentats Paris), réponse à l'appel à projets CNRS sur les attentats de 2015 :  
<http://asap.hypotheses.org/>

# La consultation des collections

# Accréditation

- Conditions particulières aux collections de dépôt légal
- Dans toutes espaces Recherche de la BnF



# AVEC - Apportez Votre Equipement personnel de Communication

- Depuis sept 2014, les archives de l'internet sont consultables directement sur l'ordinateur personnel des lecteur : AVEC <http://avec.bnf.fr>

(BnF) Bibliothèque nationale de France

avec.bnf.fr met à votre disposition, sur votre ordinateur portable, un accès sécurisé aux ressources numériques, y compris sous droits, lorsque vous êtes dans les salles de lecture de la Bibliothèque de Recherche.

Bienvenue sur le portail  
**-Avec-**

L'accès aux ressources sous droits nécessite le téléchargement d'un logiciel gratuit sur votre ordinateur portable lors de la première connexion. Vous pouvez à tout moment désinstaller facilement le logiciel.

TÉLÉCHARGER LE LOGICIEL

+ D'INFOS

# Accès distant dans les bibliothèques de dépôt légal imprimeur (BDLI)

- Un arrêté du ministère de la culture du 16 septembre 2014 autorise l'accès distant dans les 26 bibliothèques régionales chargées du DL imprimeur
- Conditions d'accréditation identiques
- Déploiement progressif :
  - Montpellier, Nancy, Strasbourg, Lyon, Limoges, Toulouse, Bordeaux, Martinique, Guadeloupe, Angers, Dijon, Poitiers, Lille, Clermont-Ferrand, Besançon
- Un dispositif sécurisé pour enregistrer, puis authentifier les postes qui se connectent au SI de la BnF



# Bienvenue !

## Vous accédez à distance aux archives de l'internet de la BnF

La Bibliothèque nationale de France met à votre disposition un accès aux **collections du dépôt légal de l'internet**, lorsque vous êtes dans les salles de lecture de ses bibliothèques partenaires.

Ces « **archives de l'internet** » sont constituées de copies de sites web français effectuées à des fréquences régulières.

Ces sites sont datés et indexés pour être restitués dans leur contexte de publication original, ce qui permet de naviguer dans les archives comme sur le web.

L'accès à ces collections, qui sont sous droits, se fait via un **navigateur sécurisé**. Vous pourrez consulter l'aide disponible dans l'application.

Pour accéder aux collections, il est nécessaire de lire attentivement et d'accepter la charte du bon usage des services à distance de la BnF.

### CHARTE DU BON USAGE DES SERVICES À DISTANCE DE LA BNF

#### Conditions d'accès et d'utilisation

**Nature des utilisations :** l'utilisation des services à distance de la BnF dans les salles de lecture des bibliothèques partenaires de la BnF est réservée à des fins de recherche documentaire et d'information conformes aux missions de recherche et d'étude de la BnF (consultation des collections du dépôt légal de l'internet).

**Respect des autres usagers :** Chaque utilisateur est invité à respecter les autres lecteurs en particulier en observant la durée des sessions sur les postes informatiques publics et en évitant

J'ai lu la charte et je l'accepte

→ **Consulter**

# Exemple de navigation dans les archives

The screenshot shows the homepage of the BnF Archives de l'internet. At the top, the logo '(BnF) Archives de l'internet' is on the left, and 'A propos | Aide' is on the right. Below the logo is a dark blue header with the text 'Remonter le temps' and 'Trouver l'archive d'un site, une page, un fichier en indiquant son adresse URL'. A search bar contains 'http://' and a red button labeled 'Rechercher l'URL'. Below the search bar is a link to '» Recherche avancée'. A navigation bar features a central text 'Découvrir les collections à travers des sélections thématiques' and a red button 'Tous les parcours guidés'. Below this are icons for 'PRESSED ET ACTUALITÉ', 'LA RONDE DES FEST...', 'LE MARCHÉ DE L'ART', 'JEAN-PHILIPPE RAM...', 'LE WEB SCIENTIFIQUE', and 'CARNETS DE VOYAGE'. The main content area displays a grid of archive cards. The first card shows a technical drawing with the text 'Tous ses états successifs de la page d'accueil du Désordre, racontés par l'auteur responsable du désordre, Philippe De Jonckheere.' and 'Archive du 04 décembre 2008 à 20:13'. The second card shows a close-up of a hairy insect with the text 'Une plongée à la découverte des profondeurs océaniques' and 'Archive du 02 juin 2010 à 02:59'. The third card shows two birds with the text 'Découvrez la très rare cigogne noire !' and 'Archive du 01 juin 2010 à 07:59'. The fourth card shows a person holding a phone with the text 'FRED FOREST MÉDIAS EN PARTAGE' and 'Archive du 14 octobre 2015 à 00:16'. Other cards include 'ARTHUR C. CLARKE est mort', a grid of small images, a close-up of a bee, and a cartoon with the text 'TU MOURRAS MOINS BÊTE' and 'MAIS TU MOURRAS COMME MÊME!'.

Page d'accueil

http://www.bnf.fr

Rechercher l'URL

» Recherche avancée

Parcours guidés

» Presse et actualité

## Recherche par URL

► URL : http://www.bnf.fr

► Date : du 1 jan. 1996 au 3 oct. 2016

1 500 résultats

» Modifier la recherche

Tout ouvrir | Tout fermer

2016

34 résultats ▼

jan.	fév.	mar.	avr.	mai	juin	jui.	août	sep.	oct.	nov.	déc.
02	02 29	+ 01 + 14	01 02 + 18 + 19 20 22 24	02	02 16	02 07 08	02 + 03 04 29	07			

2015

159 résultats ▶

2014

341 résultats ▶

2013

252 résultats ▶

2012

215 résultats ▶

2011

138 résultats ▶

Résultats de la recherche avec l'adresse *www.bnf.fr*

Joel-Peter Witkin. *Enfer ou Ciel*  
jusqu'au 1<sup>er</sup> juillet 2012  
Richelieu



Autour de l'exposition *Enfer ou Ciel* – Master class par Joel-Peter Witkin  
2 avril 2012 – François-Mitterrand

Expositions | Auditoriums | Expositions virtuelles

Le trésor du mois : un manuscrit de Marcel Pagnol  
François-Mitterrand - Espace découverte -  
jusqu'au 22 avril 2012



► Poser une question à un bibliothécaire

Gallica | Catalogue général | Collections

L'appel à chercheurs 2012-2013 est ouvert  
jusqu'au 30 mars 2012

► Conditions d'accès et tarifs | ► Réserve de place

Actualités | BnF pratique | Consulter à Richelieu



Soutenir la BnF

Événements et culture

- Auditoriums
- Conférences en ligne
- Le Cercle Littéraire de la BnF
- Expositions
- Éditions de la BnF
- Expositions virtuelles
- Action pédagogique
- Bibliothèque numérique des enfants
- Nouveautés Éditeurs
- Le magazine *Chroniques*

Collections et services

- Bibliothèques numériques - Gallica
- Catalogues
- Collections et départements
- Aides à la recherche documentaire
- Périodiques et bases en ligne
- Banque d'images
- Reproduction des documents
- Services aux lecteurs
- Les Signets de la BnF

La BnF

- Actualités
- BnF pratique
- Connaître la BnF
- Bibliothèque du Haut-de-jardin
- Visites et ateliers
- Programmes de recherche
- Rénovation de Richelieu
- Le Labo
- Vidéos

Pour les professionnels

- Actualités - Agenda
- Dépôt légal
- ISSN, ISBN, autres numéros
- Bibliographie nationale française
- Catalogue et indexation
- Produits et services bibliographiques
- Coopération nationale
- Formations et journées professionnelles
- Action internationale
- Numérisation et archivage numérique
- Conservation

Accès dédiés

- Acteurs sociaux
- Appels d'offres
- Éditeurs
- Enseignants et jeunesse
- Location d'espaces et tournages
- Mécénat
- Offres d'emploi
- Presse
- Publics handicapés

Capture de la page d'accueil du site [www.bnf.fr](http://www.bnf.fr) du 1<sup>er</sup> avril 2012