

Le Web comme source ou terrain en SHS : cadre théorique et méthodologique de l'archivage du web et de son analyse

Séminaire WebLab – Humathèque Condorcet “Le Web et les archives du Web pour la recherche en SHS”,
Médiathèque de la MMSH, Aix-en-Provence (hybride), jeudi 9 octobre 2025, 14h-16h

Maya Anderson-Gonzalez (Humathèque Condorcet)

Sophie Gebeil (TELEMMe, WebLab)

Jean-Christophe Peyssard (MMSH, WebLab)



**institut
universitaire
de France**



**CAMPUS
CONDORCET
HUMATHÈQUE**



**Maison méditerranéenne
des sciences humaines et sociales**
UAR 3125

amU Aix
Marseille
Université



Présentation des organisateurices

- Sophie GEBEIL, Maître de conférences en Histoire (TELEMME, AMU-CNRS), membre de l'IUF, VP déléguée Sciences humaines et Méditerranée pour AMU
- Maya ANDERSON-GONZALEZ, ingénieure d'étude à l'Humathèque Condorcet, chargée d'accompagnement pour les données du web au sein du service Accompagnement de projet et science ouverte (SAPSO)
- Jean-Christophe PEYSSARD, responsable de la Médiathèque de la MMSH

Présentation du WebLab

PÔLE BIBLIOTHÈQUES ET ARCHIVES DE LA MMSH

Bibliothèques et Archives à la Maison méditerranéenne des sciences de l'homme

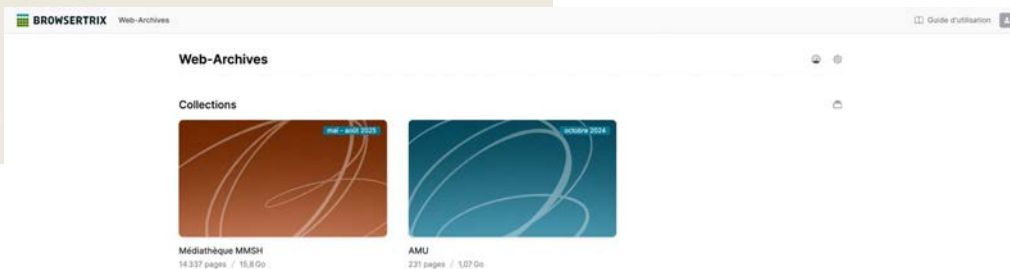
Accueil WebLab + MédiHAL Les carnets de la MMSH Frantiq Veille À propos Contact

WebLab

Le WebLab est un lieu de réflexion et de pratique pour l'étude du web archivé et des nouveaux médias impulsé par Sophie Gebeil en tant que lauréate d'une Chaire Médiation Scientifique à l'Institut Universitaire de France.

Objectifs du WebLab

Le WebLab propose de développer et de partager :



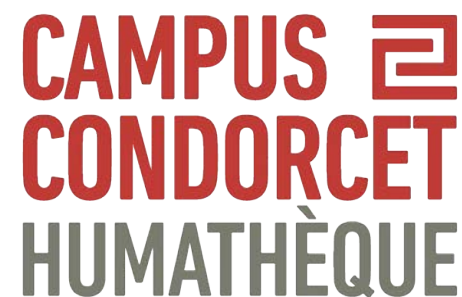
https://pba.mmsh.fr/?page_id=1465

<https://archives-poc.cedre.univ-amu.fr/explore/web-archives>

Présentation de l'Humathèque Condorcet

Le Campus Condorcet réunit onze établissements d'enseignement supérieur et de recherche.

- [Le Centre national de la recherche scientifique](#)
- [L'École des hautes études en sciences sociales](#)
- [L'École nationale des chartes](#)
- [L'École pratique des hautes études](#)
- [La Fondation maison des sciences de l'homme](#)
- [L'Institut national d'études démographiques](#)
- [L'Université Paris 1 Panthéon-Sorbonne](#)
- [L'Université Sorbonne Nouvelle](#)
- [L'Université Paris 8 Vincennes-Saint-Denis](#)
- [L'Université Paris Nanterre](#)
- [L'Université Sorbonne Paris Nord](#)



L'[Humathèque Condorcet](#) du Campus Condorcet est le résultat d'une mutualisation des collections de 50 bibliothèques issues de 8 établissements.

Offre de service de l'Humathèque Condorcet

[Actualités de l'Humathèque](#) [Agenda de l'Humathèque](#)

**CAMPUS
CONDORCET
HUMATHÈQUE**  **COLLECTIONS ET ARCHIVES**

[Accueil](#) > [Pour la recherche](#)

POUR LA RECHERCHE

19 AVRIL 2022, MODIFIÉ LE 22 NOVEMBRE 2022



SOUTIEN À LA RECHERCHE

L'Humathèque accompagne les la recherche.

CONSTITUER DES CORPUS DE DONNÉES

L'Humathèque peut s'engager de manière **opérationnelle** en intervenant dans l'ensemble des étapes du cycle de vie des données.

Accompagnement méthodologique :
définition des corpus, rédaction de guides de collecte des métadonnées adaptés aux corpus et entrepôts, etc.

Accompagnement technique :
numérisation, aide à la saisie de métadonnées, dépôt de données et métadonnées dans des entrepôts à la pièce ou en masse via les APIs, etc.

<https://www.humatheque-condorcet.fr/fr/pour-la-recherche/science-ouverte-et-donnees-de-la-recherche>

Contexte Humathèque : étude et réalisation

CAMPUS 
CONDORCET
HUMATHÈQUE

Exploiter des
données du web
à l'Humathèque
Condorcet : de
l'opportunité à la
faisabilité

Rapport d'étude

Maya ANDERSON-GONZALEZ
10/07/2025

Mission d'étude (2024-2025)

Panorama des données du web

mobilisées au sein des institutions
résidentes et membres du Campus
Condorcet

Mission de réalisation (2026-2027)

Prototyper un service

d'exploitation des données du web à
l'Humathèque, suite à l'étude d'
évaluation d'opportunité et de
faisabilité

La collaboration HTQ <> WebLab

- Les problématiques de la recherche et de la conservation des données issues du web sont véritablement deux maillons d'une même chaîne de traitement de données.

- Les bibliothèques jouent un rôle central dans l'accompagnement des nouvelles pratiques de recherche fondées sur le web et ses données.

Le WebLab et l'Humathèque associent leur expertise pour s'engager dans la transformation des pratiques de recherche via :

- la création de connaissances ;
- la formation sur le sujet ;
- la sensibilisation ;
- le partage d'outils communs.

Ensemble, nous souhaitons également renforcer la place de nos communautés dans les réseaux internationaux liés aux archives du web ([RESAW](#), [IIPC](#)).

La collaboration HTQ <> WebLab

Dans ce contexte, le WebLab et l'Humathèque Condorcet s'engagent à :

1. **Créer un écosystème (technique) commun** afin de susciter de nouveaux projets de coopération.
2. **Mener un travail collectif** sur la méthodologie d'accompagnement, de stabilisation et de pérennisation des corpus de données du web éventuellement constitués.
3. **Renforcer les liens entre les structures** existantes ayant une expertise autour des données du web et de leur analyse.
4. **Réfléchir ensemble à la constitution de corpus** de données du Web (vivant) à des fins d'expérimentation et d'analyse.

Le Web comme source ou terrain en SHS : cadre théorique et méthodologique de l'archivage du web et de son analyse

Partie 1 : Institutions SHS et données du Web

Partie 2 : Le Web et ses archives

Partie 3 : Une communauté internationale structurée

Partie 4 : Défis pour les SHS

Partie 5 : Programme prévisionnel du séminaire du WebLab

Partie 1 :

Institutions SHS et données du Web

Un réseau de partenaires à remobiliser

Participation de l'Humathèque Condorcet au ResPaDON
(2021-2023)



RÉSEAU DE PARTENAIRES POUR L'ANALYSE ET L'EXPLORATION DE DONNÉES NUMÉRIQUES, *Améliorer l'exploitation des archives du web par les communautés de recherche : Les 15 préconisations du projet ResPaDon*, <https://respadon.hypotheses.org/2768>, 18 décembre 2023.

Des données du web “vivant” ?

« Par ‘données issues du Web’, nous entendons les **données construites à partir des traces des activités sociales en ligne** des internautes, que celles-ci soient fournies par les sites web qui organisent ces activités, ou qu’elles soient construites par le chercheur à partir des informations visibles sur le Web. »

BEUSCART Jean-Samuel, « Des données du Web pour faire de la sociologie... du Web ? » in *Big data et traçabilité numérique : Les sciences sociales face à la quantification massive des individus*, Pierre-Michel Menger et Simon Paye (éd.), Paris, Collège de France, « Conférences », 2017, p. 141-161.

Etat des lieux

Une tendance :

des pratiques et méthodes propres à chaque communauté disciplinaire en SHS + des équipes et services d'accompagnement en quête de structuration et de standardisation à travers le territoire.

Un objectif :

insérer les données du web dans une chaîne de traitement des données de la recherche selon un processus structuré et une logique itérative.

Des perspectives :

des acteurs institutionnels souhaitant se constituer en communautés de pratique articulant différents métiers et profils complémentaires de l'ESR.

Compétences métiers multiples et complémentaires

- ★ Expertise thématique et technique
- ★ Ingénierie documentaire et de recherche
- ★ Conception archivistique, patrimoniale & design

L'outillage : point de jonction méthodologique ?

- ***Web scraping***

Rétroingénierie de pages web : identifier le code HTML qui nous intéresse et le « racler » pour obtenir des données brutes non-structurées.

- ***Web crawling***

Indexation automatisée de contenus trouvable sur le web (robot).

- ***Web mining***

Analyse approfondie de jeux de données massives.



<https://hyphe.medialab.sciences-po.fr/>

Panorama des institutions

Densification et diversification du paysage des acteurs de l'exploitation des données du web pour la recherche en SHS. Apparition d'infrastructures et de services dédiés au sein de structures documentaires :

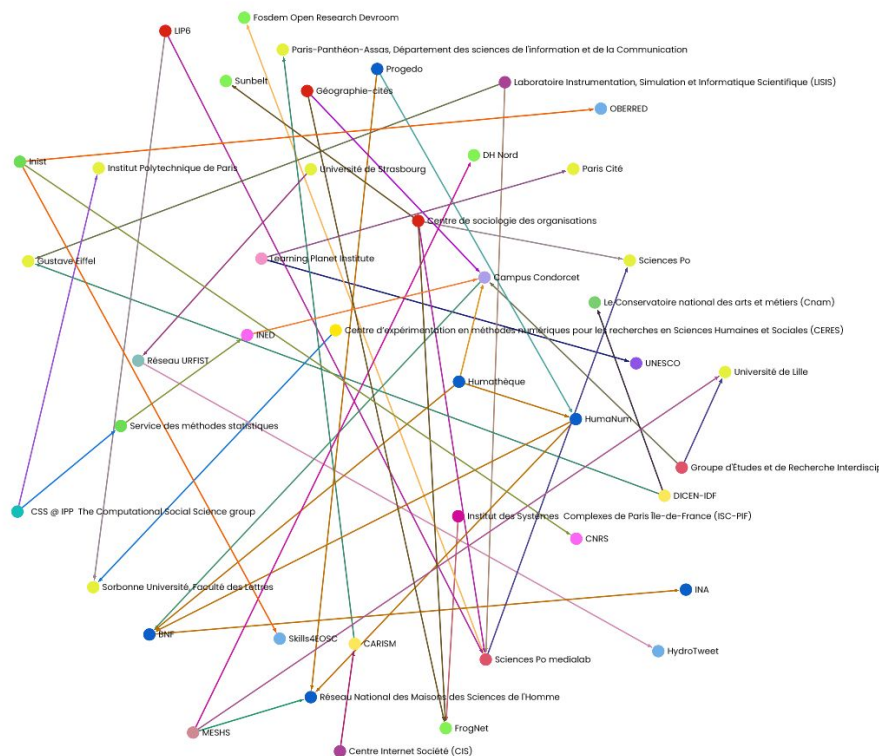
le medialab de Sciences Po ;

le BnF DataLab ;

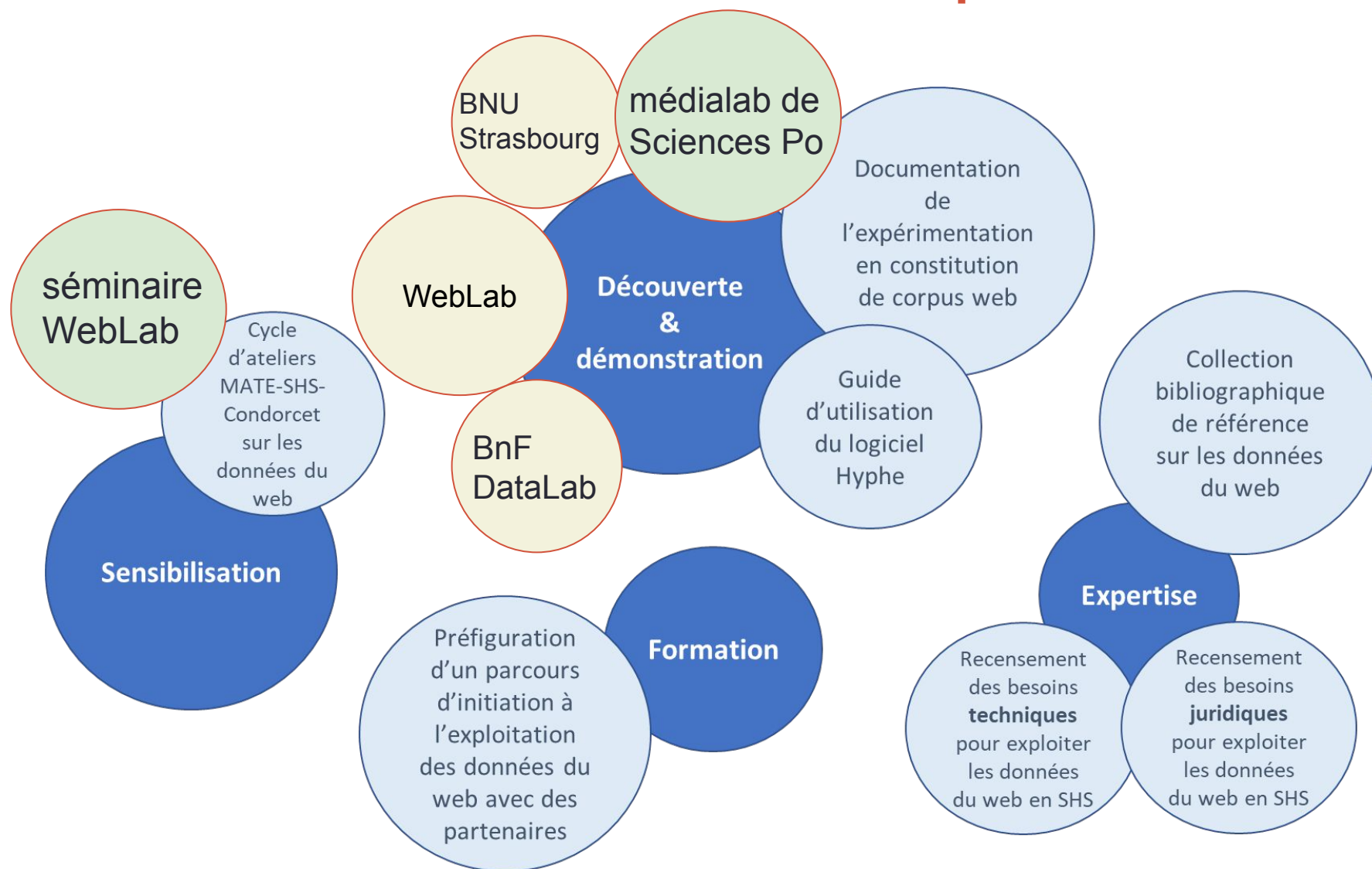
le Lab de l'INA ;

le Datalab de la BNU de Strasbourg ;

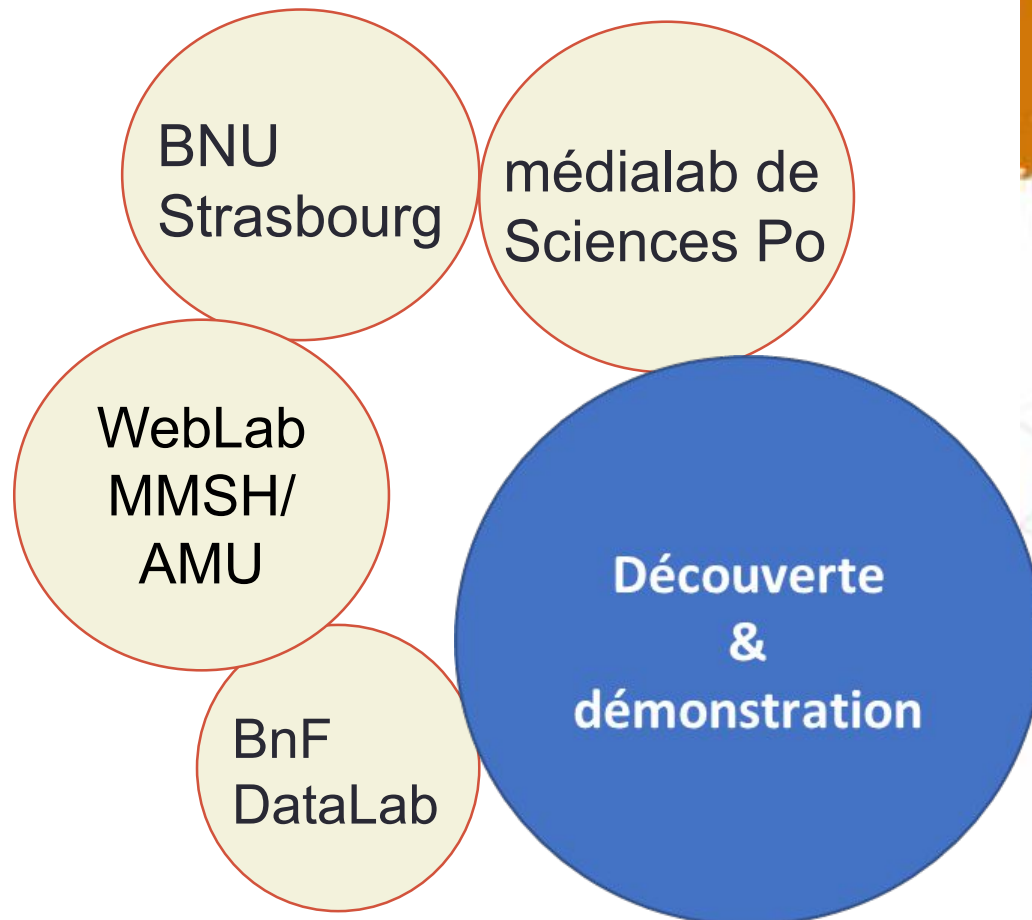
le WebLab AMU/MMSH.



Des actions à mener en partenariat

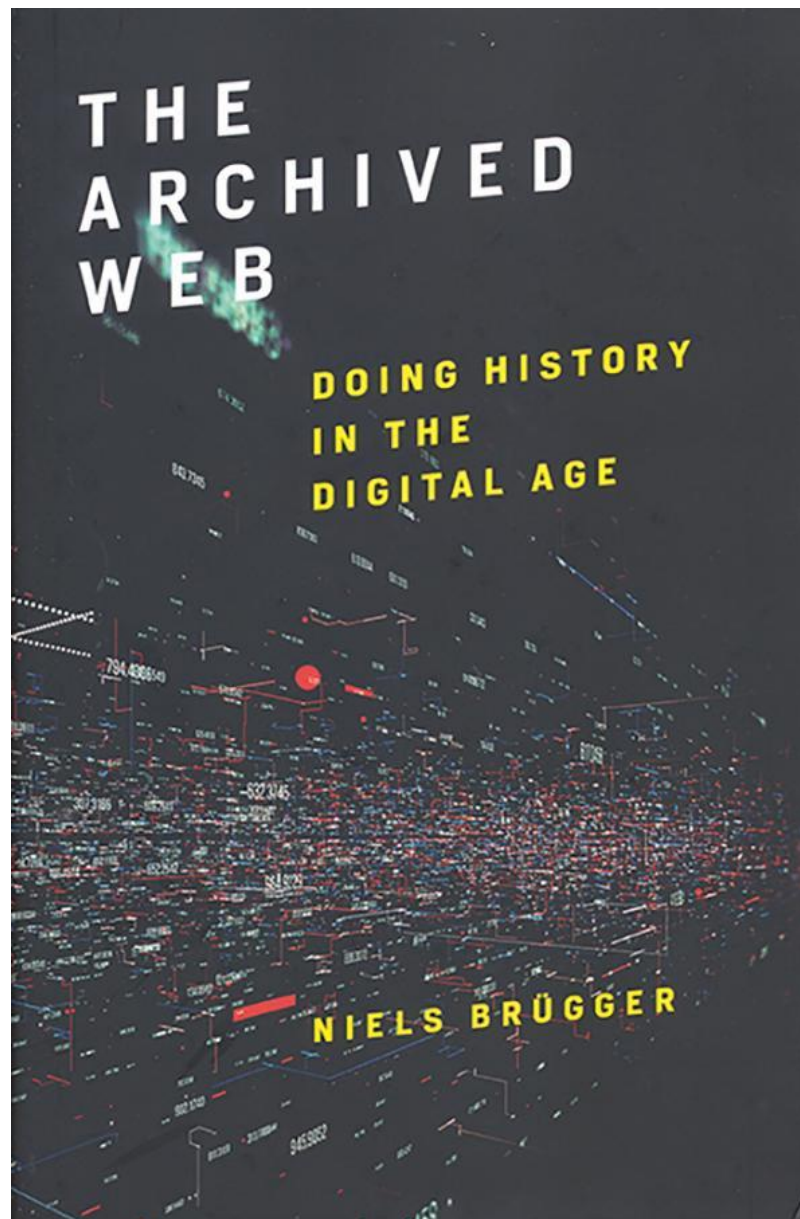


Prochain événement à l'Humathèque : 18-20 novembre 2025



Partie 2 :

Le Web et ses archives



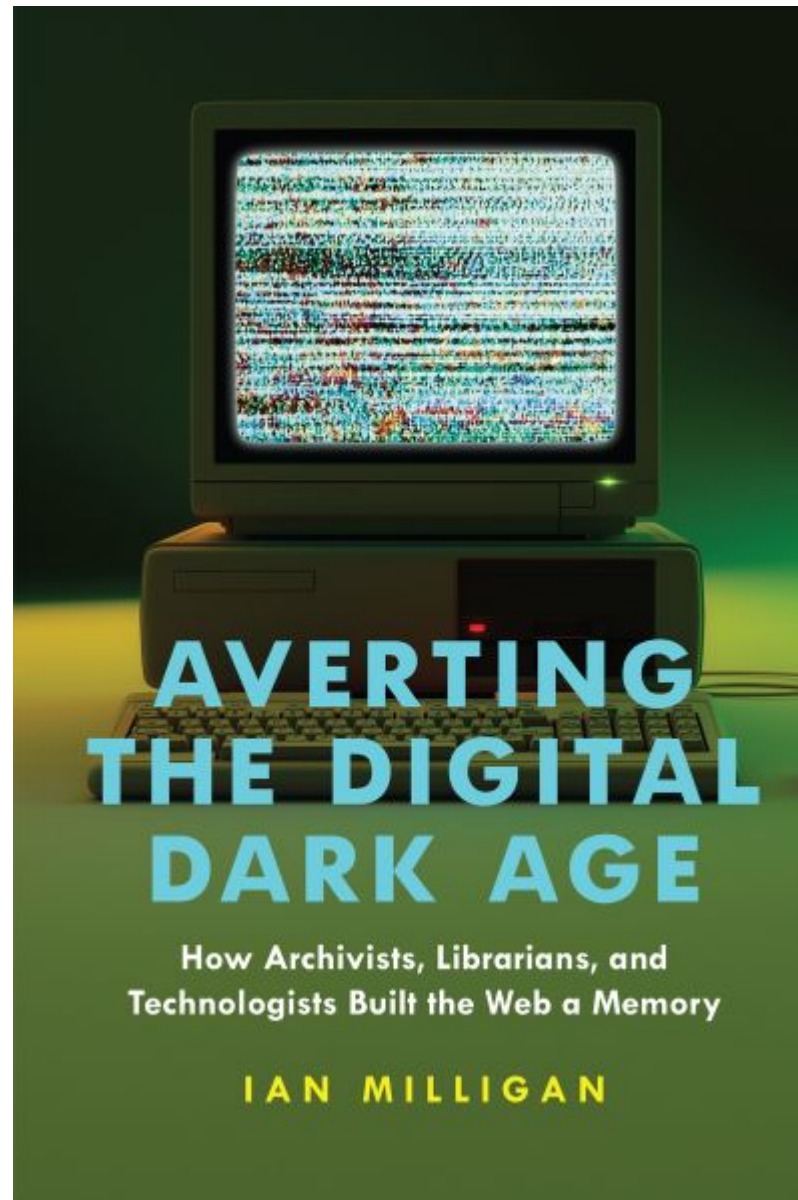
Brügger, Niels. 2018. *The Archived Web: Doing History in the Digital Age*. Cambridge, Massachusetts: The MIT Press. <https://search.worldcat.org/fr/search?q=bn:9780262039024>.



Musiani, Francesca, Camille Paloque-Bergès, Valérie Schafer, and Benjamin G. Thierry. 2019. *Qu'est-ce qu'une archive du web ?* Encyclopédie numérique. Marseille: OpenEdition Press.
<https://doi.org/10.4000/books.oep.8713>.



Gebeil, Sophie. 2021. *Website story: histoire, mémoires et archives du web*. Bry-sur-Marne, France: INA.
<https://www.cairn.info/website-story--9782869382824.htm>.



Les médias numériques

- Matériaux numérisés / Digitized material

Matériaux analogiques qui ont été numérisés : manuscrits numérisés, photographies analogiques, médias électroniques (émission de radio et télévision) ;

- Matériaux nativement numériques / Born-digital material

Qui n'a jamais existé autrement que sous une forme numérique : un texte numérique, une image numérique, un son numérique, CD Rom, DVD, sites Web ;

- Matériaux numériques « régénérés » / reborn-digital material

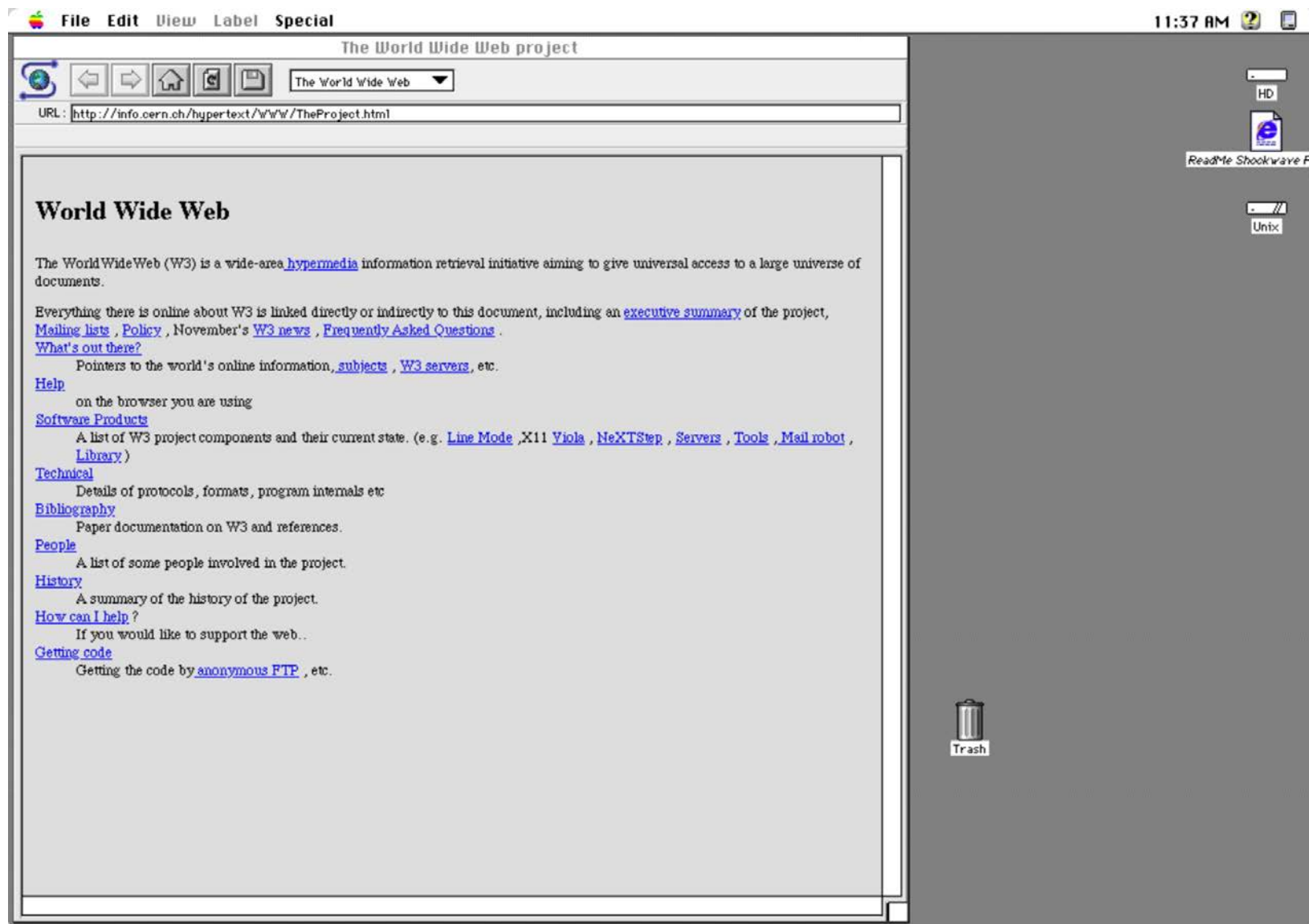
Matériaux nativement numériques collectés et conservés : jeux vidéos émulés, archives du Web : reborn digital heritage.

Brügger, Niels. 2016. « Digital Humanities in the 21st Century: Digital Material as a Driving Force ». *Digital Humanities Quarterly*, 10 (3).

<http://www.digitalhumanities.org/dhq/vol/10/3/000256/000256.html>.

World Wide Web

Le premier site Web a été publié le 6 août 1991 par le physicien britannique Tim Berners-Lee au CERN en Suisse



<http://info.cern.ch/hypertext/WWW/TheProject.html>

La structure d'une URL sur le World Wide Web (www) :

protocol:://**subdomain**.**domain**.**top-level domain**/**path**/**page**/

Exemple :

https:://**distam**.hypotheses.org/**category**/**ecole-dete**

<https://en.wikipedia.org/wiki/URL>

<http://dac.au.dk/forskning/forskningsprogrammer> p. 51

En janvier 2025, le Web indexé par les moteurs de recherche contiendrait 3,98 milliards de pages

<https://www.worldwidewebsize.com>

En août 2025, Netcraft recense 1 300 016 299 sites répartis sur 281 522 947 domaines et 13 797 308 serveurs (« web-facing computers »)

<https://www.netcraft.com/blog/august-2025-web-server-survey>

- La durée de vie moyenne d'une page web est de 44 jours ; 44% des sites web recensés en 1998, n'étaient plus trouvables en 1999 (Lyman, 2002 p. 38) ;
- 40% des contenus du web disparaît au cours d'une année, 40% sont modifiés, voilà pourquoi aujourd'hui, on peut seulement s'attendre à trouver 20% des contenus qui étaient disponibles il y a un an (Brügger, 2005, p. 15) ;
- En moyenne, une page web subira une modification ou disparaîtra, avant 100 jours (Kahle, 2015) ;
- En 2019, selon l'équipe de la Wayback Machine, la durée de vie moyenne d'une page web est de 92 jours ;

38% of webpages from 2013 are no longer accessible

% of links from each year that are no longer accessible as of October 2023



Source: Pew Research Center analysis of a random selection of URLs collected by the Common Crawl web repository (n=999,989) and checked using page and DNS response codes. Web pages defined as inaccessible if they returned a status code of 204, 400, 404, 410, 500, 501, 502, 503, 523 or did not return a valid status code.

“When Online Content Disappears”

PEW RESEARCH CENTER

The New York Times

In Supreme Court Opinions, Web Links to Nowhere

By Adam Liptak

Sept. 23, 2013



WASHINGTON — Supreme Court opinions have come down with a bad case of link rot. According to [a new study](#), 49 percent of the hyperlinks in Supreme Court decisions no longer work.

This can sometimes be amusing. A link in [a 2011 Supreme Court opinion](#) about violent video games by Justice Samuel A. Alito Jr. now leads to [a mischievous error message](#).

“Aren’t you glad you didn’t cite to this Web page?” it asks. “If you had, like Justice Alito did, the original content would have long since disappeared and someone else might have come along and purchased the domain in order to make a comment about the transience of linked information in the Internet age.”

Liptak, Adam. 2013. “In Supreme Court Opinions, Web Links to Nowhere.” *The New York Times*, September 13, 2013, sec. Politics.
[https://web.archive.org/web/20200218163337/https://www.nytimes.com/2013/09/24/us/politics/in-supreme-court-opinions-clicks-t](https://web.archive.org/web/20200218163337/https://www.nytimes.com/2013/09/24/us/politics/in-supreme-court-opinions-clicks-that-lead-nowhere.html?hp&r=3&#)
[hat-lead-nowhere.html?hp&r=3&#](https://web.archive.org/web/20200218163337/https://www.nytimes.com/2013/09/24/us/politics/in-supreme-court-opinions-clicks-t)

Support The Guardian
Available for everyone, funded by readers

Search jobs Sign in Search

The Guardian
International edition

The essential news magazine from The Guardian.
Subscribe now

Try 6 issues for \$6

Conservatives

Conservative party deletes archive of speeches from internet

Decade's worth of records is erased, including PM's speech praising internet for making more information available

Randeep Ramesh and Alex Hern
Wed 13 Nov 2013 15:40 GMT
5540 1284

▲ A speech in which David Cameron said the internet would help people hold politicians to account was among those deleted. Photograph: Barcroft Media

The **Conservatives** have removed a decade of speeches from their website and from the main internet library - including one in which David Cameron claimed that being able to search the web would democratise politics by making "more information available to more people".

Editorially

SAMSUNG

Buy any Galaxy S10 model

Galaxy S10

To get a FREE Galaxy Tab E

Buy Now

Ramesh, Randeep, and Alex Hern. 2013. “Conservative Party Deletes Archive of Speeches from Internet.” *The Guardian*, November 13, 2013, sec. Politics.
<https://web.archive.org/web/20131113221027/https://www.theguardian.com/politics/2013/nov/13/conservative-party-archive-speeches-in-ternet>.

Exemple de disparition

Paramount Global Erases Archives of MTV Website, Wipes Music, Culture History After 30 Plus Years

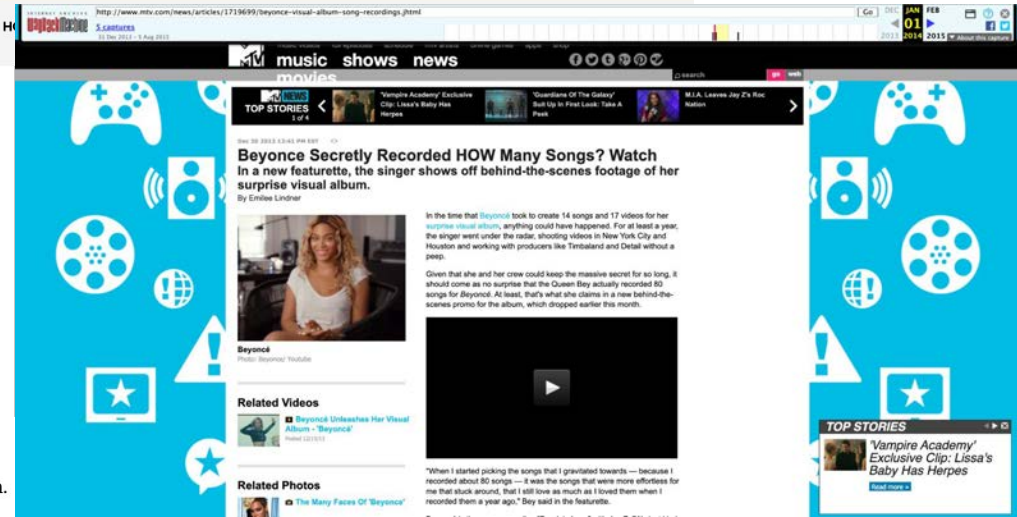
June 25, 2024 12:36 am | By Roger Friedman

Share



MTV.com is gone. Kaput. Wiped off the face of the Earth.

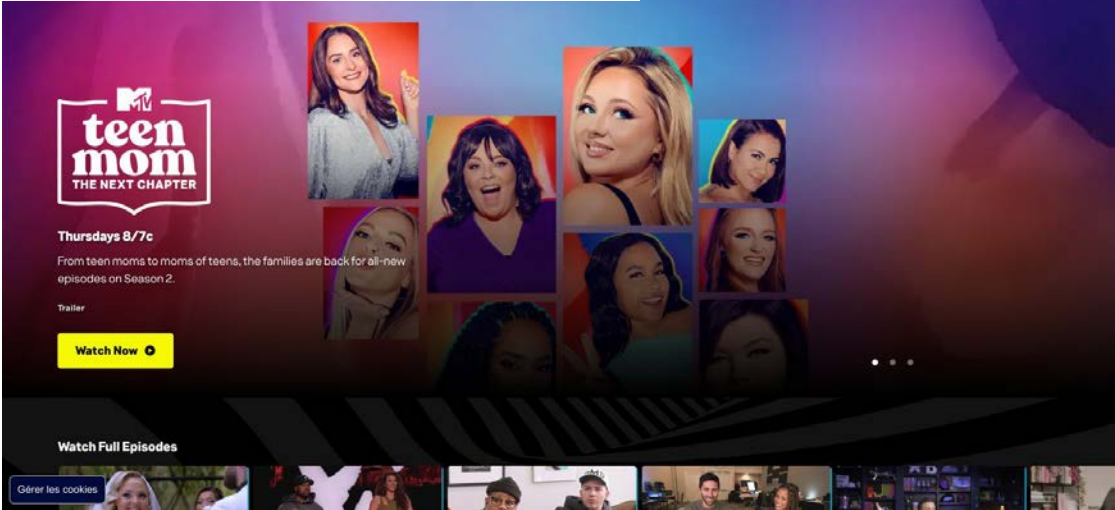
Parent company Paramount Global, formerly Viacom, has tossed twenty plus years of news archives. All that's left is a placeholder site for reality shows. The M in MTV – music – is gone, and so is all the reporting and all the journalism performed by music and political writers ever written. It's as if MTV never existed. (It's the same for VH1.com, all gone.)



Read more

Celebrity

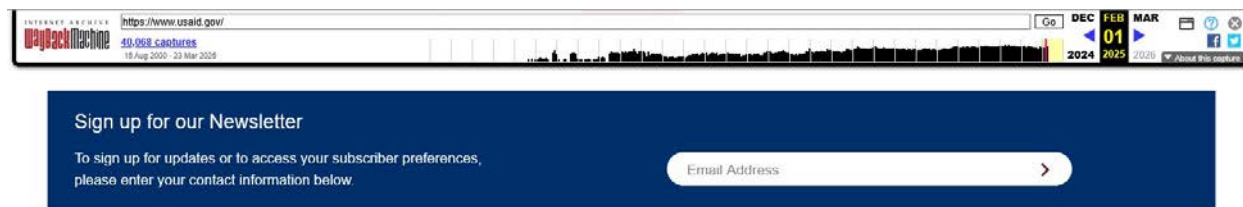
70s Feud: Singer Songwriter Stephen Bishop Goes "On and On" About "Twilight Zone" Director John Landis, Says "Nothing But Mean Spirited"



Friedman, Roger. 2024. "Paramount Erases Archives of MTV Website, Wipes Music, Culture History After 30 Plus Years." *Showbiz411* (blog). June 25, 2024.
<https://www.showbiz411.com/2024/06/25/paramount-shuts-down-mtv-website-wipes-history-after-20-plus-years>.

Exemple de disparition

USAID



Featured Focus Areas



<https://web.archive.org/web/20250201003843/https://www.usaid.gov/>



Notification of Administrative Leave

As of 11:59 p.m. EST on Sunday, February 23, 2025, all USAID direct hire personnel, with the exception of designated personnel responsible for mission-critical functions, core leadership and/or specially designated programs, will be placed on administrative leave globally.

Concurrently, USAID is beginning to implement a Reduction in Force that will affect approximately 1,600 USAID personnel with duty stations in the United States.

Individuals that are impacted will receive specific notifications on February 23, 2025, with further instructions and information about their benefits and rights.

Designated essential personnel who are expected to continue working will be informed by Agency leadership February 23, 2025, by 5 p.m. EST.

For overseas personnel, USAID intends a voluntary Agency-funded return travel program and other benefits. USAID is committed to keeping its overseas personnel safe. Until they return home, personnel will retain access to Agency systems and to diplomatic and other resources.

In the coming week, we will provide details on how to retrieve personal items from the former USAID workspaces and return government issued devices.

Additional guidance is forthcoming, and all future updates/updates will continue to be communicated through official USAID channels and posted on USAID.gov for those without access to USAID systems.

USAID Office of Inspector General

<https://www.usaid.gov/>

Un problème pour la recherche et la communication scientifique

- Error 404, Broken links, Link rot, Reference rot, Infosuicide, digital ruins, content drift, zombie media,.. *Erreur 404, lien cassé, (...), suicide numérique, (...), site non maintenu* ;
- Fermetures volontaires et administratives (« take down »), fusions et acquisitions :

Le 18 mars 2019, on apprend que *MySpace* a perdu les contenus de ses utilisateurs au cours d'un incident de migration de serveurs qui s'est mal déroulé. Plus de 50 millions de chansons et 12 années de production de contenus ont disparu pour toujours. Il n'y avait pas de sauvegarde (<https://en.wikipedia.org/wiki/Myspace>).
- Histoire :

Séparation de la Yougoslavie (.yu - Serbie and Montenegro, .rs & .me) et dissolution de la Tchécoslovaquie (.cs – maintenant République Tchèque et Slovaquie, .cz & .sk).
- **Reference rot**, a combination of:
 - **Content decay**: The content of the linked resource may change over time and, as a result, the degree to which that content remains representative of the content that was intended to be linked to may decrease over time.
- **Link rot**: The linked resource may disappear altogether. (Thoughts on Referencing, Linking, Reference Rot <http://mementoweb.org/missing-link/>)
- Neal, James G. 2014. “**The Integrity of Research Is at Risk**: Capturing and Preserving Web Sites and Web Documents and the Implications for Resource Sharing.” In . Lyon, France. <http://library.ifla.org/id/eprint/907>.)

Pourquoi archiver le Web ?

- Conserver notre patrimoine numérique culturel et scientifique ;
- Stabiliser et conserver les contenus du web en tant qu'objet de recherche ;
- Administrer la preuve et citer les sources.

- Web archiving is the process of collecting portions of the World Wide Web, preserving the collections in an archival format, and then serving the archives for access and use. (IIPC Web site, <http://netpreserve.org/web-archiving/>)
- “Web archiving is the process of gathering up data that has been recorded on the World Wide Web, storing it, ensuring the data is preserved in an archive, and making the collected data available for future research.” (Niu, Jinfang. 2012. “An Overview of Web Archiving.” *D-Lib Magazine* 18 (3/4). <https://doi.org/10.1045/march2012-niu1>.)

Brève chronologie de l'archivage du web

- 1537** Création du dépôt légal en France (Espagne, 1619 ; Royaume-Uni, 1710)
- 1989** Le World Wide Web est inventé par Tim Berners-Lee
- 1996** Fondation d'Internet Archive par Brewster Kahle
- 1996** Début de l'archivage du web en Suède (domaine national .se) avec le projet Kulturarw3
- 1998** Lancement de Google
- 2001** Création de la Wayback Machine (Internet Archive)
- 2003** Publication de la charte de l'UNESCO sur la conservation du patrimoine culturel numérique
- 2003** Création du IIPC, International Internet Preservation Consortium à la bibliothèque nationale de France, avec 12 institutions partenaires
- 2005** Lancement de Youtube
- 2006** Lancement de Facebook et Twitter
- 2006** La bibliothèque nationale de France (BnF) et l'institut national de l'audiovisuel (Ina) sont chargés d'assurer le dépôt légal du web
- 2013** Création du projet EU Web Archive (<https://op.europa.eu/fr/web/euwebarchive>)
- 2025** 1 000 milliards de pages web archivées par l'Internet Archive et accessibles via la Wayback Machine

- Macro archivage ;
- Micro archivage ;
- Archivage sélectif et thématique ;
- Moissonnages par lots, à large spectres ou de *snapshots* ;
- Tentative d'exhaustivité d'archivage, par exemple à partir d'un nom de domaine national (.se en Suède) ;
- Archivage institutionnel d'événements ou archivage en temps réel, ex l'incendie de Notre Dame en 2019 ;
- Coopération et partage d'archivage entre différentes institutions (BnF / Ina)
- ...



Dépôt légal web BnF

@DLwebBnF

Abonné

Merci à [@BNE_biblioteca](#) et à toutes les institutions membres de [@NetPreserve](#) qui ont participé à cette collecte par le biais d'une sélection de sites web !



Biblioteca Nacional de España [@BNE_biblioteca](#)

Estamos colaborando con la Biblioteca Nacional de Francia en una recolección web sobre el incendio de Notre Dame

[Afficher cette discussion](#)

17:02 - 25 avr. 2019

8 Retweets 15 J'aime



8



15



Les différentes façons d'accéder aux archives du web

- Les archives du web en Finlande (depuis 2006):
<https://www.kansalliskirjasto.fi/en/collections-and-content-online#finnish-web-archive>

Les archives du web finlandais sont accessibles uniquement depuis les terminaux du dépôt légal du web à la bibliothèque nationale et dans un réseau de bibliothèques publiques sur le territoire.

- Au Portugal (depuis 2008) les archives du web sont accessibles en libre accès Arquivo.pt
- La Wayback Machine (depuis 2006) est accessible en libre accès <https://archive.org/web/>

Un large éventail d'archives du web

Comme pour les autres types d'archives, il est utile de connaître le contexte de production de l'archive pour mieux la comprendre et l'utiliser dans un travail académique. Lorsque vous consultez une archive du web, il s'agit d'une reconstruction et pas d'une simple copie.

- “What is harvested is both a point in time (the time of harvesting) and a period of time (the period up to the time of harvesting).” (Brügger, 2008 p. 158)
- “On the one hand the archive does not look like the internet as it actually was in the past (we have lost something), but on the other hand the archive might look like the internet as it never was in the past (we get something different).” (Brügger, 2001 p. 6)

Les projets d'archivage du web ont besoins des compétences de différents acteurs et de leurs expertises : chercheur·e·s, archivistes, bibliothécaires, juristes, informaticien·ne·s, spécialistes de la données, ... usagères et usagers, partenaires et membres de la société civile.

Comme pour les autres types d'archives, il est nécessaire d'agir dans le cadre légale et éthique depuis le processus d'archivage et jusqu'à l'usage qui sera fait des archives du Web :

- Les documents contenus dans les archives Web sont protégés par les lois sur le droit d'auteur, comme ils le sont sur le Web "vivant".
- Il existe des "tensions entre les principes archivistiques de préservation des documents publics et les attentes des citoyens en matière de droit à l'oubli" (Bingam, 2018)
- Le traitement des données personnelles est soumis à des lois et plus encore à l'éthique du projet de recherche.
- règlement général sur la protection des données (RGPD) / GPDR / General Data Protection Regulation (GPDR)

Peur, incertitude et doute sur les archives du Web



Research Business What's on Learning Visit **Cyber incident** More Shop

British Library / Cyber incident

Cyber-attack update

We're continuing to experience disruption as a result of a cyber-attack that took place in October 2023. The outage is still affecting our website, online systems and services, as well as some onsite services, however our buildings are open as usual.

The attack caused substantial damage that is complex and challenging to repair, beginning with the installation of a completely new computing infrastructure for the entire Library.

As the new academic year begins we are restoring a number of services. Our CEO, Sir Roly Keating, explains more in a [new blog](#).

Our teams have been working since the cyber-attack to find ways to restore access to as much of our collection as possible, but disruption to certain services is expected to persist for some time. Find out more about the impact on our [research services](#).

We have published a paper about the attack and its impact, [introduced in this blog](#) by our Chief Executive. Its goal is to share our understanding of what happened and to help others learn from our experience.

In November the attackers released some of our data onto the dark web including some personal user information. We've contacted our users to alert them to this incident and to offer advice from the National Cyber Security Centre (NCSC) on how to protect themselves, including updating their passwords on other systems.

If you have any questions relating to your data you can email our Data Protection Officer at data.governance@bl.uk

Because our systems were badly damaged during the cyber-attack they remain unavailable, so you can't change the password for our services. However, if you use the same password for non-British Library services, we recommend that you change it as a precaution.

The NCSC offers [advice](#) on staying secure online, including how to create a strong password, as well as [specific guidance](#) for individuals who may have been impacted by a data breach.

We're really sorry for the ongoing disruption to our systems and services and we'll provide further updates when we can. Thank you for bearing with us.



internetarchive
@internetarchive@mastodon.archive.org

Latest update on the DDOS attack from @brewsterkahle (Oct 11 @ 10:22am PT):

"The data is safe.

Services are offline as we examine and strengthen them. Sorry, but needed. @internetarchive staff is working hard.

Estimated Timeline: days, not weeks.

Thank you for the offers of pizza (we are set)."

Oct 11, 2024, 20:00 · Web · 1.1K · 1.5K

Service Availability

[Wayback Machine \(provisional, read-only\) service.](#)

Other Internet Archive services are temporarily offline.

Please check our official accounts, including [Twitter/X](#), [Bluesky](#) or [Mastodon](#) for the latest information.

We apologize for the inconvenience.

<https://www.bl.uk/cyber-incident/>

<https://mastodon.archive.org/@internetarchive/113290094683712739>

Partie 3 : une communauté internationale structurée



INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM

[HOME](#) [ABOUT IIPC](#) [WEB ARCHIVING](#) [EVENTS](#) [BLOG](#) [JOIN US](#)



The web is a unique
and dynamic resource that
is of high value to current
and future researchers



[Learn about the value of our work](#)

IIPC Members

Members are organizations from over 45 countries, including national, university and regional libraries and archives.

Working groups

IIPC members join **working groups** that engage in short and long-term **projects** to advance the practice of **web archiving**.

Events

Our community comes together annually to share experiences and present solutions during the **Web Archiving Conference** and the **General Assembly**.

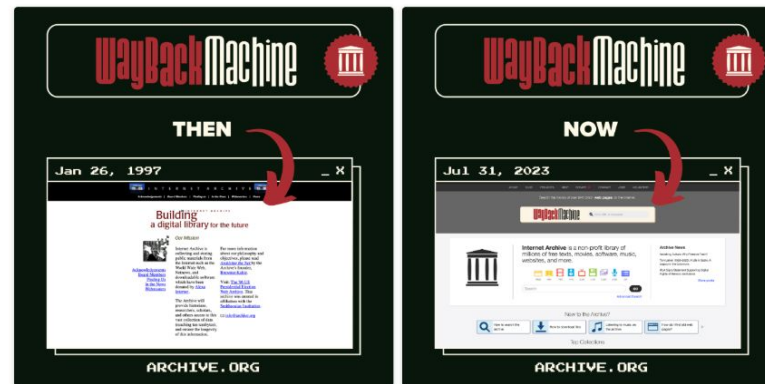
Internet Archive Blogs

A blog from the team at archive.org

[Blog](#) [Announcements](#) [archive.org](#) [About](#) [Events](#) [Developers](#) [Donate](#)

Calling All Libraries: Celebrate 1 Trillion Web Pages Archived with the Internet Archive

Posted on [October 7, 2025](#) by [Chris Freeland](#)



The Internet Archive has released a new [resource guide](#) to help libraries join in commemorating a once-in-a-generation milestone: 1 trillion web pages archived by the Internet Archive and available for use via the Wayback Machine. This historic achievement represents a collective effort to preserve our shared digital history, and libraries have been at the heart of that work from the beginning. The resource guide provides background on why this milestone matters to libraries, highlights their role in safeguarding online knowledge, and offers ready-to-use materials to help libraries celebrate with their communities.

Celebrate with the Internet Archive this October: [1 trillion web pages archived](#)

Latest Posts

- [Calling All Libraries: Celebrate 1 Trillion Web Pages Archived with the Internet Archive](#)
- [Happy College Radio Day! Explore Student Radio's Past and Present in the DLARC College Radio Collection](#)
- [Peer-to-Peer Fundraising at the Internet Archive](#)
- [Rural Libraries Receive Support from the Internet Archive to Preserve Community Stories](#)
- [Sir Tim Berners-Lee to Receive the 2025 Internet Archive Hero Award](#)

Upcoming Events

OCT 7:30 pm - 8:30 pm
9 Building and Preserving the

- Lancé en 2001
- 1000 milliards de pages web archivées (oct 2025)
- Archives remontant jusqu'en 1996

<https://web.archive.org/>

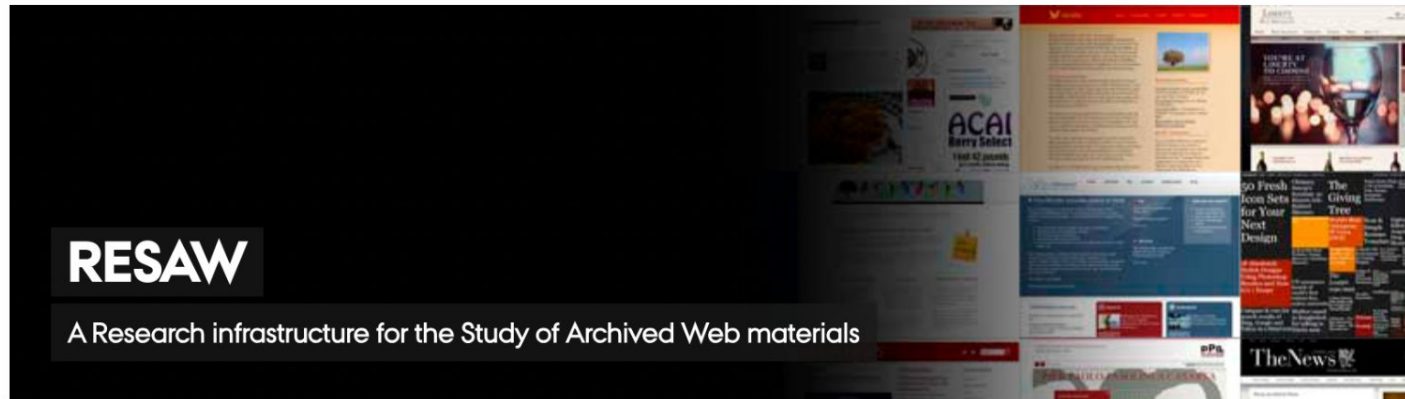
<https://cc.au.dk/en/resaw>

RESAW - Infrastructure de recherche pour l'étude des sources issues du web archivé



RESAW

- › About
- › Events
- › RESAW timeline
- › Participants
- › Web Archives



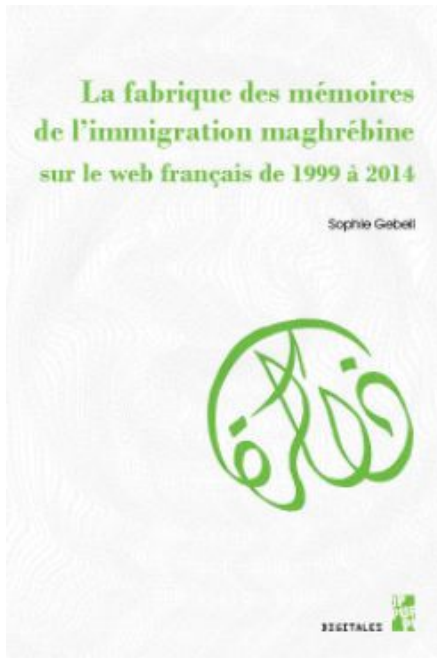
Home

RESAW, a Research Infrastructure for the Study of Archived Web Materials, is a community established in 2012, aiming at promoting a collaborative European research infrastructure for the study of archived web materials.

Partie 4 : L'archivage du Web. Quels défis pour les SHS ?



Du Web comme source aux défis épistémologiques



<https://presses-universitaires.univ-amu.fr/fabrique-memoires-limmigration-maghrebine-web-francais-1999-a-2014>

L'archivage du Web comme défi historiographique : entre fragmentation et médiation (Projet [IUF](#))
>> Questionnaire sur les usages

<https://madi.hypotheses.org/1436>

Parcours entre Histoire et SIC :

- Une première expérimentation en 2015 (thèse). Dépôt légal du Web français (BnF+INA), Internet Archive
- Réflexion méthodologique et épistémologique (2021) *Website Story, Histoire, mémoires et archives du Web*

Des travaux ancrés dans le domaine de recherche sur et avec les archives du web : international, interdisciplinaire, pluricatériel.

Principaux enjeux pour les SHS :

- Etat de l'art : besoin de développer des connaissances sur l'usage de l'archivage du web et ses implications épistémologiques
- Source / terrain / documents : accès aux documents nativement numériques et stabilisation des corpus issus du web vivant et des médias sociaux (citabilité)
- Web archivé, terrain de jeu des humanités numériques => méthodes computationnelles
- Mémoire et patrimoine nativement numériques : collectes, formats etc.

De nouveaux champs de recherche, de nouveaux réseaux et projets

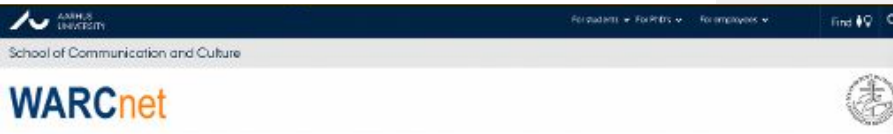
<https://web90.hypotheses.org/>

Web90 – Patrimoine, Mémoires et Histoire du Web dans les années 1990

"Technological progress has merely provided us with more efficient means for going backwards." — Aldous Huxley, Ends and Means

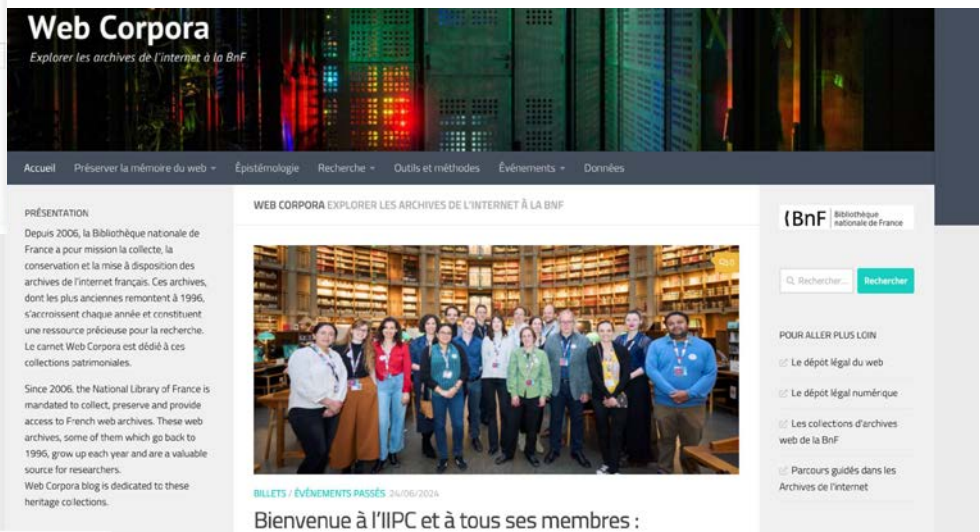
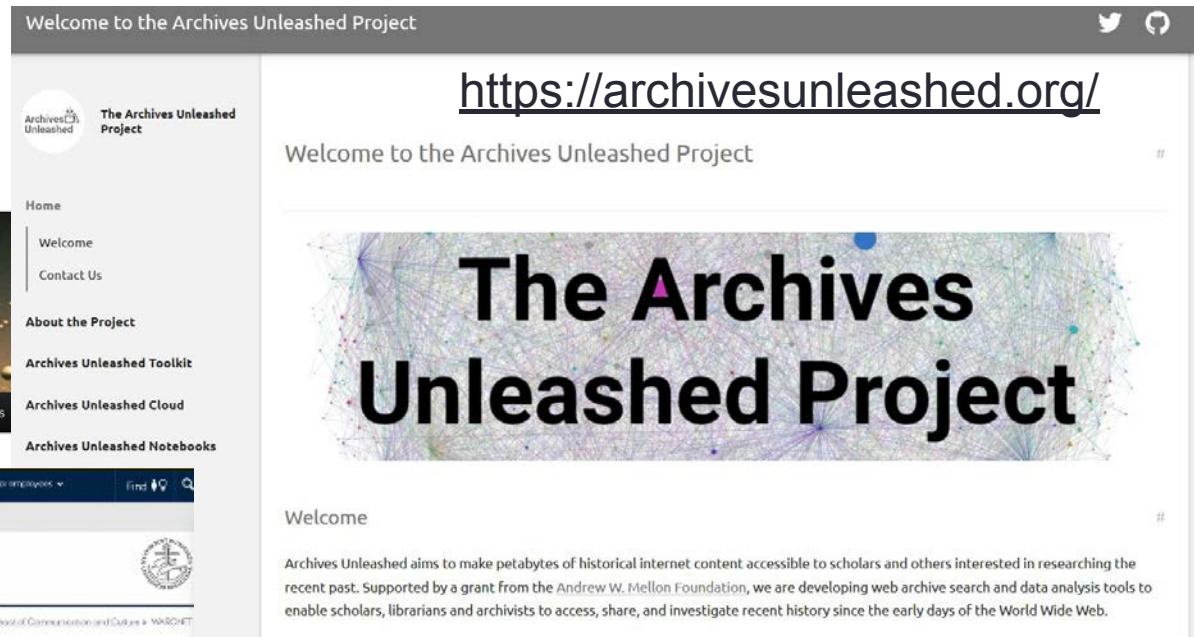


Le projet ANR Web90 Recherches en cours TTOW Symposium Veille L'équipe Crédits



<https://cc.au.dk/en/warcnet> (2021-2023)

<https://webcorpora.hypotheses.org/>



RESAW 2023

Aix-Marseille université
Société engagée

TELEMME

CNRS

A+Midex

FRANCE 2030

Mucem

Maison méditerranéenne des sciences de l'homme

PICCH

Aix-Marseille université

INSPE

FIRST RESAW CONFERENCE IN FRANCE
WITH THE SUPPORT OF THE INTERDISCIPLINARY MISSION OF AIX-MARSEILLE UNIVERSITÉ

**EXPLORING
THE ARCHIVED WEB
DURING A HIGHLY
TRANSFORMATIVE AGE**

**77 SPEAKERS - 50 SPEECHES
14 FIELDS OF RESEARCH**

TO UNDERSTAND THE CHALLENGES OF WEB ARCHIVING,
AN ESSENTIAL PART OF OUR HERITAGE, FOR CULTURE AND CIVIL SOCIETY
AT A TIME OF MULTIPLE UPEHAVALS

**5+6 JUNE 2023
MUCEM LAB, 1 ESP. J4
13002 MARSEILLE**

<https://resaw2023.sciencesconf.org>

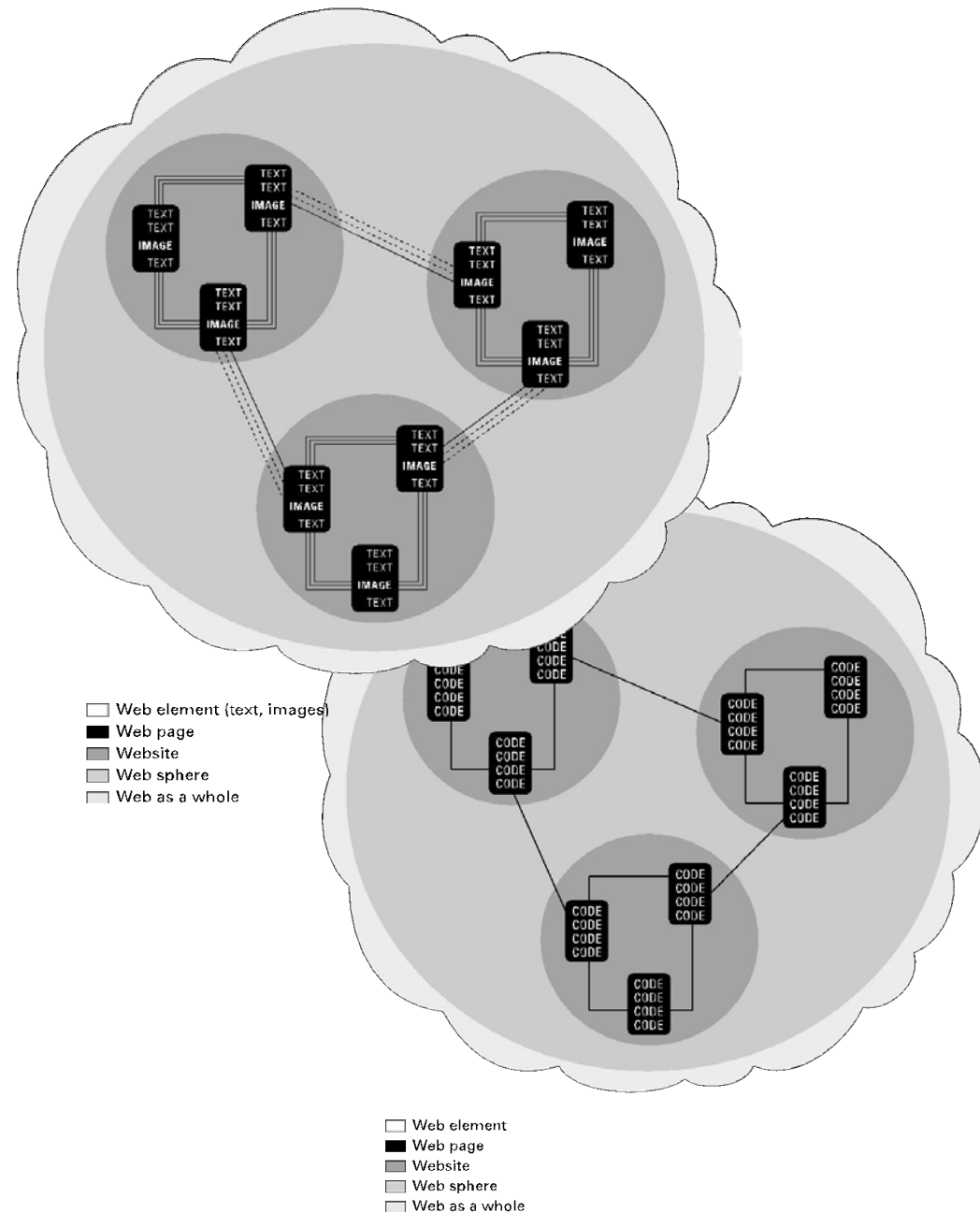
Given recent global crises, the imperative to preserve and analyze online content has never been more vital to enhancing our comprehension of contemporary changes. This book, the outcome of the 5th international RESAW conference that convened experts from fifty disciplines across seventeen countries in Marseille in June 2023, tackles the multifaceted challenges of web archiving. It underscores the dual roles of web archiving, as cultural heritage and as essential source material for researchers delving into contemporary events and the evolution of digital culture. Through twenty chapters, it explores the development of web archiving and examines how technical, cultural, geopolitical, societal, and environmental shifts impact its conception, study, and dissemination.

<https://books.fupress.com/catalogue/exploring-the-archived-web-during-a-highly-transformative-age/14127>



Les 5 couches analytiques du Web :

- Les éléments du Web
- Les pages web
- Les sites Web
- Les sphères du Web
- Le Web dans sa dimension globale



Deux couches de texte sur le Web : texte visible et texte invisible html

← → ↻ distam.hypotheses.org/11308

référence et de travail pour l'essor des humanités numériques appliquées aux études aréales.

À propos ▾ Activités ▾ Ressources ▾ Guides ▾ Nous rejoindre ▾ Contacts ▾

» ÉCOLE D'ÉTÉ «

École d'été DISTAM du 8 au 12 juillet 2024

PAR MARIE BIZAIS-LILLIG · PUBLIÉ 08/04/2024 · MIS À JOUR 04/07/2024



“Les humanités numériques en études aréales”

Le consortium DISTAM (Digital Studies Africa, Asia, and the Middle East) organise son école d'été du 8 au 12 juillet **au Couvent de la Tourette**. Le couvent est situé à 2 kilomètres de la gare de l'Arbresle (Éveux), à l'ouest de Lyon. Une navette entre la gare de l'Arbresle et le Couvent est disponible.

DevTools is now available in French!

Always match Chrome's language Switch DevTools to French Don't show again

Elements Console Sources Network >> 1

```
<!DOCTYPE html>
<html class="js" lang="fr-FR" style="position: inherit;">
  <head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <link rel="profile" href="https://qmpg.org/xfn/11">
    <link rel="pingback" href="https://distam.hypotheses.org/xmlrpc.php">
    <title>École d'été DISTAM du 8 au 12 juillet 2024 - DISTAM</title>
    <meta name="robots" content="max-image-preview:large">
    <link rel="icon" type="image/png" href="https://distam.hypotheses.org/wp-content/cleocommerce/images/favicon.png">
    <script type="text/javascript" async defer src="https://analytics.openeditio
n.org/matomo.js"></script>
    <script>
    <link rel="dns-prefetch" href="//distam.hypotheses.org">
    <link rel="alternate" type="application/rss+xml" title="DISTAM » Flux" href=
"https://distam.hypotheses.org/feed">
    <link rel="alternate" type="application/rss+xml" title="DISTAM » Flux des com
mentaires" href="https://distam.hypotheses.org/comments/feed">
    <link id="hu-user-gfont" href="//fonts.googleapis.com/css?family=Ubuntu:400,4
00italic,300italic,300,700&subset=latin,latin-ext" rel="stylesheet" type="tex
t/css">
    <link rel="alternate" type="application/rss+xml" title="DISTAM » École d'été
DISTAM du 8 au 12 juillet 2024 Flux des commentaires" href="https://distam.hyp
otheses.org/11308/feed">
    <script>
    <script id="wp-emoji-styles-inline-css">
    <link rel="stylesheet" id="wp-block-library-css" href="https://distam.hypothe
ses.org/wp-includes/css/dist/block-library/style.min.css?ver=6.5.4" media="al
l">
    <script id="classic-theme-styles-inline-css">
    <script id="global-styles-inline-css">
    <link rel="stylesheet" id="wp-hal-style1-css" href="https://distam.hypothes
e.s.org/wp-content/plugins/wp-syntax/css/wp-syntax.css?ver=1.2" media="all">
    <link rel="stylesheet" id="wp-syntax-css-css" href="https://distam.hypothes
e.s.org/wp-content/plugins/wp-syntax/css/wp-syntax.css?ver=1.2" media="all">
    <link rel="stylesheet" id="parent-style-css" href="https://distam.hypothes
e.s.org/wp-content/themes/hueman/style.css?ver=6.5.4" media="all">
  </head>
  <body>
```

html.js head title

Styles Computed Layout Event Listeners DOM Breakpoints Properties >>

Filter :hov .cls +

Ce qui est archivé est différent de ce que l'on voit en ligne



Ajout numérique des études aréales -- Afrique, Asie, Moyen-Orient et Mondes musulmans

Le carnet de consortium Huma-Num DISTAM (Digital Studies Africa, Asia, Middle East) est conçu comme un espace de référence et de travail pour l'essor des humanités numériques appliquées aux études aréales.

À propos + Activités + Ressources + Guides + Nous rejoindre + Contacts

SUIVRE

ÉCOLE D'ÉTÉ

PLUS

ARTICLE SUIVANT

Newsletter Distam #33 Avril 2024

ARTICLE PRÉCÉDENT

< [25 mars 24] Des images d'un texte à l'encodage et l'édition en TEI XML

Depuis 2022, DISTAM (Digital Studies Africa, Asia, Middle East) est un consortium d'Huma-Num destiné à accompagner et à consolider la transition numérique des études moyen-orientales, africaines et asiatiques en France. DISTAM est affilié à l'UAR 2999 d'études aréales du CNRS basé au Campus Condorcet.

DISTAM s'inscrit à des langues à graphies non-latin et dispose d'outils d'encodage et d'édition en TEI XML.

École d'été DISTAM du 8 au 12 juillet 2024

PAR MARIE BIZAS-ALLAC Publié 08/04/2024 - Mis à jour 08/07/2024



"Les humanités numériques en études aréales"

Le consortium DISTAM (Digital Studies Africa, Asia, and the Middle East) organise son

Recherche

Abonnez-vous à la newsletter DISTAM

GP BULAC

GS des Études Africaines

GS des Études Asiatiques

GS Moyen-Orient et Mondes Musulmans

Unité support CNRS aux études aréales

```
<!DOCTYPE html>
<html class="no-js" lang="fr-FR">
<head>
  <meta charset="UTF-8">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <link rel="profile" href="https://gmpg.org/xfn/11">
  <link rel="pingback" href="https://distam.hypotheses.org/xmlrpc.php">

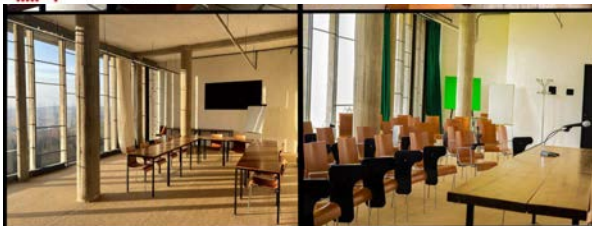
  <title>École d'été DISTAM du 8 au 12 juillet 2024 <meta name="robots" content="max-image-preview:large" />
  <link rel="icon" type="image/png" href="https://distam.hypotheses.org/wp-content/uploads/2024/07/favicon.png" />
  <link rel="dns-prefetch" href="https://distam.hypotheses.org/" />
  <link rel="alternate" type="application/rss+xml" title="DISTAM Flux des commentaires" href="https://distam.hypotheses.org/comments/feed" />
  <link id="user-gfot" href="https://fonts.googleapis.com/css?family=Ubuntu:400,400italic,300,300italic,300,700&subset=latin,latin-ext" rel="stylesheet" type="text/css">
  <script>
    window._wpemojiSettings = {"baseurl":"https://s.w.org/images/core/emoji/15.0.3/72x72/", "ext":".png", "svgurl":"https://s.w.org/images/core/emoji/15.0.3/svg/", "svgExt":".svg"}
    /*! This file is auto-generated */
    function(i,n){var o,s,e;function c(e){try{var t={supportTests:e,timestamp:(new Date).valueOf(),sessionStorage.setItem(o,JSON.stringify(t))}catch(e){}}function p(e,t,n){e.clearRect(0,0,e
  </script>
  <style id="wp-emoji-styles-inline-css">
    img.wp-smiley, img.emoji {
      display: inline !important;
      border: none !important;
      box-shadow: none !important;
      height: 1em !important;
      width: 1em !important;
      margin: 0 !important;
      vertical-align: middle !important;
      background: none !important;
      padding: 0 !important;
    }
  </style>
  <link rel="stylesheet" id="wp-block-library-css" href="https://distam.hypotheses.org/wp-includes/css/dist/block-library/style.min.css?ver=6.5.4" media="all" />
  <style id="classic-theme-styles-inline-css">
    /*! This file is auto-generated */
    .wp-block-button__link{color:inherit;background-color:#323738;border-radius:9999px;box-shadow:none;text-decoration:none;padding:calc(1.667em + 2px) calc(1.333em + 2px);font-size:1.25em;}
  </style>
  <style id="global-styles-inline-css">
    :root{--wp--preset--color--black: #000000;--wp--preset--color--cyan-blush-gray: #abb8c3;--wp--preset--color--white: #ffffff;--wp--preset--color--pale-pink: #f78da7;--wp--preset--color--violet: #c787ff;
  </style>
  <link rel="stylesheet" id="wp-hal-style-css" href="https://distam.hypotheses.org/wp-content/plugins/hal/css/style.css?ver=6.5.4" media="all" />
  <link rel="stylesheet" id="wp-syntax-css-css" href="https://distam.hypotheses.org/wp-content/plugins/wp-syntax/css/wp-syntax.css?ver=1.2" media="all" />
```

UAR2999

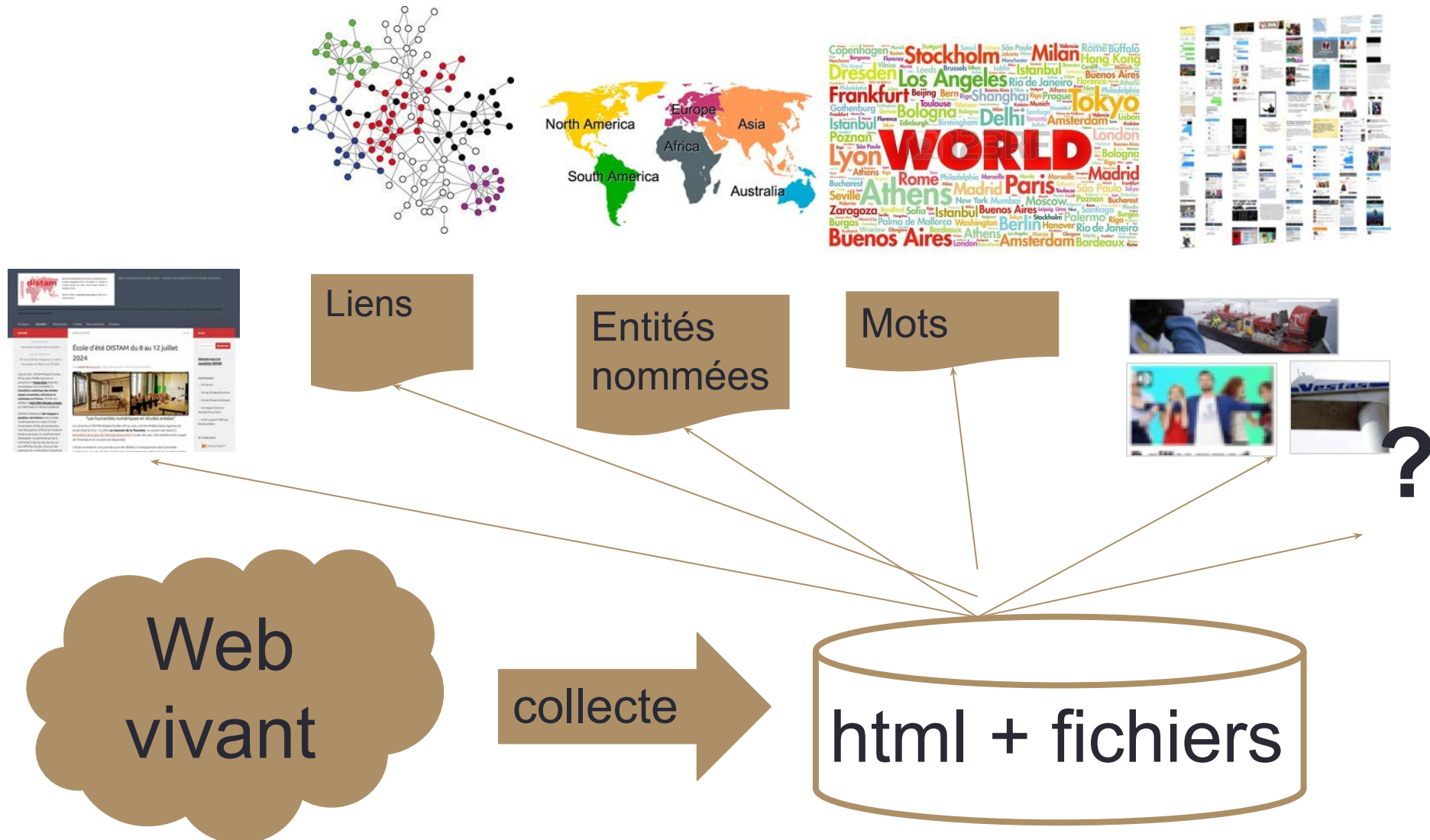


Depuis 2022, DISTAM (Digital Studies Africa, Asia, Middle East) est un consortium d'Huma-Num destiné à accompagner et à consolider la transition numérique des études moyen-orientales, africaines et asiatiques en France.

DISTAM est affilié à l'UAR 2999 d'études aréales du CNRS basé au



Une archive du Web = une masse de fichiers interconnectés





Why GitHub? ▾ Enterprise Explore ▾ Marketplace Pricing ▾

Search

Sign in

Sign up

internetarchive / heritrix3

Watch

173

★ Star

1,454

Fork

601

<> Code

🔔 Issues 19

🔗 Pull requests 2

📁 Projects 0

📖 Wiki

📊 Insights

Heritrix is the Internet Archive's open-source, extensible, web-scale, archival-quality web crawler project. <https://heritrix.readthedocs.io/>

java

webcrawling

warc

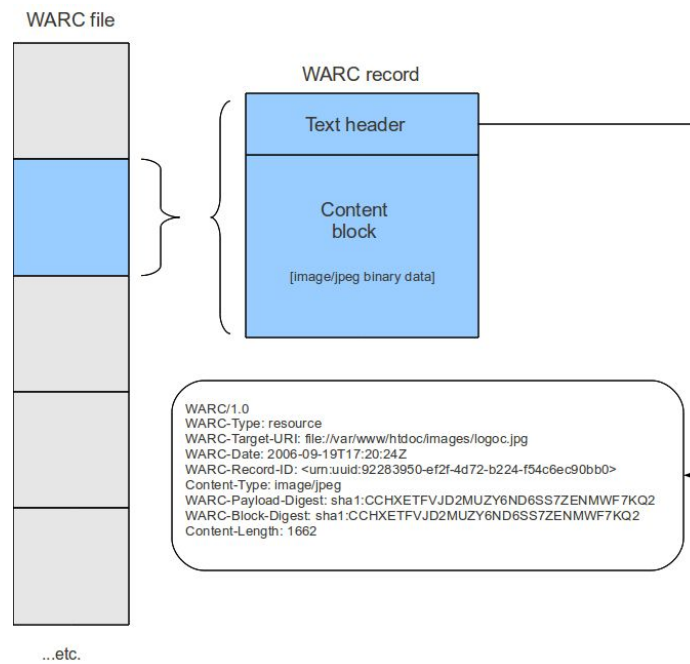
heritrix

Heritrix is the Internet Archive's open-source, extensible, web-scale, archival-quality web crawler project

📁 commons	[maven-release-plugin] prepare for next development iteration	9 days ago
📁 contrib	[maven-release-plugin] prepare for next development iteration	9 days ago
📁 dist	[maven-release-plugin] prepare for next development iteration	9 days ago
📁 docs	Note feature only applies to forthcoming 3.3 release	6 months ago
📁 engine	[maven-release-plugin] prepare for next development iteration	9 days ago
📁 modules	[maven-release-plugin] prepare for next development iteration	9 days ago
📄 .gitignore	Format HTTP request lines in API guide	10 months ago
📄 .travis.yml	allow travis-ci jdk7 failures because contrib...	a month ago
📄 CHANGELOG.md	Update changelog for 3.4.0-20190418	9 days ago

<https://github.com/internetarchive/heritrix3>

- WARC file format = Web ARChive archive format
- WARC format international depuis 2009 (ISO 28500:2017)



https://en.wikipedia.org/wiki/Web_ARChive

https://wiki.archiveteam.org/index.php/The_WARC_Ecosystem

<https://wiki.archivematica.org/File:WARCdiagram.png>

Consulter le web archivé, un nouveau rapport à l'archive

Un processus de *remediation* (Niels Brügger, 2005)

CNHI,
Page d'accueil,
Internet Archive,
05/10/2007



New Browsertrix 1.18: Large URL Lists and Beautiful Emails



WEBRECORDER

Tools ▾

Blog

About

Resources



BROWSERTRIX

Overview

Sign In

Sign Up

Web archiving for Universities & Colleges

Webrecorder provides open source solutions for everyone to archive the complex, interactive Web.

[Start a free trial of Browsertrix, our complete archiving platform →](#)



<https://webrecorder.net/>



Explore >> Publications Office of the European Union >> European Union



European Union
Collected by: [Publications Office of the European Union](#)

Archived since: avr., 2019
Description: This collection contains the websites of the EU institutions: the European Parliament, the European Council, the Council of the European Union, the European Commission, the Court of Justice of the European Union, the European Central Bank and the Court of Auditors. This collection also includes the websites of the agencies and other bodies of the European Union. Most of these are hosted on the europa.eu domain and subdomains.
Subject: [Government](#)

Narrow Your Results

Group Sort By: Count | (A-Z)

Agencies and other bodies (137)
Committee of Regions (20)
EC - Representations (31)
Economic and Social Committee (9)
European Central Bank (10)

More ▼

Topic Sort By: Count | (A-Z)

EU institutions (125)
Information and communication (52)
Covid-19 (51)

Sites for this collection are listed below. Narrow your results at left, or enter a search query below to find a site, specific URL or to search the text of archived webpages.

Enter search terms here

SearchClear

Sites Search Page Text

Page 1 of 11 (1 091 Total Results) Next Page ►

Sort By: Title (A-Z) | Title (Z-A) | URL (A-Z) | URL (Z-A)

Title: 1999 European elections

URL: <https://www.europarl.europa.eu/election/>

Description: This website provides information about the results of the 1999 European elections, as well as the electoral laws, names of candidates, entering Parliament and the highlights of the 1999

Archive-It: exemple d'une page archivée avec Archive-it

You are viewing an archived web page, collected at the request of [Digital Humanities Institute Beirut 2019](#) using [Archive-It](#). This page was captured on 12:20:14 mai 01, 2019, and is part of the [DHIB 2017](#) collection. The information on this web page may be out of date. See [All versions](#) of this archived page. [Metadata](#) [Enable QA](#)

> DHIB.

DIGITAL
HUMANITIES
INSTITUTE –
BEIRUT

HOME REGISTRATION SCHEDULE WORKSHOPS KEYNOTES
PANELS PEOPLE SPONSORS MAPS LOGO AND MEDIA
GENERAL INFO ARCHIVE ▾ CONTACT US PRE-INSTITUTE TALKS
PLACES TO VISIT IN LEBANON

Privacy & Cookies: This site uses cookies. By continuing to use this website, you agree to their use.
To find out more, including how to control cookies, see here: [Cookie Policy](#)

Close and accept

<https://wayback.archive-it.org/12102/20190501122014/https://dhibeirut.wordpress.com/archive/dhi-b-2017/>

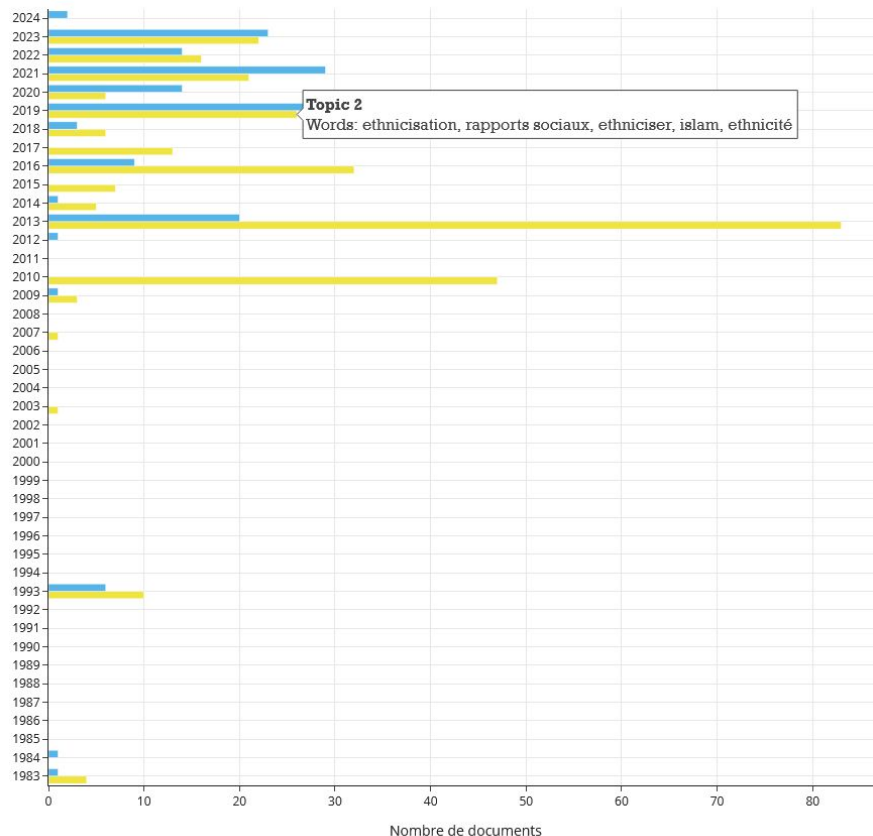
L'archivage du Web, terrain de jeu des humanités numériques

Quarante ans de médiatisation de la Marche de 1983 à la TV et en ligne

Topics par an

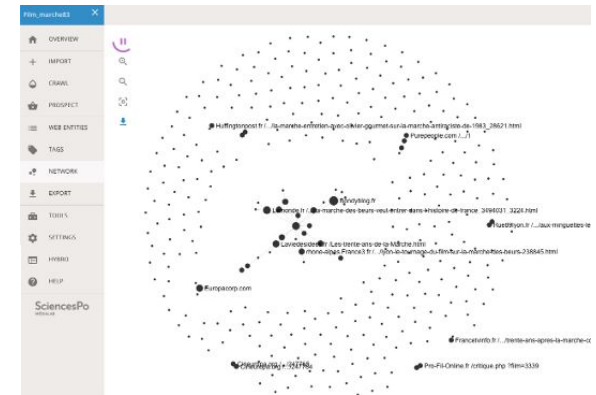
(Nombre de documents assignés au topic T pour l'année N)

Paramètres utilisés : chunk_lgc_token, 64 composants / 145 neighbors / 48 docs min per cluster



Global Topic Representation

- 1_police_algérie_marche beurs_école
- 0_mort_bruit_algérien_toumi djaidja
- 1_cinéma_jamel debbouze_cinématographique...
- 2_voile_mosquée_mosquées_foulard
- 3_marche 1983_marche 83_marche beurs_évé...
- 4_100 000 personnes_marseille 15 octobre...
- 5_mythe_main œuvre_pays origine_algérien...
- 6_parti socialiste_antiracisme_sos racis...
- 7_verlan_mot beur_mot arabe_féminin
- 8_cités_emploi_grands ensembles_chômeurs
- 9_égalité_égalité_égalité_racistes droit...
- 10_filmer_victimes_application_pratiques
- 11_pris décision_ramadan_chanté fois foi...
- 12_université_budget_bâtiment_uvsv
- 13_raptor_super_peser_petit journal
- 14_forces ordre_nuit dernière_émeutes_ro...
- 15_séjour 10 ans_carte séjour 10_carte 1...
- 16_transmission_éducation nationale_manu...
- 17_étape_valence_strasbourg_grenoble
- 18_message paix_message_paix_message éga...
- 19_micro_désolé_intervenants_réagir
- 20_maison_slip_123_facebook
- 21_cibles_actes_graphique_iran
- 22_chanson_morceau_artiste_album
- 23_contes_contes_podcast_spectacle
- 24_martin luther king_mandela_étoiles no...
- 25_main tendue_point levé_-ce bon choix_...
- 26_partis politiques_syndicats_monsieur ...



Projet Polyvocal
Interpretation of a
Contested Colonial
Heritage (2022-2024).
Collaboration avec Le
Lab (INA)

Web Archiving Aix Marseille (POC)

 **BROWSERTRIX**

Web-Archives

 Guide d'utilisation



[Dashboard](#) [En cours de collecte](#) [Items archivés](#) [Collections](#) [Profils de navigateur](#) [Paramètres](#)

Web-Archives

 [+ Créer un nouveau...](#)

Stockage

 18 Collectes	63,4 Go
 1 Chargé	2,97 Go
 0 Profils de navigateur	0 o
 19 Items archivés	66,4 Go

En cours de collecte

- 0 Collectes en cours
-  0 Flux de travail (workflow) de collectes en attente

 61 744 Pages collectées
 6 579 Pages chargées
 68 323 Pages Total

Collections

-  4 Total des collections
-  1 Collection partageable

Public Collections

[Preview public collections gallery](#)

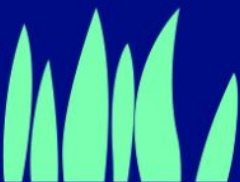
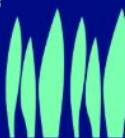
  



octobre 2024

AMU

https://festivaljeudeloie.fr/ 09/10/2024 16:30:49



REPLAY 2020 Festival des sciences sociales et des arts

Témoignages, débats scientifiques, expériences pédagogiques et créations artistiques pour investir autrement l'avenir

Édition en ligne



Confinement(s)
Tout un monde
à l'arrêt ?

Confinements au
quotidien

Performances
pédagogiques

Impacts sur le
territoire

Expositions & replay
> en ligne < tous publics

Mucem

Créations
artistiques

Interprétations et
débats

Partie 5 : Vers une exploration collective du Web archivé

- Un domaine structuré au niveau international : avec des archivistes et des communautés académiques, une ingénierie spécialisée
- Un domaine en évolution constante : platformisation, APIcalyps, défis de l'IA générative
- Une réflexion interdisciplinaire et pluri-catégorielle : séminaire avec des enjeux épistémologiques, patrimoniaux, digitaux, politiques de conservation

=> Collaboration WebLab et Humathèque

Séminaire 2025/2026 WebLab-Humathèque

Le Web et les archives du Web pour la recherche en SHS : savoirs, méthodes et outils pour la collecte, l'analyse et la pérennisation de corpus en ligne

Un site Web : https://pba.mmsh.fr/?page_id=1465

Une liste de diffusion : weblab@services.cnrs.fr

Une formation doctorale sur ADUM

Programme en cours de construction (7 séances)

Jeudi 27 nov. 14h-16h :

Institutions d'archivage et recherche en SHS : Le dépôt légal du web français (BnF et INA)

Mercredi 3 décembre 14/16h 2025

Stéphane Pouyllau (Huma-Num)

Thèmes à venir :

- Actualité du domaine
- Aspects éthiques et juridiques
- Enjeux méthodologiques : Hyphe, WAAM
- Analyse des usages

<https://www.zotero.org/groups/2908451/distam-references/collections/3ZYJKDFQ>