

Ce(tte) œuvre est mise à disposition selon les termes de la

Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International.



# L'histoire nativement numérique : défis, sources et méthodes

Sophie Gebeil

UMR TELEMMe (AMU-CNRS)

12 décembre 2019 – Data Week (TGIR PROGEDO)

Aix-en-Provence

[www.madi.hypotheses.org](http://www.madi.hypotheses.org)

39 slides



UMR 7303

Temps, Espaces, Langages,  
Europe Méridionale,  
Méditerranée



Aix-Marseille  
université  
Initiative d'excellence

## Internet et les historien·nes : une « relation singulière » (P. Mounier)

De la *computing history* à l'histoire numérique Histoire et informatique (Le Roy Ladurie)

Méthodologies de l'histoire : « La boîte à outil des historiens »

Histoire des techniques et de l'innovation : Valérie Schafer

Internet et les historiens ( P. Rygiel, S. Noiret, J.-P. Genet, F. Clavert)

Le web comme source pour une histoire du temps présent

# Etudier la fabrique des mémoires de l'immigration maghrébine ... sur le web



1:33 / 2:32:19



Mémoires d'immigrés l'héritage maghrébin 1 - 2 - 3

90 766 vues

401 18 PARTAGER



Maroane Boutayeb  
Ajoutée le 16 juin 2012

S'ABONNER 105

Enfin l'intégrale disponible sur youtube ...

À suivre

LECTURE AUTOMATIQUE



Immigration et délinquance :  
l'enquête qui dérange (1/2) -  
Infrarouge  
348 k vues



Les Chibanis oubliés -  
Documentaire  
imineo Documentaires ✓  
34 k vues



Français d'origines contrôlées  
DOCUMENTAIRE PART2  
akimdemars  
12 k vues



Communauté Magrebine de  
Barbes ( Paris )  
Canalnaf Télévision  
7,4 k vues



Emigrés Nords-Africains de  
Belgique  
emya noiram  
4,7 k vues



Sofia Essaïdi On n'est pas  
couché 9 mai 2009 #ONPC  
On n'est pas couché ✓  
57 k vues

# Mémoires de l'immigration maghrébine sur le Web français de 1999 à 2014



Enjeu épistémologique : l'histoire contemporaine face aux sources nativement numériques => comment utiliser les archives du Web en tant qu'historien.ne ?



Entre études mémorielles et histoire des médias : comment le passé est-il mise en visibilité sur la toile, comment étudier les mémoires et la patrimonialisation des groupes minoritaires ?



# **Le Web comme source pour une histoire nativement numérique**

## Une simple étape supplémentaire dans le processus d'élargissement documentaire en histoire ?

*L'histoire se fait avec des documents écrits, sans doute. Quand il y en a. Mais elle peut se faire, doit se faire, sans documents écrits s'il n'en existe point. Avec tout ce que l'ingéniosité de l'historien peut lui permettre d'utiliser pour fabriquer son miel, à défaut des fleurs usuelles. Donc avec des mots, des signes. Des paysages et des tuiles. Des formes de champs et de mauvaises herbes. Des éclipses de lune et des colliers d'attelage (...) D'un mot, avec tout ce qui, étant à l'homme dépend de l'homme, sert à l'homme exprime l'homme, signifie sa présence, l'activité, les goûts et les façons d'être de l'Homme.*

Lucien FEBVRE, *Combats pour l'histoire*, 1953, p. 428, cité par Antoine PROST, *Douze leçons sur l'histoire*, op.cit., p. 83

# Spécificités des sources nativement numériques (Camille Paloque-Berges, 2016)

En SIC, 4 types de sources nativement numériques :

1. Les données numériques : « codes symboliques sous la forme de chiffres et de lettres, produites par l'utilisateur ou générées automatiquement par les logiciels dont il fait l'usage ».
2. Le code informatique : « la série d'instructions textuelles qui prévaut au traitement des données dans les programmes et les logiciels, de l'exécution d'une action sur les données (par les algorithmes) à l'affichage d'un contenu (par des instructions appelées « balisage ») sur les médias informatisés ».
3. Document numérique
4. Les artefacts matériels qui servent de supports à la conception et à l'usage des technologies numériques (équipements, infrastructures, etc.)

Paloque-Berges Camille, « Les sources nativement numériques pour les sciences humaines et sociales », *Histoire@Politique*, 2016/3 (n° 30), p. 221-244, <https://www.cairn.info/revue-histoire-politique-2016-3-page-221.htm>

## Des documents numériques....

- L'adjectif numérique renvoie aux « méthodes technoscientifiques de traitement de l'information par le codage et le calcul à travers les programmes informatiques, et se différencie des techniques dites analogiques » (Camille Paloque-Berges, 2016)

- Documents multimédias et composites

- Documents dématérialisés ?

La distinction entre l'information et son support

“Le support numérique n'est plus un morceau de matière mais une chaîne matérielle et logicielle qui à la fois supporte et transmet le contenu : outils d'écriture, formats d'encodage, support de stockage, matériel de restitution”. (Marie-Anne Chabin, 2003)

- La possibilité de manipulation est co-originale à la nature même du support numérique [...]. Ainsi, l'essence du numérique [...] est-elle : "ça a été manipulé" » (Bachimont, 2010 : 158).

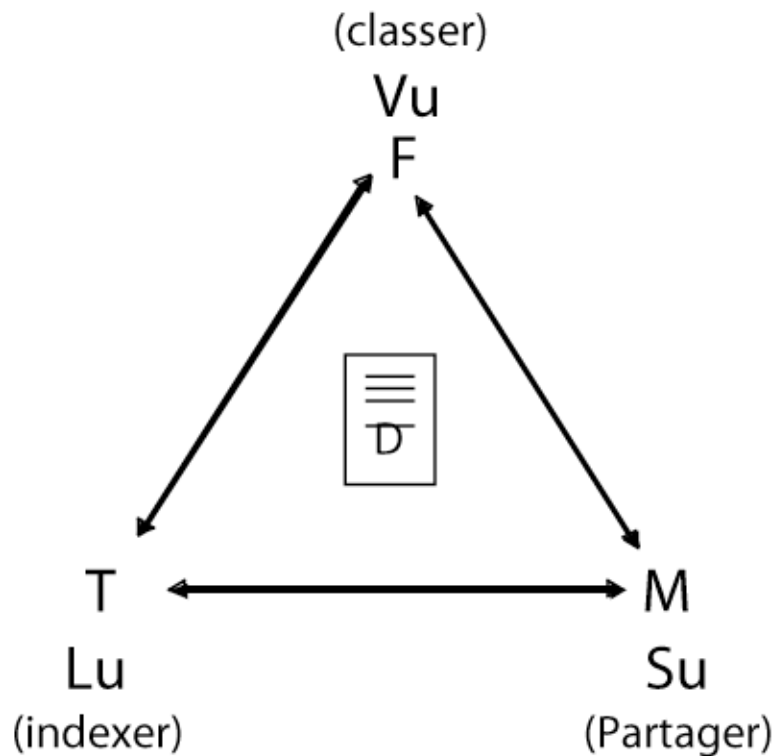
=> notion d'authenticité mise à mal



## ... en réseau

- Une page web n'existe pas isolément, elle s'inscrit dans un environnement structuré par le système interopérable de l'hypertexte sur lequel repose l'architecture du web (hyperliens, protocole d'échanges HTML)
- Une page visualisée dans un navigateur est le fruit d'un processus de médiation documentaire et informatique.
- Le document est inséré dans une multitude de traces informatiques générées par les parcours de navigation des internautes dont une partie est parfois mémorisée dans les ordinateurs des usagers (sous forme de cookies par exemple).

# Un nouvel ordre documentaire (la redocumentarisation) avec le numérique (Jean-Michel Salaün, 2012)



Dépasser  
le chaos (le silence et le  
bruit),  
la cacophonie (la  
confusion et le sensible)  
et l'oubli (l'intime et  
l'éphémère).

# L'historien·ne face aux sources nativement numériques

Rolando Minuti (2003), les technologies numériques entraînent « la fin de la matérialité » du document.

*(Les formulations des actes) font apparaître, sur un matériau quelconque et selon une forme déterminée, ce groupe de signes : la formulation est un événement qui, en droit au moins, est toujours repérable selon des coordonnées spatio-temporelles, qui peut toujours être rapporté à un auteur, et qui éventuellement peut constituer par elle-même un acte spécifique...*

(Michel Foucault, 1969)

*“Les sources primaires ne sont souvent plus reliées au contexte matériel qui leur faisait « prendre sens » et les validait dans leurs contextes : dans le monde numérique, un des grands problèmes est certainement celui de l'individuation des contextes signifiants, ce que les philologues appellent l'histoire de la construction des textes et des documents”*

(Serge Noiret et Frédéric Clavert, 2012)

## Le « contrat de vérité » mis à mal

Paul Ricoeur : l'historien, dans son travail « d'objectivation documentaire » est tenu « de répondre à la confiance que lui accorde son lecteur » (2001)

Marc Bloch :

“Car, hors des libres jeux de la fantaisie, une affirmation n'a le droit de se produire qu'à la condition de pouvoir être vérifiée ; et pour un historien, s'il emploie un document, en indiquer le plus brièvement possible la provenance, c'est à dire le moyen de le retrouver, équivaut sans plus à se soumettre à une règle universelle de probité”.

*“Le digital turn a rendu précaire un certain nombres de concepts chers aux historiens comme celui de la pérennité des sources et de la capacité de reproduire dans le temps une analyse qui s’y réfère (...) La permanence et la conservation des informations dans la toile est ainsi un problème central de son utilisation scientifique”. (Serge Noiret, 2011)*

# Le Web comme source pour une histoire du temps présent, un défis méthodologique

Accéder aux anciennes versions des sites Web

Disposer d'un corpus stable

Com'étéik  
diffusion

Objets Moyens Projets

Projet d'association  
Fonctionnement  
Territoire

Fruit de la rencontre entre globe-trotters, migrants, travailleurs sociaux, et p...  
Com'étéik diffusion a pour principaux

La culture permet de retrouver les savoirs...  
La p... com...  
effe... redon...  
colle...  
La c...  
soci...  
visar...  
en a...  
inév...

Le b...  
télé...

Cometik.info, snapshot, Internet Archive, 2007

arte

127, RUE DE LA GARENNE | Le bidonville de La Folie, Nanterre

Not Found

The requested URL / was not found on this server.

Apache/1.3.33 Server at bidonville-nanterre.arte.tv Po

01 03 05 07 09 10 12 15 16 18 20 21

« 127 rue de la Garenne, le bidonville de la Folie, Nanterre », Arte.fr, broken link : <http://bidonville-nanterre.arte.tv>



# Les archives du Web : apports et limites

---

## Le projet d'Internet Archive : la Wayback Machine (Brewster Kahle).

<http://archive.org/web/web.php>

Ici un exemple avec le site  
[www.oumma.com](http://www.oumma.com) pour la page  
du 30 novembre 2001, <  
[http://web.archive.org/web/20011130080811/  
http://oumma.com/](http://web.archive.org/web/20011130080811/http://oumma.com/)>, consultée le 02/12/12

http://oumma.com/ Go

673 captures 4 déc. 00 - 12 jul. 11

SEPT. NOV. JANV. 30 2000 2001 2002

# Oumma.com

Bienvenue sur le site de référence de l'Islam francophone 2 Décembre 64 | 7 Rabi' Al-Thani 575

Accueil Articles Communication Services Nous contacter

- Informations
  - Tous les articles
  - Les plus lus
  - Questions/Réponses
- Communication
  - Forum
  - Agence de rencontre
  - Petites annonces
- Services
  - Le Saint Coran
  - Calendrier des prières
  - Agenda
  - Cartes postales
  - Guide des Mosquées
  - Pages jaunes Musulmanes
  - Apprendre l'arabe
- Guide
  - Qu'est ce que l'Islam ?
  - Comment faire la prière ?
- Dossiers
  - Attentats du 11 septembre et leurs conséquences

Edition du 30/11/2001

### Le Ramadhan quelques aspects spirituels, sociaux et médicaux

Par Docteur Abdallah (28/11/2001)



Ramadan est un des mois sacrés du calendrier lunaire musulman. Durant cette période, il est prescrit aux croyantes et aux croyants pubères, conscients et en bonne santé de jeûner du début de l'aube au coucher de soleil.

Ce jeûne implique l'absence d'absorption d'aliments, qu'ils soient liquides ou solides, ainsi que l'interdiction de relations charnelles entre époux.

### Je suis diabétique, puis je faire le Ramadhan ?

Par Docteur Abdallah (27/11/2001)

Ta maladie te dispense de l'obligation de jeûner notamment parce que jeûner sous insuline nécessite des conditions préalables impératives et que jeûner te fais courir un risque non négligeable pour ta santé. Tu peux donc manger durant le mois sacré de Ramadan et payer l'équivalent de repas pour les pauvres en compensation, même chose pour les jours que tu n'as pas fait précédemment.

### Ils sont fous ces Ricains !

Par Farid El Asri (25/11/2001)

Farid Al Asri évoque avec sa verve habituelle le bombardement par les américains de la ville historique d'Hérat. Un bombardement qui n'a pas suscité l'indignation de la communauté internationale qui avait protesté énergiquement à l'époque contre la destruction des bouddhas de Bamiyan par les Talibans. La ville d'Hérat, fleuron asiatique et patrimoine de l'humanité a été au cours de l'histoire

Abonnez vous !  
Abonnez vous gratuitement au site Oumma.com

vosre email  
OK  
 s'abonner  
 se désabonner

Prière  
Paris et Ses Environs  
Angle = 15°  
Fajr 06:41:00  
Chourok 08:20:00  
Dhor 12:44:00  
Asr 14:40:00  
Maghreb 17:02:00  
Icha 18:38:00

Horaires des prières des autres villes en France ou à l'étranger

Rechercher

Mots clés:





**A GLOBAL NETWORK OF  
EXPERTS ARCHIVING THE WEB  
FOR FUTURE GENERATIONS.**

[Learn More About IIPC](#)



**INTERNATIONAL**

**I**

**P**

**C**



# Cadre juridique du DL Web en France

## STATISTIQUES GÉNÉRALES DE L'ARCHIVE

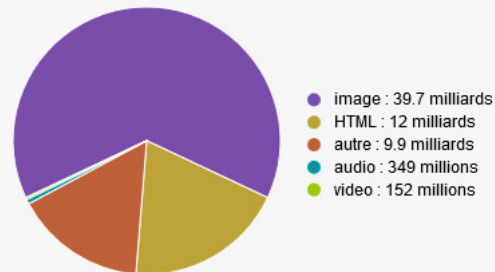
à la date du 11 octobre 2017

Nombre de sites : **14 418**  
 Nombre de versions d'URL : **62.1 milliards**  
 Taille de l'archive : **4.82 Po**  
 Compression DAFF : **95.1%** sur les données web, **84.2%** en incluant les données hors web

## NOMBRE D'URL

L'archive compte **62.1 milliards** de versions d'URL, dont 152 millions pointant vers de la vidéo et 349 millions vers de l'audio.

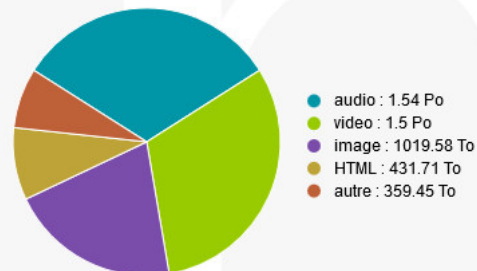
Les pages HTML et les images qu'elles intègrent constituent la plus grande partie des URL collectées.



## TAILLES

L'archive pèse **4.82 Po**, dont 1.5 Po de vidéo et 1.54 Po d'audio.

En dépit de leur nombre, les pages HTML ne constituent pas une masse énorme.



ina  
THEQUE

NAVIGATION

🔗 URL

📄 Sites

RECHERCHE

🌐 Web

📺 Vidéos

🐦 Twitter

STATISTIQUES

📊 Chiffres clés

📊 Stats Twitter

AIDE

💬 FAQ

📖 Guide

🔍 ? Assistance

e  
que

au  
et  
ion  
(11)

# Archives du Web Média – Inathèque (Borne MMSH)



Toute la TV  
Radio  
Web média

Recherche

Galerie

Guide utilisateur

FAQ

Poser une question

l'Archive en chiffres

## ACCÈS PAR URL

Accès à une page ou une ressource identifiée si vous connaissez précisément son URL. Une liste des versions archivées pour sera proposée.

CHERCHER



## ACCÈS PAR SITE

Tapez le nom d'un site pour un filtrage dans la liste des sites archivés par l'Ina. L'ensemble de la liste des sites est consultable en cliquant directement sur le bouton *Lister*.

CHERCHER



## ACCÈS PAR LE TEXTE

Requête par moteur de recherche *fulltext* dans toutes les pages et vidéos archivées.

# Sites : accès par la base de données

**ina**  
THEQUE

NAVIGATION

- URL
- Sites

RECHERCHE

- Web
- Vidéos
- Twitter

STATISTIQUES

- Chiffres clés
- Stats Twitter

AIDE

- FAQ
- Guide
- Assistance

Corpus: dlweb

Général	Documentation	Secteur d'activité	Fonction	Média	Couverture	Thématique	Personnalité	Programme	Public	Service	Statut	Notes	Première capture	Dernière capture
00-csi-manchattan-00scott002.skyrock		commentaire	site marchand	TV		programme		série	tous publics	gratuit	non mis à jour	- 25/08/2011 Blog c2010-09-14		
1-33.com		production/distrib	blog	TV		programme		série	tous publics	gratuit	non mis à jour	- 03/10/11 : blog cc2007-10-16		
1-85.com		production/distrib	site institutionnel	TV		programme		court métrage, mes	tous publics	gratuit, vidéo dispo	collecté	- 23/07/2013 : site		
1coup2d.com		commentaire	site personnel	TV		programme	réalisateur	court métrage, long	tous publics	gratuit, inscription	non mis à jour	- 23/07/2013 : site		
1gars.fille.com		commentaire	site de programme	TV		programme		série	professionnel	collecté	non mis à jour	- 05/04/13 Site du i		
1h15.blogspot.fr		commentaire	blog	TV		programme		série	tous publics	gratuit	mort	- 03/10/11 : site offi2001-04-02		2012-09-03
01net.tv	NextRadioTV	diffuseur	web TV	web		programme		interview entretien	tous publics	gratuit	non mis à jour	- 11/03/15 Blog d'in2015-03-11		
1interval.com		publicité et commu	service	TV		technique audiovis			tous publics	gratuit, vidéo dispo	collecté	- 19/03/14 01netTV		
1oakfilms.com		production/distrib	site institutionnel	TV				message publicita	professionnel	gratuit	collecté	- 10/10/2015 1nterv		
1onair.net		diffuseur	web radio	web				tous publics	tous publics	gratuit, vidéo dispo	collecté	- 25/07/2013 : site 2012-03-06		
1plus1production.c		production/distrib	site institutionnel	TV		programme		message publicita	tous publics	gratuit, audio disp	mort	- 31/08/11 : depuis 2007-06-02		2011-06-06
01podcast.com		production/distrib	site institutionnel	TV		programme		documentaire	professionnel	gratuit, vidéo dispo	collecté	- 03/10/11 : site de 2004-04-03		
2bullsproductions.c		référencement	annuaire	radio				documentaire	tous publics	gratuit, audio disp	mort	- 03/10/11 : annuaire de podca2005-10-23		2009-08-23
2d3d-animations.cc		production/distrib	site institutionnel	TV		programme		long métrage, docu	tous publics	gratuit	collecté	- 23/07/2013 : site		
2factory.com		production/distrib	site institutionnel	TV, web		programme		animation/dessin a	tous publics	gratuit	collecté	- 23/07/2013 : site		
2fproductions.fr		production/distrib	site institutionnel	TV		programme		animation/dessin a	professionnel	gratuit, vidéo dispo	collecté	- 15/12/2014 Agenc		
2gm-normandie.co		production/distrib	site institutionnel	TV		programme		documentaire	professionnel	gratuit	mort	- 03/10/11 Site de l2007-10-12		2012-11-12
2good.fr		événement/festiv	site institutionnel	web		actualité des média			tous publics	gratuit, audio disp	mort	- 09/04/14 Le site "		2014-05-16
2h56.com		production/distrib	site institutionnel	TV, radio					tous publics	gratuit, vidéo dispo	mort	- 03/10/11 : site de 2003-02-12		2010-04-22
2minutes2share.co		production/distrib	site institutionnel	TV		programme		message publicita	professionnel	gratuit, vidéo dispo	collecté	- 02/02/2016 - Site 2016-02-02		
2minutes.fr		groupe média	site institutionnel	web		programme		web fiction vidéo	tous publics	gratuit, vidéo dispo	collecté	- 24/04/15 2 minutes2015-04-24		
2p2l.com		production/distrib	site institutionnel	TV		programme		animation/dessin a	tous publics	gratuit, vidéo dispo	collecté	- 24/07/2013 : site 2001-05-19		
2p2blog.canalblog		production/distrib	site institutionnel	TV		programme		documentaire, jour	tous publics	gratuit	collecté	- Site de la société c2000-10-17		
2rives.tv		commentaire	blog	TV		programme		documentaire, jour	tous publics	gratuit	non mis à jour	- 18/08/11 Blog de 2011-08-13		
2secretstory.skyroc		diffuseur	site de chaîne	TV	internationale			documentaire	tous publics	gratuit	collecté	- 18/12/2015 2 Rive2015-12-18		
3dized.com		commentaire	blog	TV		programme		documentaire	tous publics	gratuit	mort	- 03/10/11 : blog de2008-06-11		2015-03-24
3dmedias.blogspot		production/distrib	site institutionnel	TV		technique audiovis		documentaire	tous publics	gratuit, audio dispo	collecté	- 23/07/2013 : site		
3dxf.com		entité institution	blog	TV		éducation et ensei		documentaire	professionnel	gratuit, vidéo dispo	collecté	- 03/10/11 : site d'u2009-09-04		
3emegauche.tv		commentaire	portail	TV		technique audiovis		documentaire	professionnel	gratuit, audio dispo	collecté	- 27/05/2015 3DVF2015-05-27		
3grenouilles.blogsp		diffuseur	web TV	TV	locale				tous publics	gratuit, vidéo dispo	collecté	- 22/02/12 La web '2010-03-31		
3hommeset1podca		commentaire	blog	web		programme		animation/dessin a	tous publics	gratuit, vidéo dispo	collecté	- 24/07/2013 : blog2010-07-22		
3is.fr		diffuseur	site de programme	web		programme		animation/dessin a	tous publics	gratuit, vidéo dispo	collecté	- 06/11/13 Podcast		
3lezards.com		entité institution	site institutionnel	TV, web		éducation et ensei		documentaire	professionnel	gratuit, audio dispo	collecté	- Institut Internations2001-03-10		
3s-fr.com		production/distrib	site institutionnel	TV, web		programme		documentaire	professionnel	gratuit, vidéo dispo	collecté	- 24/07/2013 : site 2007-12-13		
3toon.com		production/distrib	site institutionnel	TV		programme		documentaire	professionnel	gratuit	non mis à jour	- 3S est une société2004-02-20		
3wdoc.com		publicité et commu	site de programme	web		programme		documentaire	professionnel	gratuit, vidéo dispo	collecté	- 30/05/14 Société d		
3xplus.com		production/distrib	service	web		technique audiovis		documentaire	professionnel	gratuit	mort	- Site de 3WDOC, e2010-12-10		2014-04-06
14418 sites		production/distrib	site institutionnel	TV, web				documentaire	tous publics	gratuit, vidéo dispo	collecté	- 24/07/2013 : site 2001-04-01		

# Archives du Web Média – Inathèque (Borne MMSH)

<http://www.beurfm.net/>

CHERCHER

27 198 versions archivées pour cette URL, accessibles en cliquant sur l'heure souhaitée.  
Cliquez sur les bandeaux de mois et d'années pour déplier/replier la liste des versions jour par jour.



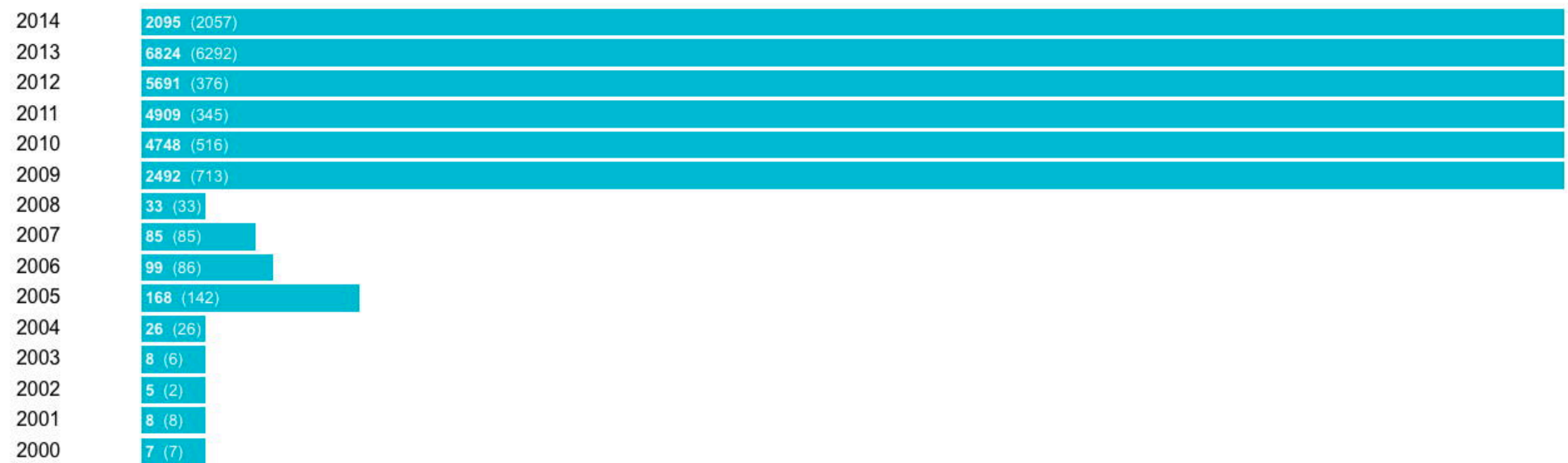


Toute la TV  
Radio  
Web média

## Versions archivées d'une URL

CHERCHER

27 198 versions archivées pour cette URL, accessibles en cliquant sur l'heure souhaitée.  
Cliquez sur les bandeaux de mois et d'années pour déplier/replier la liste des versions jour par jour.



# The BnF Web archives

The screenshot shows the BnF Archives de l'Internet website. At the top, there is a navigation bar with the BnF logo, the title 'Archives de l'Internet', and links for 'accueil' and 'aide'. Below this is a secondary navigation bar with 'Outils : Recherche par URL | Recherche par mot | Parcours guidés |' and a 'NWW' logo. The main content area is divided into three columns. The left column contains introductory text: 'Les archives sont constituées de sites internet du domaine français archivés de 1996 à aujourd'hui.' and 'à propos des archives de l'Internet...'. The middle column features 'Recherche par URL' with a search box containing 'http://', a 'Remonter le temps' button, and an 'Option' section with a dropdown menu for ' limiter la recherche à cette année :'. The right column features 'Recherche par mot' with a search box, a 'Rechercher' button, and a 'Possibles :' section listing search criteria like 'une expression : "François Bon"'. The bottom section is titled 'Parcours guidés' and includes a sub-section 'Images amateurs, amateurs d'images' with a text block and a small image gallery. A final section 'La révolution tunisienne à travers le Web' is partially visible at the bottom.

accueil | aide

BnF Archives de l'Internet

Accès expérimental

Outils : Recherche par URL | Recherche par mot | Parcours guidés |

Les archives sont constituées de sites internet du domaine français archivés de 1996 à aujourd'hui.

à propos des archives de l'Internet...

**Recherche par URL**

Retrouver un site, une page, un fichier en indiquant son adresse internet (exemple : <http://www.inventaire-invention.com>).

**Remonter le temps** Recherche avancée

Option

limiter la recherche à cette année :

**Recherche par mot**

Retrouver ces mots dans la partie indexée des archives (environ 5%, documents archivés en nov-déc 2006 et 2007).

**Rechercher** Recherche avancée

Possibles :

- une expression : "François Bon"
- un mot sur un site : site:www.zazieweb.fr éditeurs

**Parcours guidés** Tous les parcours ...

Découvrir le contenu des archives et se familiariser avec les outils de recherche et de consultation.

**Images amateurs, amateurs d'images**

Le développement du Web, associé à celui des équipements de prise de vue numérique et des logiciels de post-traitement, a offert aux simples particuliers des facilités sans précédent pour publier des photographies et des vidéos. Les archives de l'Internet constituées par la BnF fournissent la matière d'un regard rétrospectif sur une décennie d'usages et de réalisations d'amateurs dans le domaine des images fixes et animées.

**La révolution tunisienne à travers le Web**

# The BnF Web archives

Retrouver ce site ou  
cette page :

[Remonter le temps](#)

Option

limiter à :



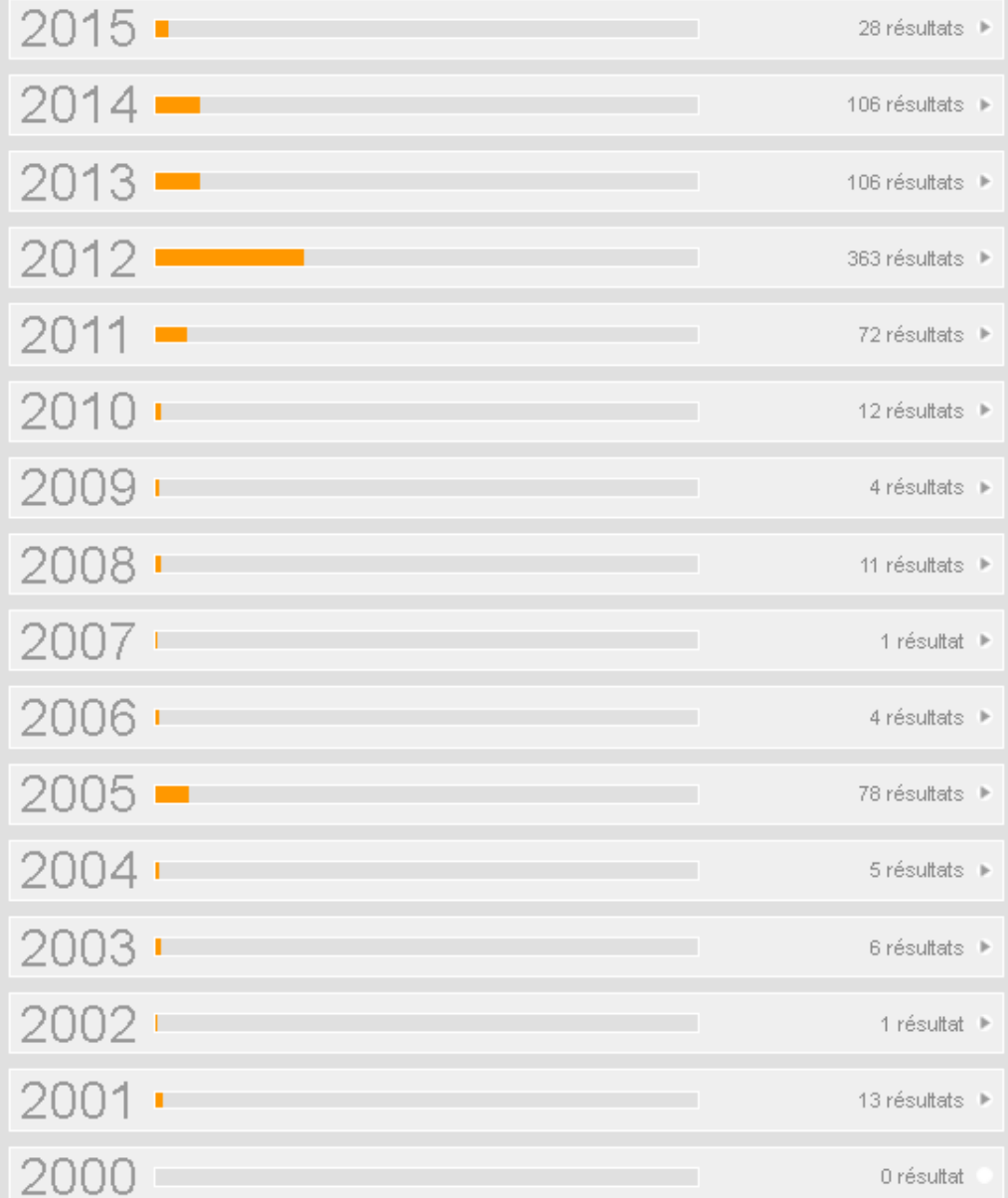
Recherche par URL

pour : <http://www.oumma.com>

810 résultats

du 1 jan. 1996 au 17 avr. 2015

[ouvrir tout](#) | [fermer tout](#)





# Page d'accueil du site Generiques.org (1999-2014), Internet Archive et BnF (Dépôt légal du Web)

[www.generiques.org](http://www.generiques.org)

Page d'accueil (2000-2014)

Wayback machine, Internet Archive

[www.web.archive.org](http://www.web.archive.org)

Pour voir la vidéo, cliquer sur le lien :  
<https://webcorpora.hypotheses.org/380>

# Les risques des Archives du Web

Un processus de *remediation* (N. Brügger)

histoire-immigration.fr Go

442 captures  
5 Dec 2003 - 18 Mar 2017

AUG OCT NOV  
2006 2007 2008

CiTE nationale de l'histoire de l'immigration

PORTES OUVERTES DU 10 AU 14 OCTOBRE 07 ENTRÉE LIBRE

LA CITÉ : L'ACTUALITÉ LE PROJET LES ACTIVITÉS INFOS PRATIQUES Recherche OK

La programmation Le musée Le réseau Les ressources À voir, à lire, à faire

OUVERTURE  
J-179  
CITÉ NATIONALE DE L'HISTOIRE DE L'IMMIGRATION

assistez aux émissions de  
inter bleu France culture  
le 12 octobre de 7h à 20h  
té nationale de l'histoire et

EXPOSITION  
UN SIÈCLE D'IMMIGRATION  
ESPAGNOLE EN FRANCE

Palais de la Porte Dorée. 293, avenue Daumesnil - 75012 Paris  
Métro Porte Dorée / Bus 46 et PC2

LE FILM LE DOSSIER PORTRAIT

ALTÉRITÉS  
Portrait : Idir  
Trans-génération  
Portrait et interview autour de la sortie du CD *La France des couleurs défendra les couleurs de la France*

HOMMES & MIGRATIONS  
Nouvelles figures de l'immigration en France et en Méditerranée  
Hommes & Migrations poursuit sa

ZOOM DE LA SEMAINE

Journées portes ouverture du 10 au 14 octobre !  
À l'occasion de son ouverture, la Cité nationale de l'histoire de l'immigration accueille gratuitement le public pendant cinq jours, du mercredi 10 octobre au dimanche 14 octobre,

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

## Les limites du DL Web

- Comme toutes les archives : elles sont lacunaires
- Territorialisation
- Difficultés d'accès
- La tentation de la généralisation
- Un environnement spécifique
- Un nouveau rapport à l'archive du fait du processus de *remediation* (N. Brügger), recréation (C. Mussou)



# Adaptations épistémologiques et méthodologiques

- Les limites de l'archive
  - Une archive recréée et en cours de constitution
  - Moteurs de recherche
  - Notion de « Web français »
- Prise en compte des spécificités de la source :
  - Web archivé/Web vivant
  - Sélection autour de la validité
  - L'archive, condition nécessaire à toute recherche

# Défi méthodologique, enjeux épistémologiques

- Une garantie pour la stabilité de la source

- Un privilège : une archive en cours d'élaboration => possibilité de déposer les sites et situation favorisant une collaboration plus étroites

Des archives inédites

Une démarche d'historien

- Comprendre la fabrique des archives du web
- Un nouveau rapport à l'archive
- Des outils spécifiques

- Une approche qualitative fondée sur le croisement des sources.
- L'inscription des mémoires en ligne dans le temps long de l'histoire de l'immigration maghrébine.



# Histoire nativement numérique : méthodes et outils

# Constitutions et analyses de corpus massifs

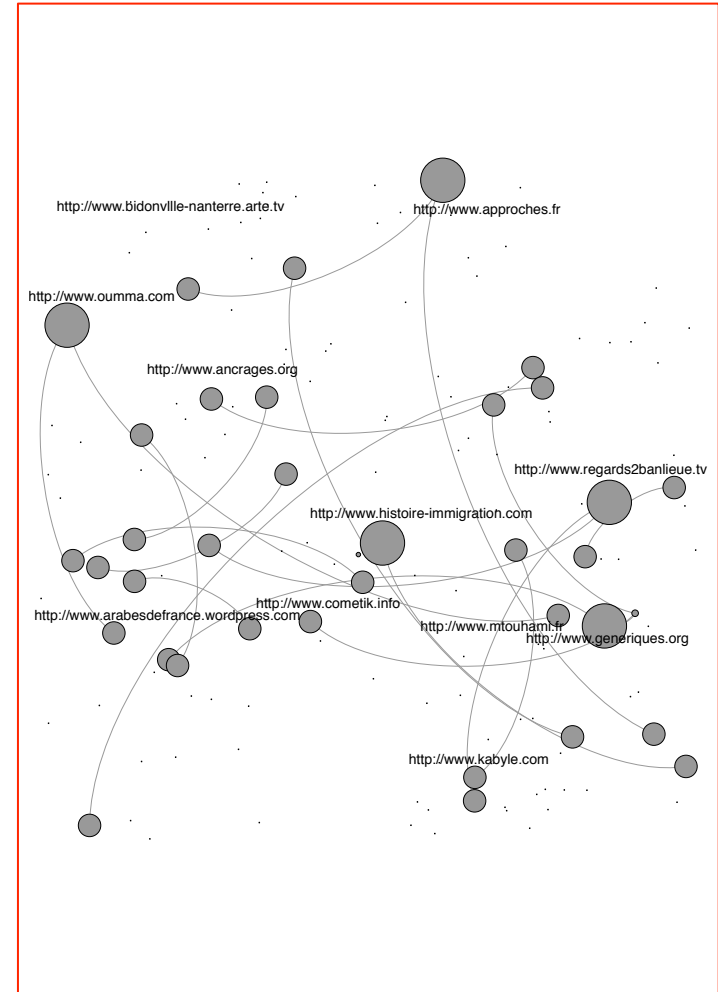
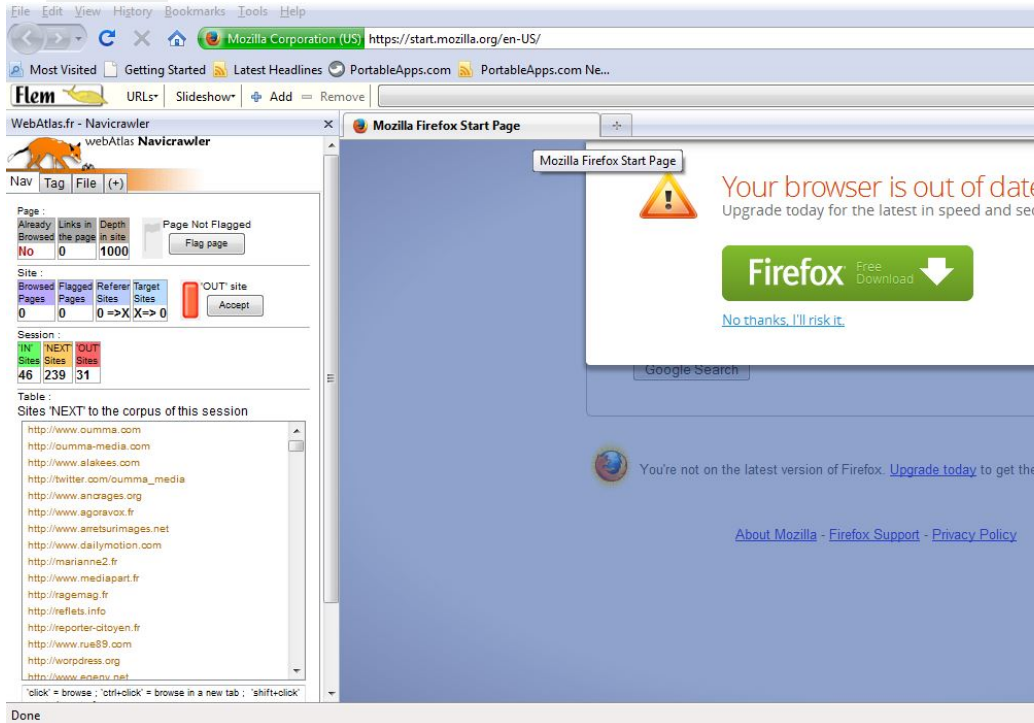
Web vivant + Profondeur diachronique dans l'archive

Base de données (Tableur)

Graphe de liens au sein du corpus

# Constitutions et analyses de corpus massifs

## Navicrawler et Gephi



Corpus online on Navicrawler (Mathieu Jacomy, Sciences Po) in 2013.  
Link Graph of my web corpus online on 2013 (Gephi)



# Constitutions et analyses de corpus massifs

**Valérie Beaudouin, Zeynep Pehlivan. Cartographie de la Grande Guerre sur le Web** : Rappo final de la phase 2 du projet "Le devenir en ligne du patrimoine numérisé : l'exemple de la Grande Guerre". [Rapport de recherche] Bibliothèque nationale de France; Bibliothèque de documentation internationale contemporaine; Télécom ParisTech. 2017. [hal-01425600](https://hal.archives-ouvertes.fr/hal-01425600)

Voir aussi **Frédéric Clavert**, **la Grande Guerre sur Twitter**

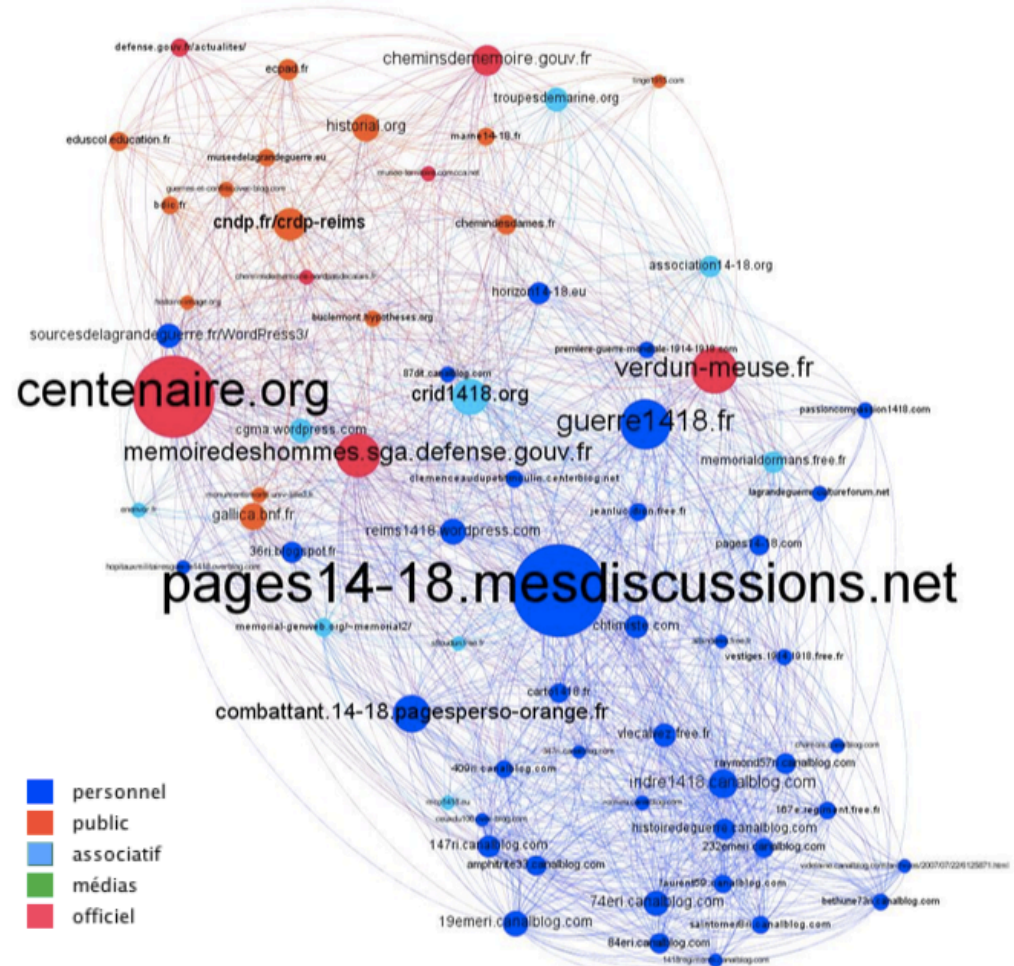


Figure 19. Carte du Web Grande Guerre (collecte de novembre 2014, sites avec degrés >30)



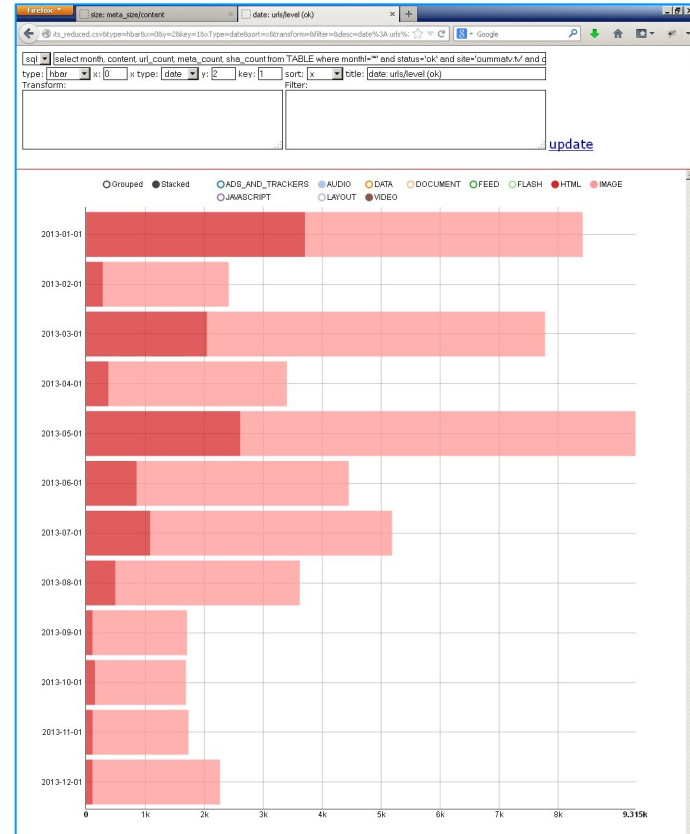
# Quid des outils au sein des archives du Web

Nom usuel du site	
Adresse URL	
Date de consultation	
Chemin de recherche	
Nature du site (fonction)	
Editeur du site	
Archive profondeur temporelle	
Nombre de versions	
Type de média	
Date création site	
Nom usuel du site	
Auteurs identifiés et activité pro.	
Type de site en fonction adresse URL	
Interactivité	
Interconnectivité	
Contenus	
Couleur dominante	
Organisation du site	

Mise en relation avec la problématique :

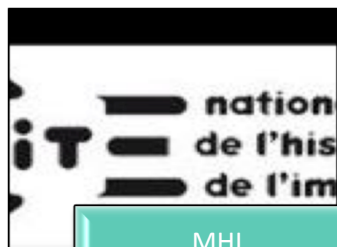
Thèmes	
Sous-thèmes	
Résumé du	

**An analysis file  
(S. Noiret & P. Rygiel)**

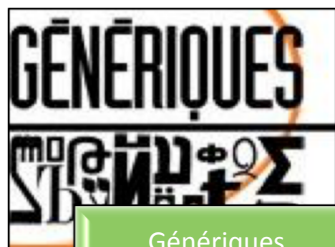


**Data-mining tools**  
David RAPIN, INA, 2013

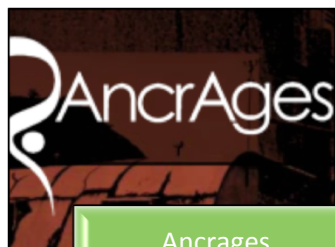
# Pour une histoire sociale et qualitative



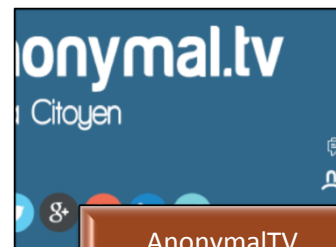
MHI



Génériques



Ancrages



AnonymaTV



Kabyle.com



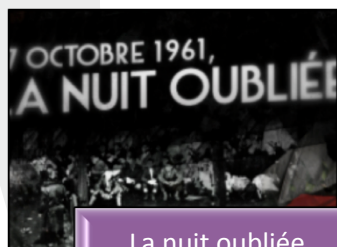
Oumma.com



OummaTV



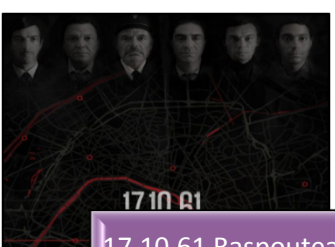
Com'etik Diffusion



La nuit oubliée



127 rue de la Garenne



17.10.61 Raspouteam





Touhami Moualek







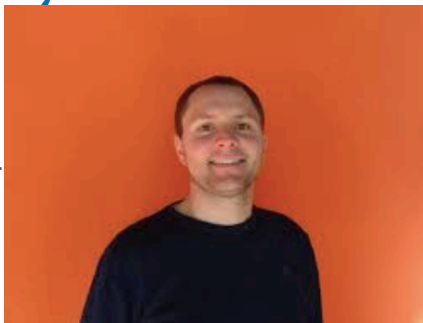
Arabes de France

## Traditionnal players

-  Institutions
-  Memorial associations

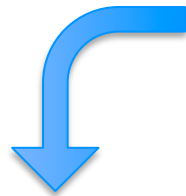
## New players

-  Local Web media
-  Minority Web media
-  Webdocumentary directors
-  Bloggers & Internet users

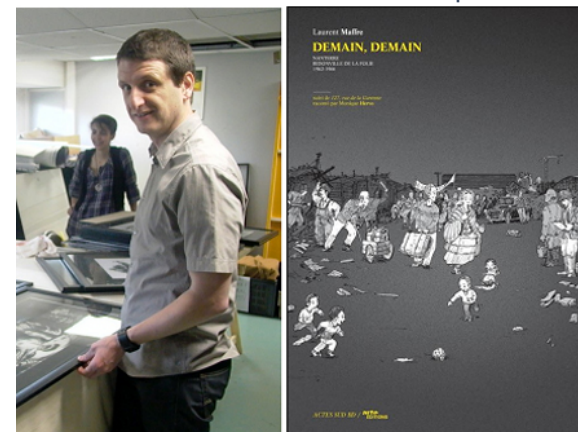
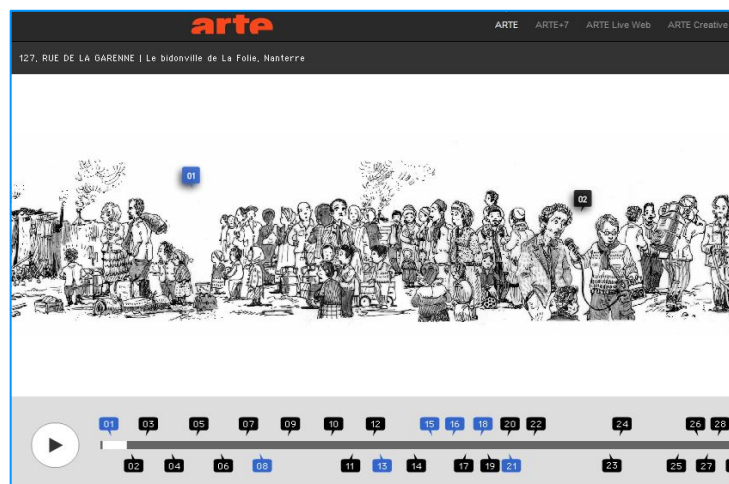


**Gilles Freissinier, ARTE**

<http://leblogdocumentaire.fr/le-pole-web-darte-avec-marianne-levy-leblong-et-gilles-freissinier/>



Enquête orale, passer « derrière l'écran »

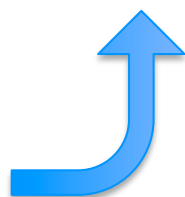


**Laurent Maffre, « Demain Demain »**



**Thomas Gabison, ACTES SUD**

<https://www.fnac.com/Angouleme-2011-Interview-de-Thomas-Gabison-editeur-de-Brechts-Evens-Les-noceurs/cp5175/w-4>



## Entretien avec Monique Hervo



Monique H  
En 1956, e  
En 1959, e  
Elle décide  
rendre cor  
décidera d  
d'outillage  
le bidonvil  
Elle revien  
sur la man

**Monique Hervo,**  
<http://www.histoire-immigration.fr/collections/entretien-avec-monique-hervo>

# Enquête orale en confrontation avec les archives du Web

*Projet e-diasporas (dir. Dana Diminescu 2003) : les enquêtés et les graphes de lien*

Ex : Sabrina Marchandise

Les étudiants Marocains sur Facebook

Espace pratique / espace imaginé

<http://maps.e-diasporas.fr/>



## Confrontation entre les témoins et les archives du web

- débusquer les erreurs d'archivage
- intégrer la question de leur relation au contenu en ligne dans l'enquête (investissement personnel, imaginaire d'internet)
- intégrer les témoins dans la lecture critique des contenus en prenant en compte les conditions de production

## L'histoire nativement numérique :

Des défis épistémologiques et méthodologiques  
(corpus, documents, temporalités)

Ouverture disciplinaire incontournable

Invite à repenser des questions anciennes (document, archive, quanti/quali) tandis que l'ontologie disciplinaire constitue un ancrage précieux.